The Operations and Design of Markets with Spatial and Incentive Considerations

Francisco Castro

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

ABSTRACT

The Operations and Design of Markets with Spatial and Incentive Considerations

Francisco Castro

Technology has greatly impacted how economic agents interact in various markets, including transportation and online display advertising. This calls for a better understanding of some of the key features of these marketplaces and the development of fundamental insights for this class of problems. In this thesis, we study markets for which spatial and incentive considerations are crucial factors for their operational and economic success. In particular, we study pricing and staffing decisions for ride-hailing platforms. We also consider the contract design problem faced by Ad Exchanges when buyers' strategic behavior and inherent business constraints limit these platforms' decisions. Firstly, we investigate the pricing challenges of ride-hailing platforms and propose a general measure-theoretical framework in which a platform selects prices for different locations, and drivers respond by choosing where to relocate based on prices, travel costs, and market congestion levels. Our results identify the revenue-maximizing pricing policy and showcase the importance of accounting for global network effects. Secondly, we develop a queuing approach to study the link between capacity and performance for a service firm with spatial operations. In a classical $M/M/n$ queueing model, the square root safety (SRS) staffing rule balances server utilization and customer wait times. By contrast, we find that the SRS rule does not lead to such a balance in spatial systems. In these settings, a service firm should use a higher safety factor, proportional to the offered load to the power of 2/3. Lastly, motivated by the online display advertising market where publishers frequently use transaction-contingent fees instead of up-front fees, we study the classic sequential screening problem and isolate the impact of buyers? ex-post participation constraints. We characterize the optimal selling mechanism and provide an intuitive necessary and sufficient condition under which screening is better than pooling.

Contents

List of Figures

List of Tables

Acknowledgements

I feel profound gratitude and admiration towards my advisors Omar Besbes, Ilan Lobel and Gabriel Weintraub. I have learned invaluable lessons from my interactions with them that have been not only a fundamental part of my doctoral adventure, but also, I'm sure, will be with me for the rest of my career and life. Omar has taught me the value of spending enough time thinking about research questions before jumping into them; the value of going the extra mile so that your research can achieve the next level; and the importance of communicating your ideas and practicing talks. I'll always be grateful for the immense support and sound life advice he provided during my years at Columbia; I could tell that he took my success personally. Ilan's highly motivated and eager spirit has been an inspiration during the last few years. He showed me the importance of digging deeper to uncover the true value of results; he was supportive and very direct during the times I needed it most; and he pushed me outside of my comfort zone—all of which have been instrumental for my success in my PhD. Gabriel's insightful way of thinking is a quality I aspire to have. He has pushed me to think in intuitive terms and express ideas in a clear fashion. It's not easy to collaborate from a distance, but he was always accessible and happy to discuss not only our research, but any matter that may arise. It has been an honor to see the three of them work and experience firsthand how passionate and smart they are. I would also like to give a special mention to Dirk Bergemann. Collaborating with him has been a privilege. The bar is high, but I hope someday to be like all four of these mentors.

I am fortunate to have been a graduate student in the DRO division of Columbia Business School, a rich intellectual environment of high standards that has constantly pushed me to be the best version of myself, but also a place where I felt like part of a team in which everyone wanted the best for me. I would particularly like to thank Carri Chan, Jing Dong, Santiago Balseiro, Fanyin Zheng, Nelson Fraiman, Costis Maglaras and Yash Kanoria for their support over the years. I always felt great admiration for and support from Carri who was happy to provide feedback on my writings and talks anytime. Jing and Santiago arrived to DRO not long ago, but they provided great feedback and asked tough questions during my job market preparation.

This adventure was even better than expected thanks to the excellent cohort of students I've had a chance to interact with. I have had a great time exchanging ideas and working on problem sets with Amine, Nico, Mauro and Pu. I have enjoyed spontaneous research conversations with Yaarit after a talk or when she would come by the cubicle. I would like to single out my cube-mate and friend Vashist. Talking about research and life with him was a pleasure, and I appreciate the honest advice he gave me for the job market. I would also like to thank Geun Hae with whom I would go for an afternoon walk around campus and talk about research, philosophy and the future.

The path to the PhD started 5 years ago in Santiago. I'm thankful for the rigorous training I received from the math department at the University of Chile. I'll never forget my measure theory class with professor Jaime San Martin who at the end of the semester spoke about inequality and our social responsibility as students of the University of Chile. I look up to my undergrad advisor Rafael Epstein who is an example of bridging the gap between operations research and practice. I would also like to thank Denis Saure for his support for my undergrad thesis and advice before starting the job market. A special mention to some of the most talented classmates

I ever had, Alf and Mauro, with whom I not only share similar academic interests but also a great and heartfelt friendship. Mauro has been my constant companion in New York and has made my life away from home easier. Although he does not live as close by, it feels as though Alf and I can pick up right where we left off; he has been my partner in exploring everywhere from Patagonia to the Grand Canyon.

I am thankful for the support I received from my childhood friends, “los bris,” throughout grad school but also, more generally, throughout my life. Pepa, who I’ve known since age 6, has always been there for me and even though I messed things up from time to time. I’m lucky to have grown up alongside Raimundo and Felipe (Gas) from whom I have learned important life lessons and who have cheered me up during grad school. I am also grateful to Eduardo (Chutz) and Carlos (Charlos) who, together with the rest of the gang, would always make time to catch up during my brief visits back home. One cannot ask for better friends.

I feel lucky to be a part of the family that I’m in. My older brother Felipe was my first teacher; he opened up my mathematical curiosity. His tenacity, intellect and passion are qualities that I hope one day to have. I admire the determination with which my sister Coni has pursued her dreams. She has been an example of how hard work and love for what you do can get you wherever you want. My father Carlos is one of the best people I know. With his good heart, he has always been there to listen and to do everything he could to help. My mother Gloria could not be more strong or more fierce; she has overcome countless obstacles in life, including cancer. I will never forget what she would say when dropping me off at school before an exam: “be judicious.” Being away from my family, especially being away from my mom while she was fighting cancer, has been one of the biggest challenges of grad school.

Lastly, I would like thank Kaitlin. She has been my pillar during grad school. She helped improve my writing and has always been willing to listen to my research struggles with the hope that by explaining things to her I would find my answer. Even today, I'm not sure what was going through her mind during those conversations. Her constant caring, love and support have not only been essential during these years but have also made me realize what's important in life. I profoundly admire and love her, and I am excited to start the next adventure holding her hand.

To my parents - Gloria Altamirano & Carlos Castro who have sacrificed a great deal to get me to where I stand today.

Introduction

Marketplaces are a fundamental part of how agents in society interact. Before the internet, most of these interactions occurred in a physical fashion. In search for basic goods consumers would go to a nearby store; for transportation they would take the bus, subway or a cab; for information they would look in newspapers or magazines. However, technological developments have fostered exciting changes in almost all marketplaces which, in turn, have forever changed the way economic agents interface with each other. Now people can shop online and have their goods delivered to their homes within two days. Instead of hailing a cab on the street, consumers can now "Uber" to anywhere they need to go right from their front door. The search for information is now at the palm of our hand, easier than ever. These innovations have impacted virtually every industry, from retail, to transportation, to advertising, and beyond. There is a great deal of excitement and interest in the academic community for understanding the new practical challenges these industries face; in turn, there is equal excitement for designing policies and selling mechanisms to address those challenges. In this thesis we explore practical economic and operational considerations for a select group of online marketplaces that have recently revolutionized their industries. In particular, we study ride-hailing systems and aim to understand how to better design pricing and staffing policies while keeping in mind the spatial nature of this market. We also explore online advertising through the lens of mechanism design, considering buyers with rational behavior particular to this market that constrains the way sellers can sell impressions.

Ride-hailing services such as Uber, Lyft, and DiDi have changed the way people move in cities. For example, from 2013 to 2017 the number of average weekday taxi trips in New York has declined by approximately 100,000, and it has nearly been matched by on-demand transportation platforms.[1] On these platforms, riders can now seamlessly request rides from their smartphones, while drivers possess information about the system that helps them make real-time strategic decisions about when and where to work. This has created an environment of unprecedented complexity that prompts exciting practical and academic questions. This complexity stems from both their spatial operational nature and the presence of strategic self-interested agents. For instance, managing supply-demand imbalances in space entails solving high-dimensional optimization problems in which complicated network effects have to be taken into account. Strategic interactions between agents add yet another layer of complexity, as the right incentives must be in place. In Chapter 1 and 2, we consider these challenges and bring a new understanding to classical questions in operations and revenue management.

In Chapter 1 we study how a revenue-maximizing ride-hailing platform should select prices across city locations while taking into account drivers' strategic repositioning behavior. We use a general game-theoretical framework that accounts for spatial frictions that arise due to congestion and driving costs to elucidate the interplay between local and global price effects. Local changes in price might have a local effect on demand but, since supply is strategic and can reposition, they might induce a non-trivial global supply response. To tackle this challenge we first establish that the platform's optimization problem can be decoupled into local subproblems associated with smaller regions of the city, each of which can be solved via a coupled, bounded knapsack relaxation. Then, by pasting these local solutions together we obtain the global optimal solution. Our solution showcases a surprising insight that

[1] Fix, N.Y.C. "Advisory Panel Report" (2018).

highlights how space impacts the design of optimal prices and drivers' strategic behavior: in order to incentivize the repositioning of drivers to high-demand areas, the platform can damage regions where drivers are not needed and, by doing so, boost revenues. These damaged regions are characterized by low prices and high congestion, the combination of which creates enough incentive to steer drivers to locations that are more profitable for the platform. The framework we develop has applications in other settings, e.g., where strategic workers must plan their working schedules or spatial equilibrium models of labor mobility.

Another central matter in operations management is capacity planning. For a traditional multi-server queueing system it is well known that in heavy traffic a square-root staffing (SRS) rule can maintain the balance between customers' waiting time and servers' efficiency (QED regime). In systems where customers arrive to random locations in space, such as ride-hailing platforms or automated warehouses, and servers have to spend time not only servicing customers but also reaching them before service starts, this balance may no longer hold. How should "capacity thinking" be adapted in such settings? In Chapter 2, we analyze this question. We consider a Markovian stochastic system that captures the key aspects of a spatial multi-server system. We establish that, in stark contrast with a standard multi-server system, the SRS rule brings the spatial multi-server system to the ED (efficiency driven) regime. The reason is that, because customers have to be reached before service starts, the time a server spends on them is larger than in a standard queuing setting and, therefore, more servers are required to achieve QED performance. In addition, we fully characterize the system's performance under a range of scalings, thereby showing how it shifts from the ED to QD (quality driven) regimes by passing through the QED regime. Interestingly, reaching the QED regime in our model is more subtle. It can only happen when the buffer term in the classic SRS staffing formula is raised to the power of 2/3 instead of 1/2, and for a specific value of the SRS parameter. Our results suggest that in a spatial setting, operating in the QED regime depends not only on the rate at which we scale the system but also on how we approach such a rate. The results in this paper imply that common rules of thumb such as the SRS rule will no longer be valid for firms that operate in space and, therefore, new staffing rules of thumb are necessary. This has implications for fleets of self-driving cars and for how to think about trade-offs for this fast-approaching technology.

A market that has drawn a great deal of attention in the Revenue Management community is online display advertising. The wide adoption of auctions as the predominant selling mechanism in this market showcases the existence of a type of "business constraint": buyers never pay more than they are willing to pay for impressions. In addition, it is common that for the same impression multiple auctions are used to provide different service levels to buyers and, by doing so, to price-discriminate them. An important practical example are the so-called "waterfall auctions," in which bidders can decide to participate in one of two auctions: (1) an auction with "first-look" priority but a high reserve price, or (2) another with access only to the leftover inventory that was not cleared in the first auction, but a low reserve price. The purpose of this mechanism is screening; high valuation buyers should select the first auction and low valuation buyers should select the second one. A natural practical question is whether this is an effective price discrimination device. This brings to the forefront the question of how to design an optimal screening selling contract assuming that buyers satisfy ex-post individual rationality; that is, like in typical auctions, buyers are always willing to participate even after learning their valuation. In Chapter 3 we isolate the essential parts of this problem and address it using a mechanism design formulation. We study the problem faced by a monopolist selling a single item to a two-type buyer who privately, and sequentially, learns her valuation in two stages. The distinctive feature of our problem is that after the buyer completely learns her valuation she is still willing to buy the item. Leveraging a connection with marginal revenues, we obtain a full characterization of the optimal selling mechanism and establish that its structure depends on an intrinsic economic quantity that we call profit-to-rent ratio. It measures the change in the seller's revenue per unit of information rents given to the buyer. We show that, depending on how this economic quantity behaves around the optimal posted price, the optimal contract can be either a simple posted price that pools types or a more elaborate randomized mechanism that separates types. The latter contract randomizes the low-type buyer and offers her a low price, while it allocates with certainty the item to the high-type buyer and offers her a high price. Importantly, despite the fact that we are in a setting with one buyer and a single item, the presence of ex-post participation constraints makes our optimal solution different from the classic bang-bang solution in mechanism design. Moreover, we establish that the randomized contract can outperform the posted price contract by up to 25%. Finally, we also provide extensions to the setting with an arbitrary number of types.

Chapter 1

Surge Pricing and Its Spatial Supply Response

1.1 Motivation and Overview of Results

Pricing and revenue management have seen significant developments over the years in both practice and the literature. At a high level, the main focus has been to investigate tactical pricing decisions given the dynamic evolution of inventories, with prototypical examples coming from the airline, hospitality and retail industries ([64]). With the emergence and multiplication of two-sided marketplaces, a new question has emerged: how to price when capacity/supply units are strategic and can decide when and where to participate. This is particularly relevant for ride-hailing platforms such as Uber and Lyft. In these platforms, drivers are independent contractors who have the ability to relocate strategically within their cities to boost their own profits. On the one hand, this leads to a more flexible supply. On the other hand, one is not able to simply reallocate supply across locations when needed, but rather a platform needs to ensure that incentives are in place for a "good" reallocation to take place. Consider the spatial pricing problem within a city faced by a platform that shares its revenues with drivers. Suppose there are different demand and supply conditions across the city. The platform may want to increase prices at locations with high demand and low supply. Such an increase would have two effects. The first effect is a local demand response, which pushes the riders who are not willing to pay a higher price away from the system. The second effect is global in nature, as drivers throughout the city may find the locations with high prices more attractive than the ones where they are currently located and may decide to relocate. In turn, this may create a deficit of drivers at some locations. In other words, prices set in *one region* of a city impact demand and supply at this region, but also potentially impact supply in *other regions*. This brings to the foreground the question of how to price in space when supply units are strategic.

The central focus of this chapter is to understand the interplay between spatial pricing and supply response. In particular, we aim to understand how to optimally set prices across locations in a city, and what the impact of those prices is on the strategic repositioning of drivers. To that end, we consider a short-term model over a given timeframe where overall supply is constant. That is, drivers respond to pricing and congestion by moving to other locations, but not by entering or exiting the system. In our short-term framework, the platform's only tool for increasing the supply of drivers at a given location is to encourage drivers to relocate from other places. In turn, this time scale permits us to isolate the spatial implications on the different agents' strategic behavior. In this sense, our model can be thought of as a building block to better understand richer temporal-dynamic environments.

In more detail, we consider a revenue-maximizing platform that sets prices to match price-sensitive riders (demand) to strategic drivers (supply) who receive a fixed commission. In making their decisions, drivers take into account prices, supply levels across the city, and transportation costs. More formally, we consider a measure-theoretical Stackelberg game with three groups of players: a platform, drivers and potential customers. Supply and demand are non-atomic agents, who are initially arbitrarily positioned. We use non-negative measures to model how these agents are distributed in the city. All the players interact with each other in two dimensional city. Every location can admit different levels of supply and demand. The platform moves first, selecting prices for the different locations around the city. Once prices are set, the mass of customers willing to pay such levels is determined. Then, drivers move in equilibrium in a simultaneous move game, choosing where to reposition based on prices, supply levels and driving costs. In fact, besides prices and transportation costs, supply levels across the city are a key element for drivers to optimize their repositioning. If too many other drivers are at a given location, a driver relocating there will be less likely to be matched to a rider, negatively affecting that driver's utility. The platform's optimization problem consists of finding prices for all locations given that drivers move in equilibrium.

Main contributions. Our first set of contributions is methodological. We propose a general framework that encompasses a wide range of environments. Our measure-theoretical setup can be used to study spatial interactions in both discrete and continuous location settings. In this general framework, our main result provides a structural characterization of the optimal prices, and resulting equilibrium driver movement in regions of the city where drivers relocate. In particular, we first establish that the platform's objective can be reformulated as a function of only the equilibrium utilities of drivers and their equilibrium post-relocation distribution. In turn, we develop structural properties on these two objects. We first characterize properties of the drivers' equilibrium utilities and prove that the city admits a form of spatial decomposition into regions where movement may emerge in equilibrium, "attraction regions," and the rest of the city. Furthermore, we establish that the equilibrium utility of drivers and the local equilibrium post-relocation supply are linked through a congestion bound. The former admits a fundamental upper bound parametrized by the latter. Driven by these properties and our objective reformulation, we derive a relaxation to the platform's problem that takes the form of *coupled continuous bounded knapsack* problems. Notably, we establish that this relaxation is tight and in turn, leveraging the knapsack structure, we obtain a crisp structural characterization of an optimal pricing solution and its supply response.

In our second set of contributions, we shed light on the scope of prices as an incentive mechanism for drivers and provide insights into the structure of an optimal policy. To that end, we study a special family of cases in a linear city environment in which a central location in the city, the origin, experiences a shock of demand. To put the optimal policy in perspective, we first characterize an optimal *local price response* policy, a pricing policy that only optimizes the price at the demand shock location. Such a policy increases prices at the demand shock location leading to an attraction region around the shock in which drivers move toward the origin.

Leveraging our earlier methodological results in conjunction with the derivation of new results, we characterize in quasi-closed form the optimal pricing policy and its corresponding supply response. The optimal policy admits a much richer structure. Quite strikingly, the optimal pricing policy induces movement toward the demand shock but potentially also *away* from the demand shock. The platform may create *damaged regions* through both prices and congestion to steer the flow of drivers toward more profitable regions. Compared to the *local price response* policy, the optimal solution or *global price response* incentivizes more drivers to travel toward the demand shock.

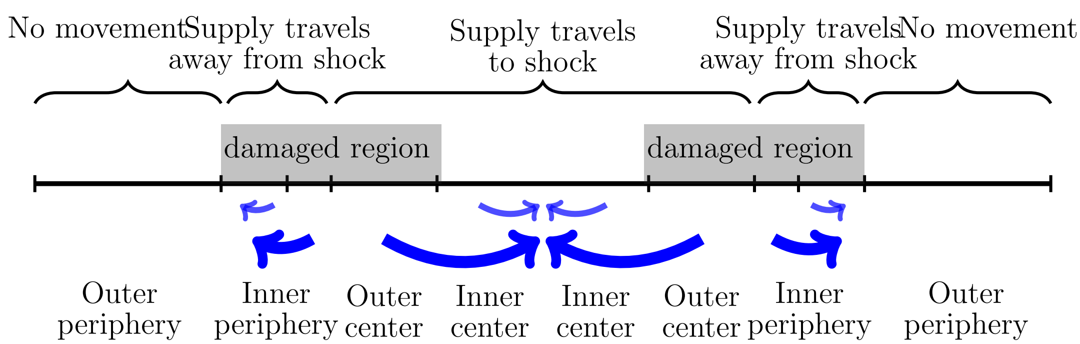


Figure 1.1: **The optimal solution creates six regions.**

The optimal pricing policy splits the city into six regions around the origin (Figure 1.1). The mass of customers needing rides at the location of the shock is serviced by three subregions around it: the origin, the inner center and the outer center. The origin is the most profitable location and so the platform surges its price, encouraging the movement of a mass of drivers to meet its high levels of demand. These drivers come from both the inner and outer center. In the former, locations are positively affected by the shock, and some drivers choose to stay in them while others travel toward the origin. In the latter, drivers are too far from the demand shock and so the platform has to *deliberately damage* this region through prices (e.g., to shut down demand) to create incentives for drivers to relocate toward the origin. However, drivers in this region have an option: instead of driving toward the demand shock at the origin, they could drive away from it. This gives rise to the next region, the inner periphery. Consider the marginal driver, i.e., the furthest driver willing to travel to the origin. To incentivize the marginal driver to move to the origin, the platform is obligated to also damage conditions in the inner periphery. The optimal solution creates two subregions within the inner periphery. In the first, conditions are degraded through prices that make it unattractive for drivers. Drivers in this region leave toward the second region. That is, they drive in the direction opposite to the demand shock. The action of the platform in the second region is more subtle. Here, the platform does not need to play with prices. The mere fact that drivers from the first region run away to this area creates congestion, and this is sufficient degradation to make the region unattractive for the marginal driver. The final region is the outer periphery, which is too far from the origin to be affected by its demand shock.

We complement our analysis with a set of numerics that highlights that the optimal policy can generate significantly more revenues than a local price response. In other words, anticipating the global supply response and taking advantage of the full flexibility of spatial pricing plays a key role in revenue optimization.

1.2 Related Literature

Several recent papers examine the operations of ride-hailing platforms from diverse perspectives. We first review works that do not take spatial considerations into account. There is a recent but significant body of work on the impact of incentive schemes on agents' participation decisions. [35] study the cost of self-scheduling capacity in a newsvendor-like model in which the firm chooses the number of agents it recruits and, in each period, selects a compensation level as well as a cap on the number of available workers. [22] analyze various compensation schemes in a setting in which the platform takes into account drivers' long-term and short-term incentives. They establish that in high-demand periods all stakeholders can benefit from dynamic pricing, and that fixed commission contracts can be nearly optimal. The performance of such contracts in two-sided markets is analyzed by [40] who derive performance guarantees. [65] considers how uncertainty affects the price and wage decisions of on-demand platforms when facing delay-sensitive customers and autonomous capacity. [53] focuses on the effect of market thickness and competition on wages, prices and welfare and shows that, in some circumstances, more supply could lead to higher wages, and that competition across platforms could lead to high prices and low consumer welfare.

In the context of matching in ride-hailing without pricing, [31] compare the waiting time performance, in a circular city, of on-demand matching versus traditional street-hailing matching. [39] analyze a dynamic matching problem as well as the structure of optimal policies. Relatedly, [54] develop a heuristic based on a continuous linear program to maximize the number of matches in a network. [1] study demand admission controls and drivers' repositioning in a two-location network, without pricing, and show that the value of the controls is large when both capacity is moderate and demand is imbalanced.

Most closely related to our work are papers that study pricing with spatial considerations. [23] take space into account, but only in reduced form through the shape of the supply curve. This chapter points out that surge pricing can help to avoid an inefficient situation termed the "wild goose chase" in which drivers' earnings are low due to long pick-up times. [12] consider a queueing network where drivers do not make decisions in the short-term (no repositioning decisions) but they do care about their long-term earning. They prove that a localized static policy is optimal as long as the system parameters are constant, but that a dynamic pricing policy is more robust to changes in these parameters. [10] find approximation methods to find source-destination prices in a network to maximize various long-run average metrics. Customers have a destination and react to prices, but supply units do not behave strategically. [17] focus on pricing for steady-state conditions in a network in which drivers behave in equilibrium and decide wether and when to provide service as well as where to reposition. They are able to isolate an interesting "balance" property of the network and establish its implications for prices, profits and consumer surplus. [20] structurally estimates a spatial model to understand the welfare costs of taxi fare regulations. These papers investigate long-term implications of spatial pricing. In contrast, our work examines how the platform should respond to short-term supply-demand imbalances given that the supply units are strategic.

From a methodological point of view, our work borrows tools from the literature on non-atomic congestion games. Our equilibrium concept is similar to the one used by [58] and [26] to analyze selfish routing under congestion in discrete settings: in equilibrium, drivers only depart for locations that yield the largest earnings. We consider a more general measure-theoretical environment that can be traced back to [61] and [48]. Our work is also related to the literature on optimal transport (see [18]). Once the platform sets prices, drivers must decide where to relocate. This creates a "flow" or a "transport plan" in the city from initial supply (initial measure) to post-relocation supply (final measure). However, in our problem, the final measure is endogenous.

Finally, some of our insights relate back to the damaged goods literature. [29] explain that a firm can strategically degrade a good in order to price discriminate. In our setting the platform can damage some regions in the city through prices and congestion to steer drivers toward more profitable locations and thus increase revenues.

Our linear city framework relates to the class of Hotelling models [38], which are typically used to study horizontal differentiation of competing firms. In contrast to this classical stream of work, we consider a monopolist who can set prices across all locations. Furthermore, these prices affect the capacity at each location and supply units can choose among all regions of the city to provide service.

1.3 Problem Formulation

Preliminaries. Throughout the chapter, we will use measure-theoretic objects to represent supply, demand and related concepts. This level of generality will enable us to capture the rich interactions that arise in the system through a continuous spatial model. The continuous nature of space simplifies our solution, enabling us to express the solution to special cases of interest in quasi-closed form. To that end, we introduce some basic notation. For an arbitrary metric set $\mathcal{X}$ equipped with a norm $\|\cdot\|$ and the Borel $\sigma-$algebra, we let $\mathcal{M}(\mathcal{X})$ denote the set of non-negative finite measures on $\mathcal{X}$. For any measure τ, we denote its restriction to a set $\mathcal{B}$ by $\tau|_{\mathcal{B}}$. The notation $\tau \ll \tau'$ represents measure τ being absolutely continuous with respect to measure τ'. The notation $\operatorname{ess\,sup}_{\mathcal{B}}$ corresponds to the essential supremum, which is the measure-theoretical version of a supremum that does not take into account sets of measure zero. To denote the support of any measure τ we use $\text{supp}(\tau)$. The notation $\tau - a.e.$ represents almost everywhere with respect to measure τ. For any measure τ in a product space $\mathcal{B} \times \mathcal{B}$, τ_1 and τ_2 will denote, respectively, the first and second marginals of τ. We use $\mathbf{1}_{\{\cdot\}}$ to denote the indicator function, and $S^{\circ}, \partial S, \overline{S}, S^c$ to represent the interior, boundary, closure and complement of a set S respectively. We denote the close and open line segment between two points by $[x, y]$ and (x, y), respectively. When x, y are in the same line segment we write $x \leq y$ or $x < y$ to denote the order in the line segment. If $F(\cdot)$ is a cumulative distribution function, then $\overline{F}(q) = 1 - F(q)$. For consistency, we use masculine pronouns to refer to drivers and feminine ones to refer to customers.

1.3.1 Model elements

Our model contains four fundamental elements: a city, a platform, drivers and potential customers. We represent the city by a convex, compact subset $\mathcal{C}$ of $\mathbb{R}^2$, and a measure Γ in $\mathcal{M}(\mathcal{C})$. We refer to this measure as the city measure and it characterizes the "size" of every location of the city. For example, if Γ has a point mass at some location then that location is large enough to admit a point mass of supply and demand.

Demand (potential customers) and supply (drivers) are assumed to be infinitesimal and initially distributed on $\mathcal{C}$. We denote the initial demand measure by $\Lambda(\cdot)$ and the supply measure by $\mu(\cdot)$, with both measures belonging to $\mathcal{M}(\mathcal{C})$. For example, if μ is the Lebesgue measure on $\mathcal{C}$, then drivers are uniformly distributed over the city. Both the demand and supply measures are assumed to be absolutely continuous with respect to the city measure, i.e., $\Lambda, \mu \ll \Gamma$. Customers at location $y \in \mathcal{C}$ have their willingness to pay drawn from a distribution $F_y(\cdot)$. For all $y \in \mathcal{C}$, we assume the revenue function $q \mapsto q \cdot \overline{F}_y(q)$ is continuous and unimodular in q and that F_y is strictly increasing over its support $\left[0, \overline{V}\right]$, for some finite positive $\overline{V}$.

We model the interactions between platform, customers and supply as a game. The first player to act in this game is the platform. The platform selects fares across locations and facilitates the matching of drivers and customers. Specifically, the platform chooses a measurable price mapping $p : \mathcal{C} \to [0, \overline{V}]$ so as to maximize its citywide revenues.

After prices are chosen, drivers select *whether* to relocate and *where* to do so. The relocation of drivers generates a flow/transportation of mass from the initial measure of drivers μ to some final endogenous measure of drivers. This final measure corresponds to the supply of drivers in the city after they have traveled to their chosen destination. The movement of drivers across the city is modeled as a measure on $\mathcal{C} \times \mathcal{C}$, which we denote by τ. Any feasible flow has to preserve the initial mass of drivers in $\mathcal{C}$. That is, the first marginal of τ should equal μ. Moreover, τ generates a new (after relocation) distribution of drivers in the city, which corresponds to the second marginal of τ, τ_2. Formally, the set of feasible flows is defined as follows

$$\mathcal{F}(\mu) = \{\tau \in \mathcal{M}(\mathcal{C} \times \mathcal{C}) : \tau_1 = \mu, \quad \tau_2 \ll \Gamma\}.$$

The first condition ensures consistency with the initial positioning of drivers, the second condition ensures that there is no mass of relocated supply at locations where the city itself has measure zero. In particular, given the latter, the Radon-Nikodym derivatives of τ_2 and Λ with respect to Γ, $d\tau_2(y)/d\Gamma$ and $d\Lambda(y)/d\Gamma$, are well defined and for ease of notation we let, for any y in $\mathcal{C}$,

$$s^\tau(y) \triangleq \frac{d\tau_2}{d\Gamma}(y), \text{ and } \lambda(y) \triangleq \frac{d\Lambda}{d\Gamma}(y).$$

Physically, $s^\tau(y)$ represents the *post-relocation supply* at location y normalized by the size of location y, and $\lambda(y)$ corresponds to the potential demand at location y also normalized by the size of such location. Here and in what follows, we will refer to $s^\tau(y)$ and $\lambda(y)$ as the *post-relocation supply* and potential demand at y, respectively. We use the notation $\mathcal{C}_\lambda$ to represent the set of locations with positive potential demand in the city, i.e., $\mathcal{C}_\lambda = \{y \in \mathcal{C} : \lambda(y) > 0\}$.

Given the prices in place, the effective demand at a location y is given by $\lambda(y) \cdot \overline{F}_y(p(y))$, as at location y, only the fraction $\overline{F}_y(p(y))$ is willing to purchase at price $p(y)$. At the same time, the supply at y is given by $s^\tau(y)$. Therefore, the ratio of effective (as opposed to potential) demand to supply at y is given by

$$\frac{\lambda(y) \cdot \overline{F}_y(p(y))}{s^\tau(y)},$$

assuming $s^\tau(y) > 0$. Since a driver can pick up at most one customer within the time frame of our game, a driver relocating to y will face a utilization rate of

$\min\left\{1, \lambda(y) \cdot \overline{F}_y(p(y))/s^\tau(y)\right\}$, assuming $s^\tau(y) > 0$. The effective utilization can be interpreted as the probability that a driver who relocated to y will be matched to a customer within the time frame of our game. In particular, if $s^\tau(y) > \lambda(y) \cdot \overline{F}_y(p(y))$, there is driver congestion at location y, and not all drivers will be matched to a customer. If $s^\tau(y) = 0$ at location y, we say the utilization rate is one if the effective demand at y is positive and zero if the effective demand is zero. Formally, the utilization rate at location y is given by

$$R\Big(y, p(y), s^\tau(y)\Big) \triangleq \begin{cases} \min\left\{1, \frac{\lambda(y)\cdot\overline{F}_y(p(y))}{s^\tau(y)}\right\} & \text{if } s^\tau(y) > 0; \\ 1 & \text{if } s^\tau(y) = 0,\ \lambda(y) \cdot \overline{F}_y(p(y)) > 0; \\ 0 & \text{if } \lambda(y) \cdot \overline{F}_y(p(y)) = 0. \end{cases}$$

When deciding whether to relocate, drivers take three effects into account: prices, travel distance and congestion. The driver congestion effect (or utilization rate) is the one described in the paragraph above. We assume that the platform uses a commission model and transfers a fraction α in $(0, 1)$ of the fare to the driver. As a result, a driver who starts in location y and chooses to remain there earns utility equal to

$$U\Big(y, p(y), s^\tau(y)\Big) \triangleq \alpha \cdot p(y) \cdot R\Big(y, p(y), s^\tau(y)\Big). \tag{1.1}$$

That is, the utility is given by the compensation per ride times the probability of a match. We model the cost for drivers of repositioning from location x to location y through the distance between the locations, $\|y - x\|$. Therefore, a driver originating in x who repositions to y earns utility

$$\Pi\Big(x, y, p(y), s^\tau(y)\Big) \triangleq U\Big(y, p(y), s^\tau(y)\Big) - \|y - x\|. \tag{1.2}$$

When clear from context, and with some abuse of notation, we omit the dependence on price and the supply-demand ratio, writing $U(y)$ and $\Pi(x, y)$. We are now ready to define the notion of a supply equilibrium.

Definition 1.1 (Supply Equilibrium) *A flow* $\tau \in \mathcal{F}(\mu)$ *is an equilibrium if it satisfies*

$$\tau\left(\left\{(x,y) \in \mathcal{C} \times \mathcal{C} : \Pi(x,y,p(y),s^\tau(y)) = \operatorname*{ess\,sup}_{\mathcal{C}} \Pi\Big(x,\cdot,p(\cdot),s^\tau(\cdot)\Big)\right\}\right) = \mu(\mathcal{C}),$$

where the essential supremum is taken with respect to the city measure Γ*.*

That is, an equilibrium flow of supply is a feasible flow such that essentially no driver wishes to unilaterally change his destination. As a result, the mass of drivers selecting the best location for themselves has to equal the original mass of drivers in the system.

The platform's objective is to maximize the revenues it garners across all locations in $\mathcal{C}$. From a given location y, it earns $(1-\alpha) \cdot p(y) \cdot \min\{s^\tau(y), \lambda(y) \cdot \overline{F}_y(p(y))\}$. The term $(1-\alpha)\cdot p(y)$ corresponds to the platform's share of each fare at location y, and the term $\min\{s^\tau(y), \lambda(y) \cdot \overline{F}_y(p(y))\}$ denotes the quantity of matches of potential customers to drivers at location y. If location y is demand constrained, then $\min\{s^\tau(y), \lambda(y) \cdot \overline{F}_y(p(y))\}$ equals $\lambda(y) \cdot \overline{F}_y(p(y))$, while if location y is supply constrained, then $\min\{s^\tau(y), \lambda(y) \cdot \overline{F}_y(p(y))\}$ amounts to $s^\tau(y)$. The platform's price optimization problem can in turn be written as

$$\begin{aligned} \sup_{p(\cdot),\, \tau \in \mathcal{F}(\mu)} \quad & (1-\alpha) \int_{\mathcal{C}} p(y) \cdot \min\{s^\tau(y), \lambda(y) \cdot \overline{F}_y(p(y))\}\, d\Gamma(y) \qquad (\mathcal{P}_1) \\ s.t. \quad & \tau \text{ is a supply equilibrium,} \\ & s^\tau = \frac{d\tau_2}{d\Gamma}. \end{aligned}$$

Remark. Our model may be interpreted as a basic model to understand the short-term operations of a ride-hailing company. In particular, each driver completes at most one customer pickup within the time frame of our game and there is not enough time for the entry of new drivers into the system. In the present model, we do not account explicitly for the destinations of the rides. We do so in order to isolate the interplay of supply incentives and pricing. In that regard, one could view our model as capturing origin-based pricing, a common practice in the ride-hailing industry.

1.4 Structural Properties and Spatial Decomposition

A key challenge in solving the optimization problem presented in ($\mathcal{P}_1$) is that the decision variables, the flow τ and the price function $p(\cdot)$, are complicated objects. The flow τ, being a measure over a two-dimensional space, is obviously a complex object to manipulate. The price function will turn out to be a difficult object to manipulate as well in that the optimal price function will often be discontinuous. In order to analyze our problem, we will need to introduce a better-behaved object. This object, which will be central to our analysis, is the (after movement) driver equilibrium utility.

Drivers' utilities. For a given price function p and flow τ, we denote by $V_{\mathcal{B}}(x|\, p,\tau)$ the essential maximum utility that a driver departing from location x can garner by going anywhere within a measurable region $\mathcal{B} \subseteq \mathcal{C}$. In particular, the mapping $V_{\mathcal{B}}(\cdot|\, p,\tau) : \mathcal{C} \to \mathbb{R}$ is defined as

$$V_{\mathcal{B}}(x|\, p,\tau) \triangleq \operatorname*{ess\,sup}_{\mathcal{B}} \Pi\left(x, \cdot, p(\cdot), s^{\tau}(\cdot)\right). \tag{1.3}$$

When $\mathcal{B} = \mathcal{C}$, we use V instead of $V_{\mathcal{C}}$. By the definition of a supply equilibrium, essentially all drivers departing from location x earn $V(x|\, p,\tau)$ utility in equilibrium.

We now show that the equilibrium utility $V_{\mathcal{B}}(\cdot|\, p,\tau)$ must be 1-Lipschitz continuous. Intuitively, drivers from two different locations x and y that consider relocating to $\mathcal{B}$ see exactly the same potential destinations. Hence, the largest utility drivers departing from x can garner must be greater or equal to that of the drivers departing from y minus the disutility stemming from relocating from x to y, that is,

$V_{\mathcal{B}}(x) \geq V_{\mathcal{B}}(y) - \|x - y\|$. Since this argument is symmetric, we deduce the 1-Lipschitz property.

Lemma 1.1 *(Lipschitz) Consider a measurable set $\mathcal{B} \subseteq \mathcal{C}$ such that $\Gamma(\mathcal{B}) > 0$. Let p be a measurable mapping $p : \mathcal{B} \to \mathbb{R}_+$, and let $\tau \in \mathcal{F}(\mu)$. Then, the function $V_{\mathcal{B}}(\cdot|p, \tau)$ is 1-Lipschitz continuous.*

We now introduce a reformulation of $(\mathcal{P}_1)$ that focuses on the equilibrium utility V and the post-relocation supply s^τ as the central elements. We then establish important structural properties of V and establish a spatial decomposition result that is based on the equilibrium behavior of drivers.

1.4.1 Reformulating the Platform's problem

In what follows, we define $\gamma \triangleq (1 - \alpha)/\alpha$. In the next result, we establish that the platform's objective can be rewritten in terms of the utility function $V(\cdot|\, p, \tau)$ and the post-relocation supply s^τ, yielding an alternative optimization problem.

Proposition 1.1 (Problem Reformulation) *The following problem*

$$\sup_{p(\cdot)\,,\,\tau \in \mathcal{F}(\mu)} \quad \gamma \cdot \int_{\mathcal{C}_\lambda} V(x|\, p, \tau) \cdot s^\tau(x)\, d\Gamma(x) \tag{$\mathcal{P}_2$}$$

$$\text{s.t.} \qquad \tau \text{ is an equilibrium flow,}$$

$$V(x|\, p, \tau) = \operatorname*{ess\,sup}_{\mathcal{C}} \Pi\Big(x, \cdot, p(\cdot), s^\tau(\cdot)\Big), \quad s^\tau = \frac{d\tau_2}{d\Gamma},$$

admits the same value as the platform's optimization problem $(\mathcal{P}_1)$, and a pair (p, τ) that solves $(\mathcal{P}_2)$ also solves $(\mathcal{P}_1)$.

The first step in the proof of the proposition above is to rewrite the platform's objective in terms of the post-relocation supply $s^\tau(x)$ and the pre-movement utility function $U(x, p(x), s^\tau(x))$ (see Eq. (1.1)). This transformation is not particularly useful per se, since the function $U(x, p(x), s^\tau(x))$ is not necessarily well-behaved.

The next step consists of establishing that $U\left(x, p(x), s^{\tau}(x)\right)$ coincides with $V(x|\, p, \tau)$ whenever a location has positive post-movement equilibrium supply (see Lemma A.2 in the Appendix). Indeed, whenever the equilibrium outcome is such that a location has positive supply, the utility generated by staying at that location has to be equal to the best utility one could obtain by traveling to any other location. This is intuitive in that if it were not the case, no driver would be willing to stay at or travel to that location. In turn, one can effectively replace $U\left(x, p(x), s^{\tau}(x)\right)$ with $V(x|\, p, \tau)$ in the objective, which yields the alternative problem. The main advantage of this new formulation is that the equilibrium utility $V(x|\, p, \tau)$ connects our problem to the theory of optimal problem and it admits significant structure, as we show in the next two subsections.

1.4.2 Connection to Optimal Transport

Our equilibrium concept is closely related to the notion of optimal transport plan in the theory of optimal transport. In any equilibrium τ the total mass of drivers repositions in the most efficient way as to minimize the total transportation cost.

Let τ be an equilibrium flow with second marginal τ_2 then

$$
\begin{aligned}
\tau \in \arg & \min_{\gamma \in \mathcal{M}(\mathcal{C}\times\mathcal{C})} \int_{\mathcal{C}\times\mathcal{C}} \|x-y\| d\gamma(x,y) \\
\text{s.t} \quad & \gamma_1 = \mu, \quad \gamma_2 = \tau_2
\end{aligned}
$$

Indeed, let γ be a feasible transport plan and let us use $\mathcal{W}(\gamma)$ to denote the optimal transport objective under the plan γ then

$$
\begin{aligned}
\mathcal{W}(\gamma) &= -\int_{\mathcal{C}\times\mathcal{C}} \Big(U\Big(y, p(y), s^{\tau}(y)\Big) - \|y-x\|\Big) d\gamma + \int_{\mathcal{C}\times\mathcal{C}} U\Big(y, p(y), s^{\tau}(y)\Big) d\gamma \\
&= -\int_{\mathcal{C}\times\mathcal{C}} \Big(U\Big(y, p(y), s^{\tau}(y)\Big) - \|y-x\|\Big) d\gamma + \int_{\mathcal{C}} U\Big(y, p(y), s^{\tau}(y)\Big) d\tau_2(y) \\
&\geq -\int_{\mathcal{C}} V(x|\, p, \tau) d\mu(x) + \int_{\mathcal{C}} U\Big(y, p(y), s^{\tau}(y)\Big) d\tau_2(y) \\
&= -\int_{\mathcal{C}\times\mathcal{C}} V(x|\, p, \tau) d\tau + \int_{\mathcal{C}} U\Big(y, p(y), s^{\tau}(y)\Big) d\tau_2(y) \\
&= -\int_{\mathcal{C}\times\mathcal{C}} \Big(U\Big(y, p(y), s^{\tau}(y)\Big) - \|y-x\|\Big) d\tau + \int_{\mathcal{C}} U\Big(y, p(y), s^{\tau}(y)\Big) d\tau_2(y) \\
&= \mathcal{W}(\tau)
\end{aligned}
$$

This establishes that given the final supply of drivers τ_2 then an equilibrium flow with second marginal τ_2 minimizes the total transportation cost. In our problem, τ_2 is an endogenous object that we need to find via optimization.

1.4.3 Indifference and Attraction Regions

A key feature of the problem at hand is that, in equilibrium, conditions at different locations are inherently linked as drivers select their destination among all locations. An important object that will help capture the link across various locations is the *indifference region* of a driver departing location x. The indifference region of x represents all the destinations to which drivers from x are willing to travel to. Formally, the indifference region for a driver departing from $x \in \mathcal{C}$ under prices p and flow τ is given by

$$
\mathcal{IR}(x|\, p, \tau) \triangleq \Big\{ y \in \mathcal{C} : \lim_{\delta \downarrow 0} V_{B(y,\delta)}(x|p, \tau) = V(x|p, \tau) \Big\},
$$

where $B(y, \delta)$ is the open ball in $\mathcal{C}$ of center y and radius δ. Intuitively, the definition above says that if $y \in \mathcal{IR}(x|\, p, \tau)$, then drivers departing from x maximize their utility by relocating to y.

Indifference regions describe the set of best possible destination for a given location. The converse concept which will turn out to be fundamental in our analysis is the *attraction region* of a location z. The attraction region of z represents the set of all possible sources for which location z is their best option. In addition, location z is called a *sink* if it is not willing to travel to any other location. These regions are rich in the sense that they enjoy several appealing properties and, as we will see in Section 1.5, we can solve for the platform's optimal solution within them. Below we provide a formal definition for an attraction region and a sink location.

In line with the literature on optimal transport, see e.g [5], it will be useful in our analysis to study the behavior of drivers along rays around a particular location z. We use R_z to denote the set of all rays originating from z (excluding z) and index the elements of R_z by a. The advantage of this is that now we can disintegrate the city measure into a family of measures concentrated along the rays, $\{\Gamma_a\}$, which we can integrate with respect to another measure Γ^{p} in R_z to obtain Γ, that is,

$$\Gamma(\mathcal{B}) = \Gamma(\{z\})\mathbf{1}_{\{z\in\mathcal{B}\}} + \int_{R_z} \Gamma_a(\mathcal{B})d\Gamma^{\mathrm{p}}(a). \tag{1.4}$$

In what follows we will use interchangeable Γ and Eq. (1.4).

Definition 1.2 (Attraction Region) *Let (p,τ) be a feasible solution of ($\mathcal{P}_2$). For any location $z \in \mathcal{C}$, its* attraction region *$A(z|\, p,\tau)$ is the set of locations from which drivers are willing to relocate to z, i.e.,*

$$A(z|\, p,\tau) \triangleq \{x \in \mathcal{C} : z \in \mathcal{IR}(x|\, p,\tau)\}.$$

We call a location $z \in \mathcal{C}$ a sink *if its attraction region $A(z|\, p,\tau)$ is non-empty and $z \notin A(z'|\, p,\tau)$ for all $z' \neq z$. When z is a sink, we represent the endpoints of its attraction region along a ray $a \in R_z$ by*

$$X_a(z|\, p,\tau) \triangleq \sup\{x \in A_a(z|\, p,\tau)\},$$

where $A_a(z|\, p,\tau)$ is the restriction of $A(z|\, p,\tau)$ in the direction of ray a.

Definition 1.3 (In-demand location) *We say a location z is in-demand whenever $\forall Q \subset R_z$ such that $\Gamma^{\mathfrak{p}}(Q) > 0$*

$$\Gamma(\{z\})\mathbf{1}_{\{\lambda(z)>0\}} + \int_Q \int_{(z,z+\delta]} \mathbf{1}_{\{\lambda(x)>0\}} d\Gamma_a(x) d\Gamma^{\mathfrak{p}}(a) > 0, \quad \forall \delta > 0.$$

The next result characterizes the shape of attraction regions.

Lemma 1.2 (Attraction Region) *Let (p, τ) be a feasible solution of ($\mathcal{P}_2$). For any sink $z \in \mathcal{C}$, its attraction region $A(z|\, p, \tau)$ is a closed set containing z, $A_a(z|\, p, \tau) = [z, X_a(z|\, p, \tau)]$ and*

$$A(z|\, p, \tau) = \bigcup_{a \in R_z} A_a(z|\, p, \tau).$$

The lemma above establishes an intuitive but important transitivity result. Let $x < y < z$ be such that x is in the attraction region of z. Then, y must also be in the attraction region of z.

The structure of the utility function V at a supply equilibrium will play a central role in our analysis. The following lemma establishes the shape of V within attraction regions.

Lemma 1.3 (Utility Within an Attraction Region) *Let (p, τ) be a feasible solution of ($\mathcal{P}_2$), then for any $z \in \mathcal{C}$ the equilibrium utility satisfies*

$$V(x|\, p, \tau) = V(z|\, p, \tau) - \|z - x\|, \quad \textit{for all} \quad x \in A(z|\, p, \tau).$$

This result is closely related to the Envelope Theorem, which is widely used in mechanism design (see [49]). If a driver originating from x is indifferent to relocating to z, then $V(z|\, p, \tau) - V(x|\, p, \tau)$ must be equal to the relocation cost $\|z - x\|$.

Importantly, attraction regions emerge as soon as drivers move in the city, as formalized in the next proposition.

Proposition 1.2 (Existence of attraction regions) *Let (p, τ) be a feasible solution of ($\mathcal{P}_2$) and suppose that $y \in \mathcal{IR}(x|\, p, \tau)$ for some $x \neq y$. Then, there exists a sink location $z \in \mathcal{C}$ such that $x, y \in A(z|\, p, \tau)$ and x, y, z are collinear points.*

In other words, as soon as there is potential for movement, in the sense that drivers at some location weakly prefer to travel to another location, necessarily an attraction region exists.

1.4.4 Spatial Decomposition

Next, we show that attraction regions lead to a natural decoupling of the platform's problem, as they provide a natural way of segmenting the city. The next result establishes a flow separation property induced by attraction regions.

Proposition 1.3 (Flow Separation) *Let (p, τ) be a feasible solution of $(\mathcal{P}_2)$, and let $z \in \mathcal{C}$ be a sink. Then, there is no flow crossing the endpoints of the attraction region, and there is no flow crossing the sink, z. Formally, with some abuse of notation, let $L(z|\, p, \tau)$ denote $\bigcup_{a \in R_z} \{X_a(z|\, p, \tau)\}$ then*

(i) $\tau(A(z|\, p, \tau)^c \times A(z|\, p, \tau)) = 0$ *and*

$$\tau(\textstyle\bigcup_{a \in R_z} [z, X_a(z|\, p, \tau) \times \Big(A(z|\, p, \tau)^c \cup L(z|\, p, \tau) \setminus \{z\}\Big)) = 0.$$

(ii) Let $R_1, R_2 \subset R_z$ with $R_1 \cap R_2 = \emptyset$ then

$$\tau\Big(\bigcup_{a \in R_1} (z, X_a(z|\, p, \tau)] \times \bigcup_{a \in R_2} (z, X_a(z|\, p, \tau)]\Big) = 0.$$

The first part of this result characterizes attraction regions as flow-isolated sets. There is no flow of drivers traveling to an attraction region from outside of it. And drivers in the interior of an attraction region do not travel outside the region.[1] In this sense, attraction regions are flow-separated subsets of $\mathcal{C}$. This will enable us to "decouple" the platform's problem in an attraction region from the rest of the city in Section 1.5.2. The second part of the proposition establishes that in an attraction region, no flow crosses between rays. However, there could be flow stemming from

[1] We clarify here that Proposition 1.3 does not impose anything on the direction of flow emerging from the end points $X_a(z|\, p, \tau)$ for $a \in R_z$. That is, if there is a mass of drivers starting from one of these boundary points, these drivers could move either into or out of the attraction region.

any ray that travels to the sink. That is, the segments $\{(z, X_a(z|\, p, \tau]\}_{a\in R_z}$ of the attraction region are flow-separated regions coupled by the sink location. Figure 1.2 illustrates this proposition.

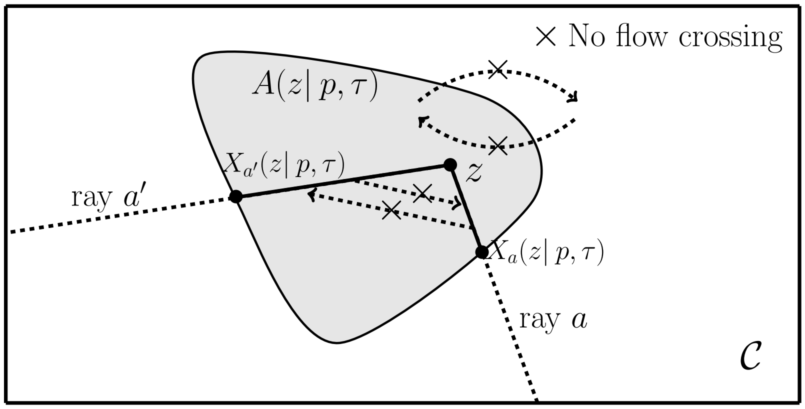


Figure 1.2: **Flow separation.** Illustration of the result in Proposition 1.3. No flow crosses the boundaries of $A(z|\, p, \tau)$.

This flow separation result will enable us to geographically decompose the platform's problem into multiple weakly coupled local problems. To that end, we introduce some additional notation that will allow us to "localize the analysis". Formally, for any measurable $\mathcal{B} \subset \mathcal{C}$ and measure $\tilde{\mu} \in \mathcal{M}(\mathcal{B})$, we define the set of feasible flows restricted to $\mathcal{B}$ to be

$$\mathcal{F}_{\mathcal{B}}(\tilde{\mu}) = \{\tau \in \mathcal{M}(\mathcal{B} \times \mathcal{B}) : \tau_1 = \tilde{\mu}, \quad \tau_2 \ll \Gamma|_{\mathcal{B}}\}.$$

In addition, we define local equilibria as follows.

Definition 1.4 (Local Equilibrium) *For any $\mathcal{B} \subset \mathcal{C}$ such that $\Gamma(\mathcal{B}) > 0$ and $\tilde{\mu} \in \mathcal{M}(\mathcal{B})$, a flow $\tau \in \mathcal{F}_{\mathcal{B}}(\tilde{\mu})$ is a local equilibrium in $\mathcal{B}$ if it satisfies*

$$\tau\left(\left\{(x,y) \in \mathcal{B} \times \mathcal{B} : \Pi(x, y, p(y), s^{\tau}_{|\mathcal{B}}(y)) = \operatorname*{ess\,sup}_{\mathcal{B}} \Pi\Big(x, \cdot, p(\cdot), s^{\tau}_{|\mathcal{B}}(\cdot)\Big)\right\}\right) = \tilde{\mu}(\mathcal{B}).$$

That is, a local equilibrium in $\mathcal{B}$ is a feasible flow such that no driver wishes to unilaterally change his destination when restricting attention to the set $\mathcal{B}$. With this definition in hand, we may now state our next result. Informally, this result states the following "pasting" property. Suppose we start from a price-equilibrium pair (p, τ) and a sink z and its attraction region $A(z|\, p, \tau)$. Then, we can replace the flow that occurs within $A(z|p, \tau)$ with any other local equilibrium within that attraction region as long as we maintain the same conditions at the boundary $\partial A(z|\, p, \tau)$.

Proposition 1.4 *(Pasting) Let (p, τ) be a feasible solution of ($\mathcal{P}_2$), and let $z \in \mathcal{C}$ be a sink. Denote $\mathcal{A} = A(z|\, p, \tau)$ and $\mathcal{L} = \bigcup_{a \in R_z} \{X_a(z|\, p, \tau)\}$. Let $\tilde{\mu} \in \mathcal{M}(\mathcal{A})$ be the measure representing drivers that stay within $\mathcal{A}$ according to flow τ, i.e., $\tilde{\mu}(\mathcal{B}) \triangleq \tau(\mathcal{B} \times \mathcal{A})$ for any measurable set $\mathcal{B} \subseteq \mathcal{A}$. Suppose there exists a measurable price mapping $\tilde{p} : \mathcal{A} \to [0, \overline{V}]$ and a flow $\tilde{\tau} \in \mathcal{F}_{\mathcal{A}}(\tilde{\mu})$ such that $\tilde{\tau}$ is a local equilibrium in $\mathcal{A}$ under pricing $\tilde{p}$. Furthermore, suppose $V_{\mathcal{A}}(\cdot|\, \tilde{p}, \tilde{\tau})$ equals $V(\cdot|\, p, \tau)$ in $\partial\mathcal{A}$. Define the pasted pricing function $\hat{p} : \mathcal{C} \to [0, \overline{V}]$,*

$$\hat{p}(x) \triangleq \begin{cases} \tilde{p}(x) & \text{if } x \in \mathcal{A}; \\ p(x) & \text{if } x \in \mathcal{A}^c, \end{cases}$$

and the pasted flow $\hat{\tau} \in \mathcal{F}(\mu)$, where for any measurable $\mathcal{B} \subseteq \mathcal{C} \times \mathcal{C}$

$$\hat{\tau}(\mathcal{B}) \triangleq \tau(\mathcal{B} \cap ((\mathcal{A}^c \cup \mathcal{L}) \times \mathcal{A}^c)) + \tilde{\tau}(\mathcal{B} \cap (\mathcal{A} \times \mathcal{A})).$$

Then, the pasted solution $(\hat{p}, \hat{\tau})$ is a feasible solution of problem ($\mathcal{P}_2$) such that

$$s^{\hat{\tau}} = \begin{cases} s^{\tilde{\tau}}(x) & \text{if } x \in \mathcal{A}; \\ s^{\tau}(x) & \text{if } x \in \mathcal{A}^c, \end{cases} \quad \text{and} \quad V(x|\, \hat{p}, \hat{\tau}) = \begin{cases} V_{\mathcal{A}}(x|\, \tilde{p}, \tilde{\tau}) & \text{if } x \in \mathcal{A}; \\ V(x|\, p, \tau) & \text{if } x \in \mathcal{A}^c. \end{cases}$$

Propositions 1.3 and 1.4 suggest a natural structure for the induced flows by any pricing policy. For a given sink z, Proposition 1.3 establishes that the attraction region of z and its complement are flow separated. Now Proposition 1.4 applies this flow separation result and shows how to make local deviations to a feasible solution while maintaining feasibility. More precisely, an equilibrium in $\mathcal{C}$ can be locally modified in the attraction region of z, without losing feasibility, as long the equilibrium utilities of drivers in the boundaries of the attraction region are not modified. The new solution $(\hat{p}, \hat{\tau})$ in $\mathcal{C}$ merges the old solution (p, τ) in $A(z|p,\tau)^c$ with the modified solution $(\tilde{p}, \tilde{\tau})$ in the attraction region $A(z|p,\tau)$.

1.5 Congestion Bound and Optimal Flows

In the prior section, we showed that the platform's optimization problem can be reformulated as a problem over equilibrium utilities V and post-relocation supply s^τ. We also showed that V is a well-behaved function: it is 1-Lipschitz continuous and it has derivative equal to +1 or -1 over attraction regions. Furthermore, we demonstrated how to use attraction regions to decompose the platform's global problem into localized problems. In this section, we focus on the optimal relocation of drivers within attraction regions. That is, we will prove that, without loss of optimality, we can restrict attention to flows within attraction regions that take a very specific form. In order to do so, we first need to formalize the notion of congestion level of a given location.

1.5.1 Congestion Bound

We first introduce some quantities that will be useful throughout our analysis. These quantities emerge from a classical capacitated monopoly pricing problem. Let us consider any location $x \in \mathcal{C}$ and ignore all other locations in the city. The problem that a monopolist faces when supply at x is s and demand is λ_x can be cast as

$$R_x^{loc}(s) \triangleq \max_{q \in [0,\overline{V}]} \; q \cdot \min\{s, \lambda_x \cdot \overline{F}_x(q)\}, \tag{1.5}$$

with the price $\rho_x^{loc}(s)$ being defined as the argument that maximizes the equation above. Since $q \cdot \overline{F}_x(q)$ is assumed to be unimodular in q, the optimal price $\rho_x^{loc}(s)$ is uniquely determined and is characterized as follows

$$\rho_x^{loc}(s) = \max\{\rho_x^{bal}(s), \rho_x^u\}, \qquad \text{where } s = \lambda_x \cdot \overline{F}_x(\rho_x^{bal}(s)), \quad \rho_x^u \in \arg\max_{\rho \in [0,\overline{V}]} \{\rho \cdot \overline{F}_x(\rho)\}. \tag{1.6}$$

That is, the optimal local price either balances supply and demand or maximizes the unconstrained local revenue.

For a given local supply s, the maximum revenue that can be generated at location x is $R_x^{loc}(s)$, with a fraction α of that revenue being paid to the drivers. Therefore, $\alpha \cdot R_x^{loc}(s)/s$ is the maximum revenue a driver staying at this location can earn. To capture this notion, we introduce for every location x the supply *congestion* function $\psi_x : \mathbb{R}_+ \to [0, \alpha \cdot \overline{V}]$, which is defined as:

$$\psi_x(s) \triangleq \begin{cases} \alpha \cdot R_x^{loc}(s)/s & \text{if } s > 0; \\ \alpha \cdot \overline{V} & \text{if } s = 0,\ \lambda(x) > 0; \\ 0 & \text{if } s = 0,\ \lambda(x) = 0. \end{cases}$$

The congestion function ψ_x must be decreasing since more drivers (in a single location problem) imply lower revenues per driver.

Lemma 1.4 *For any $x \in \mathcal{C}_\lambda$ the congestion function $\psi_x(\cdot)$ is a strictly decreasing function.*

More importantly, the congestion function ψ_x yields an upper bound for the utility of drivers at almost any location with respect to the city measure.

Proposition 1.5 (Congestion Bound) *Let (p, τ) be a feasible solution of $(\mathcal{P}_2)$. Then the equilibrium driver utility function is bounded as follows:*

$$V(x|\, p, \tau) \;\leq\; \psi_x\left(s^\tau(x)\right) \qquad \Gamma - a.e.\ x \ \ in\ \mathcal{C}_\lambda.$$

When there is a single location, the inequality above is an equality by the definition of ψ_x. For multiple locations, drivers may travel to any location and there is no a priori connection between the utility that drivers originating from x can garner, $V(x|\, p, \tau)$, and $\psi_x(s^\tau(x))$. The result above establishes that the latter upper bounds the former. The bound captures the structural property that as equilibrium supply increases at a location, and hence driver congestion increases, the drivers originating from that location will earn less utility.

1.5.2 Optimal Supply Reallocation in Attraction Regions

We now consider the problem of how to optimize flows within an attraction region. The key idea is to use the structural properties about the equilibrium utility function as well as the pasting result developed in Section 1.4, in conjunction with a relaxation to the platform's problem within an attraction region that leverages the congestion bound established in Proposition 1.5.

Consider a feasible solution (p,τ) of $(\mathcal{P}_2)$. Let $z \in \mathcal{C}$ be a sink and $A(z|\, p,\tau)$ its corresponding attraction region. We will now show how to construct a second feasible solution of $(\mathcal{P}_2)$ for which the revenue is weakly larger and we can fully characterize its prices and flows within the attraction region $A(z|\, p,\tau)$ as defined by the original solution (p,τ).

Theorem 1.1 *(Optimal Supply Within an Attraction Region) Consider a feasible solution (p,τ) of $(\mathcal{P}_2)$, and let $z \in \mathcal{C}$ be an in-demand sink. Then, there exists another feasible solution $(\hat{p},\hat{\tau})$ that weakly revenue dominates (p,τ), and is such that $V(\cdot|\, \hat{p},\hat{\tau})$ coincides with $V(\cdot|\, p,\tau)$ in $A(z|\, p,\tau)$ and its supply $s^{\hat{\tau}}$ in $A(z|\, p,\tau)$ is given by:*

$$s^{\hat{\tau}}(x) = \begin{cases} \psi_x^{-1}(V(z|\, p,\tau) - \|x-z\|)\cdot \mathbf{1}_{\{\lambda(x)>0\}} & \text{if } x \in \bigcup_{a\in R_z}[z,r_a); \\ s_i & \text{if } x = r_a,\ a \in R_z; \\ 0 & \text{otherwise,} \end{cases}$$

for a set of values $\{r_a\}$ such that $r_a \in [z, X_a(z|p,\tau)]$ and $s_a \geq 0, a \in R_z$. Furthermore,

$$\hat{p}(x) = \begin{cases} \rho_x^{loc}(s^{\hat{\tau}}(x)) & \text{if } x \in A(z|\, p,\tau) \setminus \bigcup_{a\in R_z}\{r_a\}; \\ p_i & \text{if } x = r_a,\ a \in R_z, \end{cases}$$

where p_a is such that $U(r_a,p_a,s_a) = V(r_a|\, p,\tau)\cdot \mathbf{1}_{\{\lambda(r_a)>0\}}$ for $a \in R_z$.

The theorem above characterizes an optimal solution, including both prices and flows, within an attraction region. In particular, the optimality of a pricing policy implies that it is sufficient to focus on solutions that have post-movement equilibrium supply around the sink z in $\bigcup_{a \in R_z} [z, r_a]$ while potentially creating regions with zero equilibrium supply away from the sink, in the segments $\{(r_a, X_a]\}_{a \in R_z}$. These regions "feed" the region around the sink z with drivers. Furthermore, the optimal prices are fully characterized in any attraction region through the post-relocation supply. We will highlight the main implications of Theorem 1.1 through a prototypical family of instances in Section 1.6, where we will characterize the optimal solution across the city in quasi-closed form.

Key ideas for Theorem 1.1. The key idea underlying the proof of the result is based on optimizing the contribution of the attraction region $A(z|\, p, \tau)$ to the overall objective by reallocating the supply around the sink, and then showing that this reallocation of supply constitutes an equilibrium flow in the original problem.

In order to optimize the supply around the sink we consider the following optimization problem which, as explained below, is a relaxation of $(\mathcal{P}_2)$ within $A(z|\, p, \tau)$:

$$\max_{\tilde{s}(\cdot) \geq 0} \int_{A(z|\, p,\tau)} V(x|\, p, \tau) \cdot \tilde{s}(x)\, d\Gamma(x) \qquad (\mathcal{P}_{KP}(z))$$

$$\text{s.t} \quad \tilde{s}(x) \leq \psi_x^{-1}(V(x)) \qquad \Gamma - a.e.\ x \text{ in } \mathcal{C}_\lambda, \qquad \text{(Congestion Bound)}$$

$$\int_{A(z|\, p,\tau)} \tilde{s}(x) d\Gamma(x) = \tau_c, \qquad \text{(Flow Conservation)}$$

$$\int_{(z, X_a]} \tilde{s}(x) d\Gamma_a(x) \leq \tau_a, \qquad \Gamma^{\mathrm{p}} - a.e.\ a \in R_z. \qquad \text{(No Flow Crossing Rays)}$$

where τ_c corresponds to the total flow that τ transports from $A(z|\, p, \tau)$ to $A(z|\, p, \tau)$, and τ_a correspond to the total flow in $A(z|p, \tau)$ that is transported to ray a, excluding z. Recall that given the post-relocation supply, $\tilde{s}$, the quantity

$$\int_{\mathcal{B}} \tilde{s}(x)\, d\Gamma(x),$$

represents the post-relocation supply induced by $\tilde{s}$ in $\mathcal{B}$. Thus, the last three constraints in ($\mathcal{P}_{KP}(z)$) stand for consistency of the total post-relocation supply in each one the relevant subregions of $A(z|\, p, \tau)$. The key is to observe that this is a relaxation of the original problem in the attraction region. In particular, the equilibrium constraint implies the conservation constraint (see Proposition 1.3(i)), and the no-flow-crossing constraints (see Proposition 1.3(ii)). The congestion bound is also a consequence of the equilibrium constraint (see Proposition 1.5). In words, in this formulation, we relax the equilibrium constraint but impose implications of it. We constrain the amount of mass that we can allocate on each direction around z but we fix the total amount of mass in $A(z|\, p, \tau)$.

In ($\mathcal{P}_{KP}(z)$), we fix the driver utilities and ask what should be the optimal allocation of drivers while satisfying flow balance in the regions $\{[z, X_a]\}_{z \in R_z}$ and imposing the congestion bound. Clearly selecting $\tilde{s} = s^\tau$ is feasible for the problem above and hence the optimal value upper bounds the value generated by the initial price-equilibrium pair (p, τ) in the region $A(z|\, p, \tau)$. In the proof, we show that this relaxation is tight. Namely, it is possible to construct prices and equilibrium flows achieving the value of Problem ($\mathcal{P}_{KP}(z)$). The proof consists of two main steps: 1) solving problem ($\mathcal{P}_{KP}(z)$) and 2) showing that the post-relocation supply that solves the relaxation can actually be obtained from appropriate prices and flows. For step 1), the main idea relies on recognizing that Problem ($\mathcal{P}_{KP}(z)$) is a measure-theoretical instance of a coupled collection of *Continuous Bounded Knapsack Problems.* In particular, the congestion constraint corresponds to the availability constraint in the classical knapsack problem. The solution to ($\mathcal{P}_{KP}(z)$) is obtained by allocating as much as possible at locations where we can make the most revenue per unit of volume, i.e., we would like to make $\tilde{s}(x)$ as large as possible at locations where $V(x|\, p, \tau)$ is the largest. Hence the solution starts by allocating as much supply as possible at location z. The challenge here is that flow-crossing conditions need also to be satisfied and hence whether flow is sent to z from one ray or another is key and needs to be tracked. For step 2), we explicitly construct prices, and the flow correspond to the integration of the solution of a collection of optimal transport problems. Along each segment $(z, X_a]$ we solve an optimal transport problem with cost function equal to the distance between any two points, initial measure equal to the reminder mass that was not sent to z, and final measure equal to the restriction of the solution of Problem $(\mathcal{P}_{KP}(z))$ in $(z, X_a]$. Finally, we apply the pasting result (Proposition 1.4) to obtain a feasible price-equilibrium in the whole city $\mathcal{C}$.

1.6 Response to Demand Shock: Optimal Solution and Insights

The results derived in the previous sections characterize the structure of an optimal pricing policy and the corresponding supply response in attraction regions for general demand and supply conditions in a two dimensional region. In this section, to crisply isolate the interplay of spatial supply incentives and spatial pricing, we focus on a special family of instances that will be rich enough to capture spatial supply-demand imbalances while isolating the interplay above.

In particular, to simplify exposition we focus on a one dimensional city and a family of models that captures a potential local surge in demand. Namely, we specialize the model to the case where the city measure is supported on the interval $[-H, H]$ and is given by

$$\Gamma(\mathcal{B}) = \mathbf{1}_{\{0 \in \mathcal{B}\}} + \int_{\mathcal{B}} dx, \quad \text{for any measurable set } \mathcal{B} \subseteq [-H, H]^2$$

that is, the origin may admit point masses of supply and demand while the rest of the locations in $[-H, H]$ only admit infinitesimal amounts of supply and demand. In what follows, without loss of generality we will use $\mathcal{C}$ to denote $[-H, H]$, that is, the

[2] Observe that thanks to the generality of our measure theoretical framework, all the theoretical results develop thus far apply to this one dimensional setting.

city now corresponds to the one dimensional interval over which the city measure is supported. We fix the city measure throughout, but we parametrize the supply and demand measures.

Supply is initially evenly distributed throughout the city, with a density of drivers equal to μ_1 everywhere. Potential demand will be also assumed to have a uniform density on the line interval, except potentially at the origin.

We analyze what happens when a potential demand shock at the origin (the potential high demand location) materializes and, in particular, we investigate the optimal pricing policy in response to such a shock. We represent the demand shock by a Dirac delta at this location. Therefore, for any measurable set $B \subseteq \mathcal{C}$, the potential demand measure (after the shock) is given by

$$\Lambda(\mathcal{B}) = \lambda_0 \cdot \mathbf{1}_{\{0 \in \mathcal{B}\}} + \int_{\mathcal{B}} \lambda_1 dx,$$

where $\lambda_0 \geq 0$ and $\lambda_1 > 0$. In particular, we refer to the case $\lambda_0 = 0$ as the *pre-demand shock environment* and the case $\lambda_0 > 0$ as the *demand shock environment.*

For this family of models, we assume that customer willingness to pay is drawn from the same distribution $F(\cdot)$ for all locations in the city (and this function is assumed to satisfy the regularity conditions of Section 3.3). Figure 1.3 provides a visual representation of this family of cases.

This special structure will enable us to elucidate the spatial supply response induced by surge pricing and the structural insights on the optimal policies that emerge.

Throughout this section we will use short-hand notation to present the optimal solution in a streamlined fashion. Let (p, τ) be a price equilibrium pair we use $A(0), X_l$ and X_r to denote $A(0|\, p, \tau)$, and the end points of the left and right rays around z, respectively. Moreover, when clear from context, we write $V(\cdot)$ instead of $V(\cdot|\, p, \tau)$.

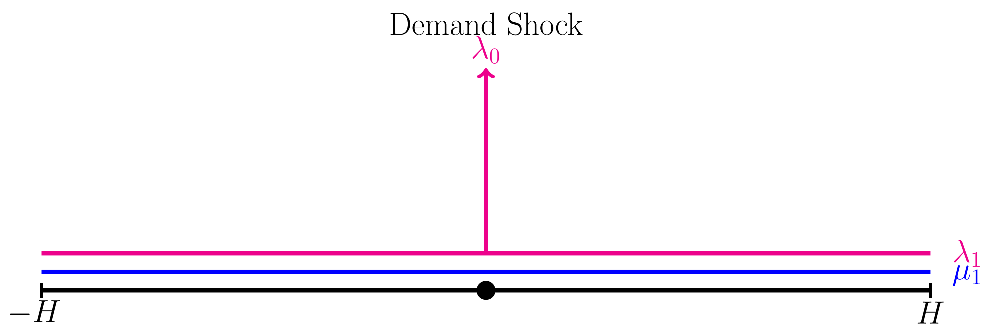


Figure 1.3: **Prototypical family of models with demand surge.** The supply is initially uniformly distributed in the city with density μ_1, and potential demand is uniformly distributed in the city with density λ_1, with a sudden demand surge at location 0.

1.6.1 The Pre-demand Shock Environment

We start by analyzing the pre-shock environment. In this environment, there is no demand shock, $\lambda_0 = 0$, and both demand and supply are uniformly distributed along the city, with respective densities λ_1 and μ_1. If one were to look at each location in isolation, the optimal local price at a location x with demand density λ_1 and supply density μ_1 is $\rho_x^{loc}(\mu_1)$, as defined in Eq. (1.6). Note that in the current environment $\rho_x^{loc}(\mu_1)$ is not location dependent and we denote it by ρ_1 throughout, we do the same with $\psi_x(\mu_1)$ which we denote by ψ_1.

Proposition 1.6 (Pre-demand Shock Environment) *Suppose $\lambda_0 = 0$. Then, the optimal policy and corresponding supply equilibrium and flows can be characterized as follows.*

(i) (Prices) The optimal pricing policy is given by $p(x) = \rho_1$, for all x in $\mathcal{C}$.

(ii) (Flow) All supply units stay at their original locations.

Furthermore, the optimal revenue equals $\gamma \cdot \psi_1 \cdot \mu_1 \cdot 2H$.

This result simply says that if the initial demand-supply conditions are identical across the city, then the optimal price policy does not induce any movement for supply, and the optimal price at each location is simply that of a single location capacitated pricing problem. In such a solution, the expected utilization of all drivers is equal to 1 if $\mu_1 \leq \lambda_1 \cdot \overline{F}(\rho^u)$, and otherwise is strictly below 1. In the latter case, there is oversupply and driver congestion at all locations. The optimal revenue, recalling the reformulation in Proposition 1.1, is given by the equilibrium utility of drivers ψ_1, times the density of equilibrium supply, integrated across all locations (times a scaling factor).

1.6.2 Benchmark: Local Price Response to a Demand Shock

We next start our analysis of the demand shock environment. Before turning our attention to an optimal policy in Section 1.6.3, we first focus on a simple type of pricing heuristic which responds to changes in demand conditions through changes in prices *only* where these changes occur. In particular, in the context of the demand shock model, this corresponds to responding to a shock in demand at the origin by only adjusting the price at the origin; we call this policy the *local price response.* This provides a benchmark to better understand the structure and performance of an optimal policy. We next characterize an optimal local price response, when prices are fixed everywhere at the pre-demand shock environment solution, except at the origin.

Proposition 1.7 (Local Price Response to a Demand Shock) *Fix $\lambda_0 > 0$. Suppose that $p(x) = \rho_1$ for all x in $\mathcal{C} \setminus \{0\}$ and that the firm optimizes for the price $p(0)$. Then,*

(i) (Prices) The optimal price at the origin is given by $p(0) = \rho_0^{loc}(s^\tau(0))$, and $p(0) \geq \rho_1$.

(ii) (Movement) There exists two thresholds $X_r \geq X_r^0 \geq 0$, such that $X_r > 0$ and:

- *for all x in $[-X_r^0, X_r^0]$, all of the supply units move to the origin,*
- *for all x in $[-X_r, -X_r^0]$ and all x in $[X_r^0, X_r]$, a fraction of the supply units move to the origin and the other fraction does not move,*
- *for all x in $\mathcal{C} \setminus [-X_r, X_r]$, no supply unit moves.*

Furthermore, the platform's revenue is strictly larger than in the pre-demand shock environment.

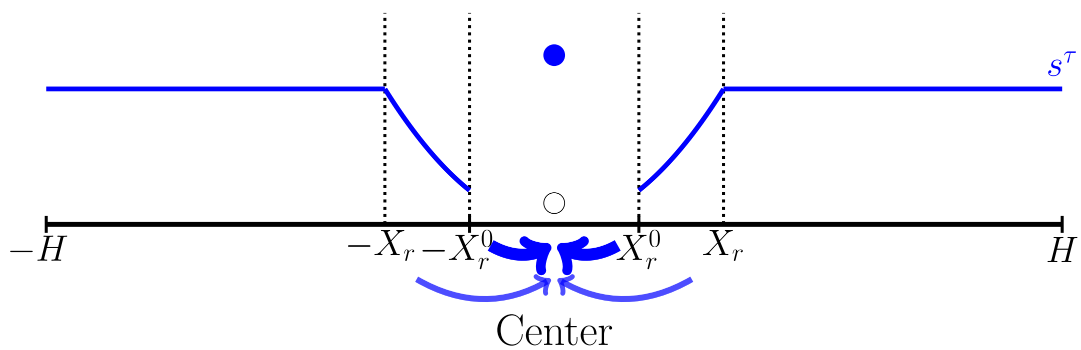


Figure 1.4: **Optimal local price response: induced supply response for a case with $\mu_1 > \lambda_1 \cdot \overline{F}(\rho^u)$.**

The result above characterizes the structure of an optimal local price response as well as the structure of the supply movement it induces. Figure 1.4 depicts the structure of the supply response. In particular, the optimal local price response leads to a higher price at the origin to respond to the surge of demand at that location. In turn, this higher price attracts drivers from a symmetric region around the origin. In that region, for locations close to the origin, all supply units move to the origin. After a given threshold X_r^0, only a fraction of the drivers will move to the origin. Intuitively, as one gets further from the origin, traveling to the origin becomes a less attractive option, compared to staying put or traveling elsewhere. As that becomes the case, a smaller and smaller fraction of units travels to the origin. Furthermore, we establish that supply units have no incentive to travel anywhere else in the city and, as a result, units that do not travel to the origin stay put and serve local demand.

Beyond the threshold X_r, no supply units move in the equilibrium induced by the optimal local price response.

The threshold X_r corresponds to the location of the last drivers willing to travel to the origin. In the current environment, prices are not flexible and, therefore, X_r must equals $V(0) - \psi_1$ since drivers who are further than that will prefer to earn ψ_1 by staying put compared to driving to the origin to earn $V(0)$ minus driving costs. If we are in a supply constrained regime, $\mu_1 \leq \lambda_1 \cdot \overline{F}(\rho^u)$, then all drivers within $[-X_r, X_r]$ drive to the origin, i.e., $X_r^0 = X_r$. However, in a supply unconstrained regime, $\mu_1 > \lambda_1 \cdot \overline{F}(\rho^u)$, the two thresholds are different, $X_r^0 < X_r$, as depicted in Figure 1.4. This occurs because in locations further from the origin but still within $[-X_r, X_r]$, as underutilized drivers drive toward the origin, conditions at the departing point improve and in equilibrium, staying put becomes competitive with driving to the origin.

1.6.3 Optimal Solution

The previous subsection provided an optimal local price response to a demand shock and the supply movement it induces. In this subsection, we focus on the optimal *global* price response across all locations in the city. To that end, we will leverage the results developed for the general model to obtain a quasi-closed form solution to the platform's problem in this specialized setting.

We begin by showing that the origin is an in-demand sink location and, therefore, the results from Sections 1.4 and 1.5 apply to the attraction region of the origin.

By leveraging structural properties of the equilibrium utility function, the congestion bound, and a novel flow-mimicking technique, we next fully characterize in Theorem 1.2 the optimal equilibrium utility of supply units $V(\cdot)$, not only in the attraction region of the origin, but across the entire city. In particular, this characterization yields a spatial separation of the city into three attractions regions and regions of no-movement. Leveraging Theorem 1.1 and a symmetry argument, we solve for the optimal s^τ and the corresponding prices in each attraction region. The solution for the no-movement regions reduces to the pre-shock environment. Leveraging the pasting result (cf. Proposition 1.4) yields the optimal solution to the platform's problem as presented in Theorem 1.3.

Our first result in this section demonstrates that we can focus on price-equilibrium pairs such that the high demand location is a sink that has drivers coming towards it from left and right.

Lemma 1.5 (Origin is in-demand sink) *Without loss of optimality, one can restrict attention to price-equilibrium pairs* (p, τ) *such that the origin is an in-demand sink such that* $X_l < 0 < X_r$.

The intuition behind this proposition harks back to the fact that the performance of the pre-shock environment is dominated by that of the local price response solution. Solutions for which the origin is not an in-demand sink have revenues capped by that of the pre-demand shock environment. At a high-level, in those solutions, there is no positive mass of drivers willing to travel to the demand shock location and, thus, the city resembles a city without a demand shock. However, the local price response solution incentivizes drivers from both sides to travel to the demand shock and has a strictly larger revenue. This implies that at optimality we must have drivers coming from both sides to the origin, that is, $X_l < 0 < X_r$.

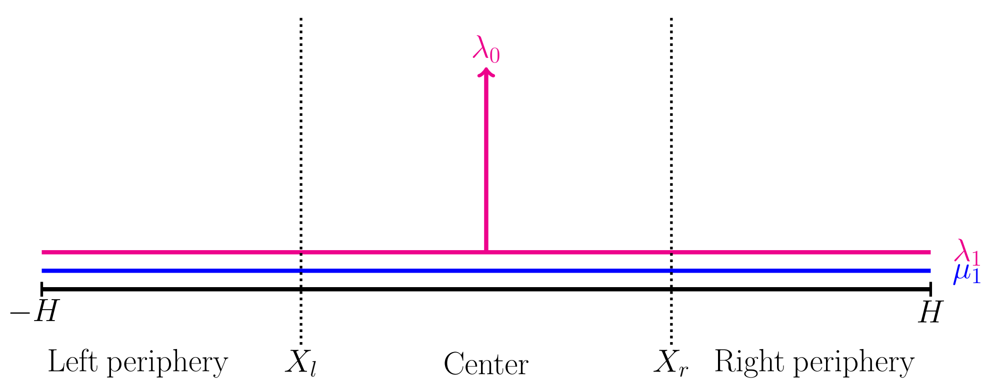


Figure 1.5: Three region-decomposition.

In what follows we solve for the key objects of the platform's optimization problem $(\mathcal{P}_2)$. To make our exposition clear and highlight the solution's spatial aspects, we call the interval $[X_l, X_r]$ the *center* region, and the region outside of it will be referred to as the *periphery* (see Figure 1.5).

1.6.3.1 Equilibrium Utilities

In this subsection we characterize $V(\cdot)$ throughout $\mathcal{C}$. We begin by stating the main result of this subsection. We then we discuss some of the implications and associated intuition.

Theorem 1.2 *(Equilibrium utilities) Under an optimal price-equilibrium pair* (p, τ)*, the equilibrium utility function* $V(\cdot)$ *is fully parametrized by the three values* $V(0)$ *and* X_l, X_r *as follows:*

$$V(x) = \begin{cases} V(0) - |x| & \text{if } x \in [X_l, X_r], \\ \min\{V(0) - 2X_r + x, \psi_1\} & \text{if } x > X_r, \\ \min\{V(0) - 2|X_l| + |x|, \psi_1\} & \text{if } x < X_l. \end{cases}$$

Moreover, $V(0) > \psi_1$ *and* $V(X_l), V(X_r) \leq \psi_1$*.*

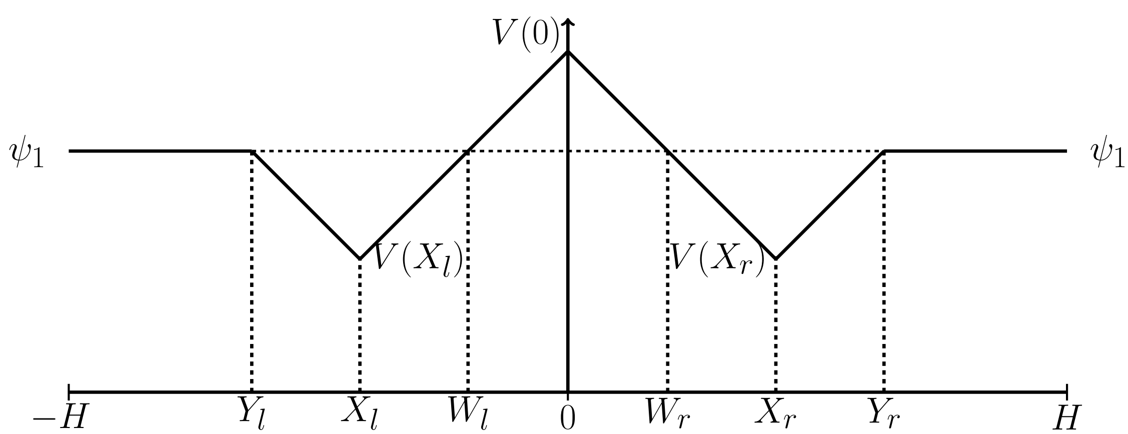


Figure 1.6: **Drivers' equilibrium utility under an optimal pricing policy.** The equilibrium utility is fully characterized up to $V(0)$, X_l and X_r.

The first main implication of this result is that we know exactly how much utility each supply unit garners under optimal prices throughout the entire city. Quite strikingly, the characterization of $V(\cdot)$ is "independent" of the flows. That is, in order to characterize the equilibrium utility we did not need to pin down the distribution of after-movement supply.

The second implication is that the city has at most three types of regions. Figure 1.6 depicts the equilibrium utility function. The center $[X_l, X_r]$ is by definition an attraction region. Let W_r and Y_r be defined as the points to the left and to the right of X_r where the driver's equilibrium utility function equals the pre-shock utility level ψ_1. To the right of the origin (and similarly to the left), we can observe three main regions. We first have the interval $[0, W_r]$, where drivers' utilities are above the pre-shock utility level. Drivers in this region are positively impacted by the shock of demand at the origin (and the global optimal prices). The second region $[W_r, Y_r]$ is notable. Here, drivers garner strictly less utility compared to the pre-shock environment. In $[W_r, X_r]$ drivers are "too far" from the origin so their utilities are negatively affected by the cost of driving to the origin. Drivers in $[X_r, Y_r]$ are outside the origin's attraction region and, thus, do not relocate to the origin. Interestingly, drivers in $[X_r, Y_r]$ suffer because the platform has to make sure that drivers in $[0, X_r]$ stay within the attraction region of the origin. For the marginal drivers at X_r to be willing to travel to the origin, the conditions to the right of X_r should not be too attractive. The final region corresponds to $[Y_r, H]$; this region is not affected by the shock of demand as it is effectively too far from the origin.

Key ideas for the proof of Theorem 1.2. We now present the main arguments that enable us to establish Theorem 1.2. At a high level, we focus on each region separately, center and periphery, and solve for $V(\cdot)$ in each of these regions.

We start by considering the center region, which is easy to analyze. Lemma 1.5 establishes that we can focus on solutions such that $A(0) = [X_l, X_r]$ is a non-empty interval that strictly contains the origin. Our envelope result (Lemma 1.3) characterizes the equilibrium utility function in any attraction region. In turn, this implies that

$$V(x) = V(0) - |x|, \qquad \text{for all } x \in [X_l, X_r].$$

Importantly, the characterization of $V(\cdot)$ in this region only depends on three parameters, namely, $V(0)$, X_l and X_r. In Section 1.7, we will leverage this fact to numerically compute the optimal value for these parameters.

We now switch our attention to the periphery. Consider the right periphery $(X_r, H]$. We first argue that, in this region, the drivers' equilibrium utility has a non-trivial upper bound, and then establish that this upper bound is achieved. The treatment for the left periphery is analogous.

Lemma 1.6 *(Upper bound) An optimal price-equilibrium pair (p, τ) satisfies*

$$V(x) \leq \min\{V(X_r) + x - X_r, \psi_1\}, \qquad \textit{for all } x \in (X_r, H]. \tag{1.7}$$

The upper bound above follows from two bounds. A first upper bound can be derived using the 1-Lipschitz property of V (Lemma 1.1), which ensures that V can grow at a rate of at most 1. Thus, $V(x)$ is bounded by $V(X_r)+x-X_r$. A second bound may be obtained by leveraging the congestion bound (Proposition 1.5). One may show that that drivers from almost any location that do not have an incentive to travel to the origin have their utilities capped by the pre-demand shock utility level ψ_1. Locations different than the origin that receive supply increase their driver congestion with respect to the initial congestion level which, in turn, reduces the driver utility at that location. In addition, drivers traveling to these locations have to incur a transportation cost further decreasing their utilities. Thus $V(\cdot)$ has to be bounded by ψ_1 in $(X_r, H]$.

The core of the argument toward characterizing the equilibrium utilities in the periphery resides in establishing that the upper bound in Eq. (1.7) is always binding for any x in $(X_r, H]$, a result we will present in Proposition 1.9. We show this result in two steps: we first establish that the value function has to be non-decreasing in $[X_r, H]$ and then leverage this to establish that the upper bound is achieved under an optimal pricing policy.

By our characterization of a driver's utility in an attraction region (see Lemma 1.3), the upper bound would not be binding if there were drivers willing to move left in $(X_r, H]$. That would imply the existence of an attraction region (see Proposition 1.2) inside of which $V(\cdot)$ is decreasing. Our first proposition proves this cannot happen by establishing that, in an optimal solution, $V(\cdot)$ is a non-decreasing function in the right periphery.

Proposition 1.8 *(Monotonicity in the periphery) Without loss of optimality, we can focus on price-equilibrium pairs (p, τ) such that $V(\cdot)$ is non-decreasing in $(X_r, H]$. Furthermore, if $V(X_r) = \psi_1$, then $V(x) = \psi_1$ for all $x \geq X_r$.*

We first observe that the attraction region around the origin of the demand shock location is always wider under the optimal solution than under the local best response. That is, $A^{\text{lr}}(0) \subset A^{\text{opt}}(0)$. In particular, this means that more locations are affected by a demand shock in the optimal solution than under the local price response. Hence, the largest interval in which both solutions differ corresponds to $[-Y_r^{\text{opt}}, Y_r^{\text{opt}}]$. We denote this interval by $\mathcal{C}_{\text{diff}}$.

The key argument behind the proof of Proposition 1.8 is to construct a (strictly) profitable deviation whenever $V(\cdot)$ is decreasing in some region. We illustrate the main idea of the argument in Figure 1.7. Suppose the value function is decreasing in some interval as illustrated in Figure 1.7(a). We will construct a deviation over a superset of that interval, denoted by $[y_0, y_1]$ in the figure. The construction of a deviation contains three main ideas.

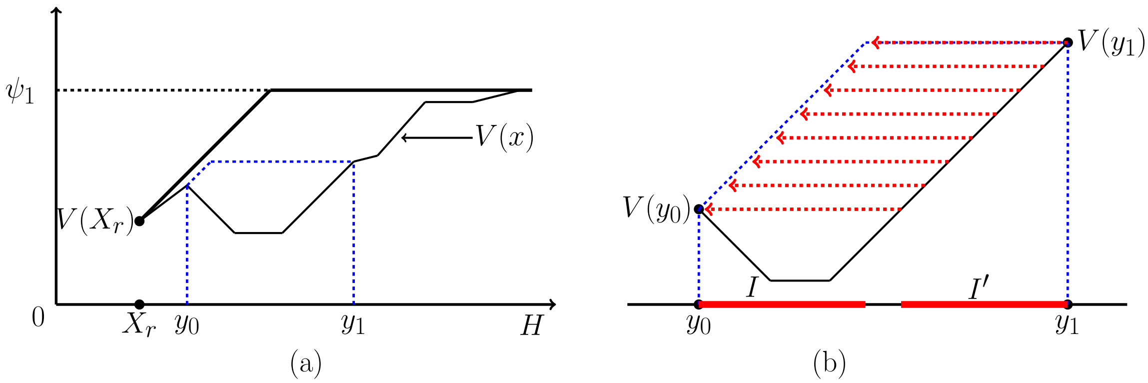


Figure 1.7: Illustration of the main argument in the proof of Proposition 1.8.

First, the interval $[y_0, y_1]$ is constructed in such a way that it is flow separated. That is, there is no flow of drivers leaving this interval and no drivers coming in ($\tau - a.e$). This separation permits us to analyze this region as an individual sub-problem, where the behavior of drivers is relatively "controlled". In particular, we construct the interval $[y_0, y_1]$ in such a way there is at most one maximal subinterval where $V(\cdot)$ decreases at rate -1, and at most one maximal subinterval where where V increases at rate 1. Where $V(\cdot)$ decreases at rate -1 drivers can only move left, and where $V(\cdot)$ increases at rate 1 drivers can only move right.

Second, the best incentive compatible deviation that ensures a non-decreasing value function coincides with the dashed blue line. Because V can increase at most at a rate of 1, after y_0 the best deviation equals $V(y_0) + (x - y_0)$ (recall Eq. (1.7)). Moreover, since the interval ends at y_1 and we want the deviation to be a non-decreasing function, it has to be bounded by $V(y_1)$.

The final idea is a subtle, but critical one. We know from Proposition 1.1 that the platform earns revenues from a location x proportionally to $V(x) \cdot s^\tau(x)$. As a result, one needs to focus on *both* $V(\cdot)$ and the post-movement supply s^τ to establish a profitable deviation. We need to argue that overall the platform will earn higher revenues after the drivers move. Our argument, which relies on judicious price setting as well as a proper mapping of revenue contributions in different space regions between the old and new flows, is illustrated in Figure 1.7(b). We set prices in such a way that it is incentive compatible for drivers not to move within the interval $[y_0, y_1]$, except for the region we denote by I near y_0. In this region, we set prices to incentivize the drivers to behave as they did in region I' in the old (non-monotone) solution. This enables us not only to achieve the upper bound constructed, but also to obtain a strict revenue improvement for the platform.

In brief, at the optimal solution, $V(\cdot)$ must be a non-decreasing function in $(X_r, H]$. This implies that drivers only move right (or do not move) in the right peripheral region. Our next result shows that Eq. (1.7) is indeed binding.

Proposition 1.9 *(Tight upper bound) Without loss of optimality, we can focus on price-equilibrium pairs* (p, τ) *such that the upper bound in Eq.* (1.7) *is tight.*

The proof of Proposition 1.9 relies on the monotonicity in the periphery of $V(\cdot)$ to construct a strict improvement whenever we have a solution (p, τ) for which the upper bound in Eq. (1.7) is not tight. We start by separating intervals that form maximal attraction regions, that is, attraction regions with a sink at an end point. In these regions, $V(\cdot)$ is differentiable and has slope equal to 1. Such intervals can be mapped onto the interval where the upper bound in Eq. (1.7) also has slope 1. This mapping in represented by dashed lines and arrows in Figure 1.8.

We can then use a flow mimicking argument similar to the one used in Figure 1.7(b). The solutions in the initial intervals in the mapping can be replicated in the new intervals, which we illustrate in Figure 1.8. Thus, this mapping preserves the platform's revenue in the intervals being mapped. The regions that are left after the mapping (thick black lines in the figure) are given prices such that drivers in them prefer not to relocate, and V coincides with the upper bound. By pasting the solutions in the intervals we obtain then a solution for which the upper bound is tight and whose revenue is strictly larger than that of (p, τ).

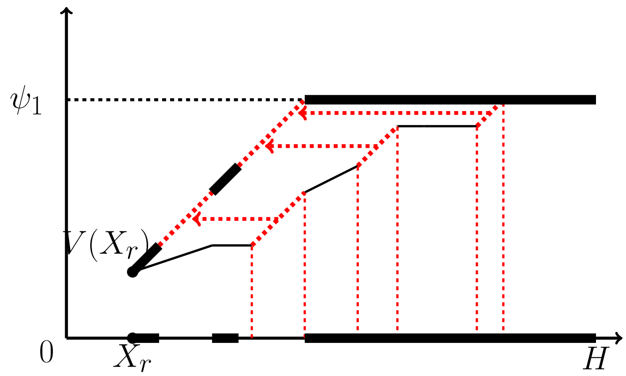


Figure 1.8: Illustration of the main idea underlying the proof of Proposition 1.9. The dashed lines in $V(x)$ correspond with interval where $dV(x)/dx = 1$. These intervals are mapped onto the intervals in $[X_r, H]$ where the upper bound in Eq. (1.7) has slope 1. The thick black lines correspond to both the intervals and parts of the upper bound that are left after the mapping.

1.6.3.2 From Equilibrium Utilities to Supply Distribution and Optimal Prices

Given that we pinned down the equilibrium utility function across the city, the natural next step as prescribed by the problem reformulation in Proposition 1.1 is to solve for prices and supply.

Theorem 1.3 *(Optimal prices and flows) An optimal price-equilibrium pair (p, τ) is such that $V(\cdot)$ is as in Theorem 1.2, $X_r = -X_l$, and prices and flows are characterized as follows.*

1. *(Prices) The optimal prices are given by $p(x) = \rho_x^{loc}(s^\tau(x))$, where $s^\tau(x)$ is as below.*

2. *(Post-relocation supply) There exists unique $\beta_c \in [0, W_r]$ and $\beta_p \in [X_r, Y_r]$ such that*

$$\int_{-\beta_c}^{\beta_c} \psi_x^{-1}(V(x))d\Gamma(x) = \mu_1 \cdot 2 \cdot X_r \quad \textit{and} \quad \int_{\beta_p}^{Y_r} \psi_x^{-1}(V(x))d\Gamma(x) = \mu_1 \cdot (Y_r - X_r),$$

and the optimal post-relocation supply is given by

$$s^{\tau}(x) = \begin{cases} 0 & \text{if } x \in (\beta_c, \beta_p) \cup (-\beta_p, \beta_c), \\ \psi_x^{-1}(V(x)) & \text{otherwise.} \end{cases}$$

3. *(Movement)*

- *for all x in $[-\beta_c, \beta_c]$, drivers move in the direction of the origin,*
- *for all x in $[-X_r, -\beta_c) \cup (\beta_c, X_r]$, all drivers move to $[-\beta_c, \beta_c]$,*
- *for all x in $[X_r, \beta_p)$, all drivers move to $[\beta_p, Y_r]$.*
- *for all x in $(-\beta_p, -X_r]$, all drivers move to $[-Y_r, -\beta_p]$.*
- *for all x in $[\beta_p, Y_r]$, drivers move in the direction of Y_r,*
- *for all x in $[-Y_r, -\beta_p]$, drivers move in the direction of $-Y_r$,*
- *for all x in $[-H, -Y_r) \cup (Y_r, H]$, drivers do not relocate.*

The key idea underlying Theorem 1.3 is to recognize the structure of the regions. The center $[X_l, X_r]$ is by definition an attraction region. The other two attraction regions correspond to the intervals $[Y_l, X_l]$ and $[X_r, Y_r]$ (to recall the definitions of these terms, please revisit Figure 1.6). Consider the last of these intervals. In it, $V(\cdot)$ increases at a rate of 1 and drivers only move towards Y_r but not beyond it. The shape of $V(\cdot)$ then ensures that all drivers in this region are willing to travel to Y_r and, therefore, this location has to be a sink with its associated attraction region being $[X_r, Y_r]$. We can thus leverage Theorem 1.1 to characterize the flow structure within attraction regions and then paste solutions appropriately. Finally, we show that the optimal solution has to be symmetric around the origin. In particular, now all the relevant quantities that characterize the optimal solution depend only on two values: $V(0)$ and X_r.

Discussion. We depict in Figure 1.9 the structure of the solution obtained in Theorem 1.3. The main feature of the optimal solution is that it separates each side

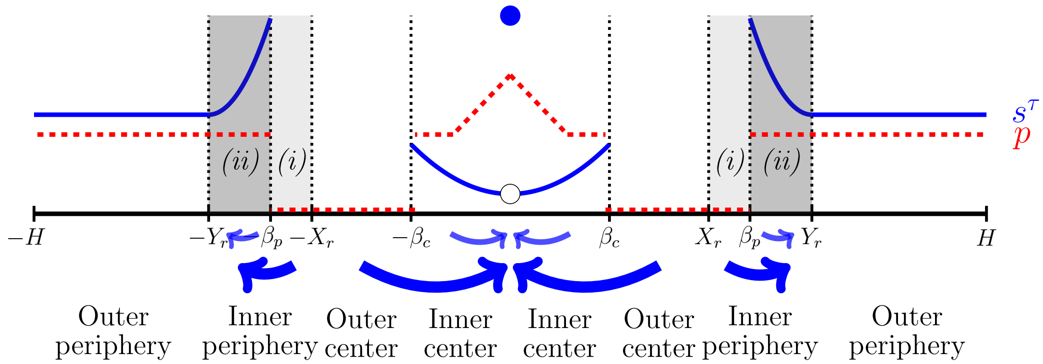


Figure 1.9: **Supply response (solid-blue line) induced by optimal prices (dashed-red line).**

of the city, with respect to the origin, into six regions. Without loss of generality, we focus our discussion on the right side of the city.

The origin receives a mass of supply equal to $\psi_0^{-1}(V(0))$. This mass of drivers comes from two regions, the inner and the outer center, which we now define. The first corresponds to the interval $(0, \beta_c]$. Some drivers in this region choose to stay put while others, attracted by the favorable conditions at the center of the city, choose to drive to the origin. In equilibrium, drivers staying or traveling to the origin garner the same utility. The outer center is the interval $(\beta_c, X_r]$. Here, the platform sets prices to $\overline{V}$ (or 0) and therefore supply is equal to zero. That is, the platform chooses prices to shut down demand, giving no incentive for drivers to stay there (or alternatively sets prices at zero to again give no incentive for drivers to stay there). In turn, this incentivizes all drivers in this region to move somewhere else. In order to incentivize these drivers to move towards the origin, the platform creates one more region: the inner periphery.

The inner periphery corresponds to the interval $(X_r, Y_r]$. The platforms "artificially" degrades the conditions for drivers in this interval in two different ways, leading to the two sub regions, *(i)* and *(ii)* in Figure 1.9. In region (i), the platform sets prices equal to $\overline{V}$ (or 0) in $(X_r, \beta_p]$, shutting down demand, so no drivers want to either travel to or stay in this region. As a result the interval $(\beta_p, Y_r]$ receives all drivers from $(X_r, \beta_p]$. This creates driver congestion and, thus, endogenously worsens driver conditions in the interval $(\beta_p, Y_r]$. The reason the platform selects these inner periphery prices is to discourage drivers in the outer center from driving towards the periphery. Quite strikingly, the optimal global price response to a demand shock at the origin induces supply movement *away* from the origin in the inner periphery. The final region is the outer periphery. All drivers in this region stay put, leading to $s^\tau(x) = \mu_1$. Here, drivers collect the same utility they would make if there was no demand shock at the origin.

In sum, the optimal global price response to a demand shock, while correcting the supply-demand imbalance at the origin, also creates significant imbalances across the city. This is driven by the self-interested nature of capacity units and the need to incentivize them through spatial pricing. In particular, we observe that the structure of the optimal pricing policy is very different from that of the local price response (cf. Proposition 1.7).

1.7 Local Price Response versus Optimal (Global) Prices

In this section, we will use the optimal local price response solution as a benchmark for comparison to put the optimal solution in perspective. The objective is to illustrate through several metrics the different features of the optimal solution as well as its performance in terms of revenue maximization and welfare. Throughout this section, we use superscripts *lr* and *opt* to label relevant quantities associated with the local price response and optimal solution, respectively (except when obvious from the context).

We first observe that the attraction region around the origin of the demand shock location is always wider under the optimal solution than under the local best response. That is, $A^{\text{lr}}(0) \subset A^{\text{opt}}(0)$. In particular, this means that more locations are affected by a demand shock in the optimal solution than under the local price response. Hence, the largest interval in which both solutions differ corresponds to $[-Y_r^{\text{opt}}, Y_r^{\text{opt}}]$. We denote this interval by $\mathcal{C}_{\text{diff}}$.

Next, we illustrate and discuss through a set of numerics the differences between the two policies. We consider a range of instances that includes various levels of supply availability. We fix the city to be characterized by $H = 1$ and assume that the demand is uniformly distributed across locations with $\lambda_1 = 4$. The origin experiences a shock of demand ranging from low to high: $\lambda_0 \in \{3, 6, 9\}$. We vary the initial supply $\mu_1 \in \{1, 1.5, 2, \ldots, 4.5, 5\}$ so that when low, the city (excluding the origin) is supply constrained, and when high, the city is supply unconstrained. Consumer valuation is uniformly distributed in the unit interval. Note that the city (excluding the origin) is supply constrained whenever $\mu_1 < \lambda_1 \cdot \overline{F}(p^u) = 2$. To eliminate any strong dependence on the choice of H, for each instance, we compare the local price response performance and optimal solution performances within the sub-region of the city corresponding to the largest interval in which both solutions differ, $\mathcal{C}_{\text{diff}}$. Given the symmetry of the solutions, in all that follows we focus on the right side of the city $[0, H]$.

Policy structure. Figure 1.10 depicts the core spatial thresholds characterizing the optimal pricing policy and the local price response as the supply conditions μ_1 changes (on the y-axis). In particular, we track the changes in X_r, β_p, β_c and Y_r for the optimal solution (cf. Theorem 1.3) and the changes in X_r and X_r^0 for the local price response (cf. Proposition 1.7).

The first thing to note is that the structure of supply in the attraction region of 0 differs significantly between the local price response and the optimal policy. In the local price response, there are no drivers who stay put around the origin; and post-relocation, drivers are either at the origin or in $[X_r^{0,\text{lr}}, X_r^{\text{lr}}]$. In contrast for the the optimal policy, there are no drivers in a region separated from the origin $[\beta_c, X_r^{\text{opt}}]$

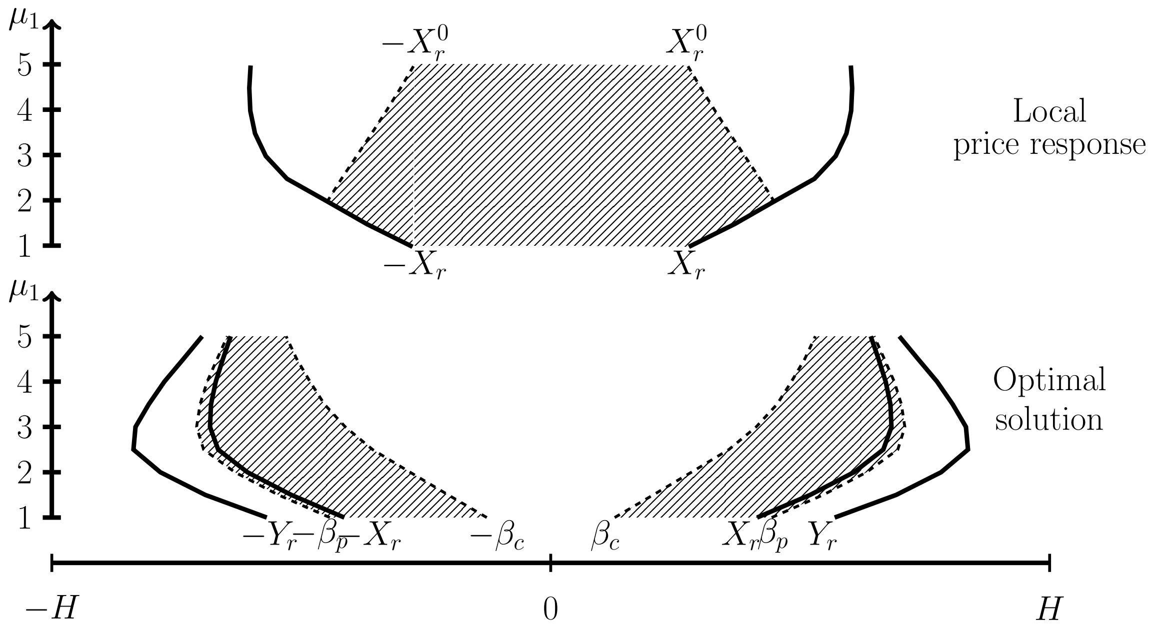


Figure 1.10: **Policy structure.** Spatial thresholds characterizing the optimal pricing policy and the local price response as the the supply conditions change. The shaded regions have no supply in equilibrium. The figure assumes $\lambda_0 = 9$ and $\lambda_1 = 4$.

but there are drivers in $[0, \beta_c]$. This contrast can be better understood through the reformulation of the objective in Proposition 1.1, in conjunction with the shape of the equilibrium utility function in the attraction region of 0 (cf. Lemma 1.3). Given the objective, the platform would ideally like to have supply as close to the origin as possible (subject to the congestion bound constraint) as it maximizes the integral of $V(x) \cdot s^{\tau}(x)$. With a local price response, as a result of the lack of flexibility in setting prices throughout the city, the platform is unable to "optimize" the supply in the attraction region and ends up with drivers at locations with low V in $[X_r^{0,\mathrm{lr}}, X_r^{\mathrm{lr}}]$ while locations with higher V's have no drivers in $(0, X_r^{\mathrm{lr}}]$. Meanwhile, the optimal policy is able to set prices so as to induces the best possible distribution of supply in the attraction region.

In the periphery of the optimal solution, which is outside the origin's attraction region under pricing policy, the local price response behaves exactly as in the pre-demand shock environment. In stark contrast, the optimal solution incentivizes movement of drivers from the periphery away from the demand shock. In particular, the region $[X_r, Y_r]$, which has a non-trivial size, is artificially damaged. This region is needed for the optimal solution to steer more drivers towards the origin, an issue we address in more detail in the revenue improvement discussion below.

Revenue Improvement. The revenue performance of the optimal solution with respect to our benchmark in $\mathcal{C}_{\text{diff}}$ is shown in Table 1.1.

Table 1.1: Revenue improvement (in %) of optimal solution over optimal local prices response solution in $\mathcal{C}_{\text{diff}}$.

μ_1	1	1.5	2	2.5	3	3.5	4	4.5	5
$\lambda_0 = 3$	2.05	4.64	9.59	13.02	13.87	12.92	11.00	8.60	5.91
$\lambda_0 = 6$	2.17	3.11	4.99	8.73	9.96	10.01	9.56	8.92	8.21
$\lambda_0 = 9$	2.69	3.51	4.69	8.75	10.16	10.30	9.81	9.10	8.29

For any level of demand shock, we observe that the revenue improvement reaches its maximum value for medium to high levels of supply, and can be significant, above 10%.

In order to appreciate where the revenue gains stem from, consider Figure 1.10 and Table 1.2 below, which summarizes some key quantities for the case $\mu_1 = 3$, $\lambda_0 = 9$ (so that ψ_1 equals 0.27). Let us analyze the various contributions to revenues

$V^{\text{opt}}(0)$	$s^{\text{opt}}(0)$	$p^{\text{opt}}(0)$	X_r^{opt}	Y_r^{opt}	$V^{\text{lr}}(0)$	$s^{\text{lr}}(0)$	$p^{\text{lr}}(0)$	X_r^{lr}	$X_r^{0,\text{lr}}$
0.62	1.97	0.78	0.46	0.57	0.65	1.66	0.81	0.38	0.25

Table 1.2: Metrics for the local response and optimal solution for the case $\mu_1 = 3$, $\lambda_0 = 9$.

under both policies. We start by noticing that the drivers' equilibrium utility at the shock location is lower under the optimal solution than under the local price response, $V^{\text{opt}}(0) = 0.62$ and $V^{\text{lr}}(0) = 0.65$. However, since $X_r^{\text{opt}} = 0.46$ and $X_r^{\text{lr}} = 0.38$, the optimal solution is able to incentivize the movement of a larger mass of drivers towards the demand shock, leading to a mass $s^{\text{opt}}(0) = 1.97$ and $s^{\text{lr}}(0) = 1.66$. Focusing on the objective reformulation in Proposition 1.1, this extra mass of drivers delivers

0.14 units $(0.62 \times 1.97 - 0.65 \times 1.66)$ of extra revenue to the platform. The revenue difference is further increased by the fact that the remainder 0.79 units of drivers in the attraction region of zero $(2 \times 3 \times 0.46 - 1.97)$ in the optimal solution travel to locations nearby the demand shock, where $V(\cdot)$ is close to 0.62. In contrast, the benchmark solution has the remainder 0.62 drivers $(2 \times 3 \times 0.38 - 1.66)$ staying within $[X_r^{0,\text{lr}}, X_r^{\text{lr}}]$ where $V(\cdot)$ is below 0.37 $(V^{\text{lr}}(0) - X_r^{0,\text{lr}})$. Through these two mechanisms, the optimal policy garners more revenue than the benchmark solution in the region $[-X_r^{\text{opt}}, X_r^{\text{opt}}]$.

However, the benefits come at a cost. In particular, to induce the "right" incentives in the shock's attraction region, the platform has to alter conditions to the right of the attraction region. In order to incentivize the movement of drivers in $[-X_r^{\text{opt}}, X_r^{\text{opt}}]$ towards the demand shock, the region $[X_r^{\text{opt}}, Y_r^{\text{opt}}]$ is damaged by having the 0.22 units of drivers in it $(2 \times (0.57 - 0.46))$ contributing values strictly below $\psi_1 = 0.27$ to the platform's objective. The same units of drivers in the benchmark solution contribute exactly 0.27 per unit to the platform's revenue. This cost is offset by the proceeds that incentivizing the movement of a larger amount of drivers toward the demand shock generates.

Welfare Implications. The revenue improvement of the optimal solutions relies on creating a special region in which drivers' utilities are below of what they could earn if the platform responded only locally to the demand shock. This raises the question of whether revenue-optimal pricing leads to lower or higher surpluses for drivers and consumers compared to the benchmark solution.

The social welfare (SW) equals the sum of the platform's revenue, and the driver (DS) and consumer surpluses (CS), as given by

$$DS = \int_{\mathcal{C}_{\text{diff}}} V(x) d\mu(x), \quad CS = \int_{\mathcal{C}_{\text{diff}}} \mathbf{E}[(v - p(x)) | v \geq p(x)] \cdot \min \left\{ s^\tau(x), \lambda_x \cdot \overline{F}(p(x)) \right\} d\Gamma(x).$$

Driver surplus corresponds to nothing more than the integral of driver equilibrium utilities across all locations in $\mathcal{C}_{\text{diff}}$. Similarly, consumer surplus corresponds to the gains enjoyed across $\mathcal{C}_{\text{diff}}$ by all those consumers who are willing to pay and are matched to some driver.

In Table 1.3 we display the percentage differences of driver and consumer surpluses, as well as social welfare between the optimal and benchmark solutions. We note that there are instances where the optimal solution is a Pareto improvement over the local price response, in the sense that it is better for the platform, drivers and consumers. There are also instances where the platform's revenue gain is at the expense of both drivers and consumers.

Table 1.3: Driver surplus, consumer surplus and social welfare difference (in %) of optimal solution over optimal local prices response solution in $\mathcal{C}_{\text{diff}}$.

μ_1		1	1.5	2	2.5	3	3.5	4	4.5	5
	$\lambda_0 = 3$	-0.67	3.09	11.3	13.64	14.6	12.44	10.00	7.53	4.92
DS	$\lambda_0 = 6$	-4.15	-3.99	-1.62	-2.01	-0.82	0.74	3.00	5.35	7.80
	$\lambda_0 = 9$	-6.22	-7.35	-7.48	-9.45	-9.72	-9.02	-8.14	-6.36	-4.32
	$\lambda_0 = 3$	-10.96	-14.1	-18.48	-7.24	-3.15	-0.44	1.01	1.57	1.58
CS	$\lambda_0 = 6$	-12.03	-10.58	-17.15	-6.32	1.18	4.18	4.24	2.85	0.69
	$\lambda_0 = 9$	-14.33	-11.94	-22.43	-12.58	-1.39	5.77	9.73	10.98	10.44
	$\lambda_0 = 3$	-1.04	0.81	4.26	8.28	9.70	8.83	7.44	5.8	3.96
SW	$\lambda_0 = 6$	-3.60	-3.56	-3.49	-1.05	1.50	3.16	4.43	5.29	5.87
	$\lambda_0 = 9$	-5.24	-5.95	-8.16	-6.84	-4.40	-2.32	-0.86	0.51	1.58

For a given level of supply, the driver surplus degrades with respect to the benchmark as the demand shock becomes more intense. We also find that, independently of the size of the demand shock, the optimal solution performs better than the benchmark in terms of consumer surplus when the supply level is high. More drivers in the city imply more matches and lower prices and, thus, higher consumer surplus.

Chapter 2

Spatial Capacity Planning

2.1 Motivation and Overview of Results

Many traditional service systems are characterized by static servers and customers that arrive stochastically and line up in a queue before receiving service. These include call centers, health-care facilities, and amusement parks, among others. In designing such systems, one faces a tradeoff between the cost of servers and the quality of service as measured through the characteristics of wait time. The prevalence of such systems has led to an extensive literature on capacity sizing that has provided important practical guidelines about how to set capacity levels in service systems. Typically, there is a fine balance between the two objectives. A central rule, the so-called square root safety (SRS) staffing rule, emerges naturally from different performance considerations. In the SRS rule, the capacity is set at the nominal offered load plus a safety factor proportional to the square root of the offered load. If one considers a social planner's problem attempting to minimize the system's total cost measured by the aggregate of capacity and waiting costs, the SRS rule is optimal in large systems. Another central metric in the literature and in practice is the probability that a customer waits before being attended by a server, which has led to the coining of various terms to describe the regimes of interest. Quality driven (QD) is the regime where customer quality is paramount and, thus, the probability of waiting is vanishingly small. Efficiency driven (ED) refers to the regime where cost concerns prevail. In ED, a customer's probability of having to wait approaches one. Quality and Efficiency driven (QED) is the intermediate regime, where the probability that a customer waits is separated from both zero and one, leading to a fine balance between utilization and quality of service. The latter is achieved through the SRS staffing rule. The latter capacity is sufficient to ensure that a positive fraction of customers do not wait at all before receiving service.

However, there are other service systems in which customers arrive to random locations in space and servers have to spend time not only servicing customers, but also reaching them before service starts. This includes, for example, ride-hailing systems such as Uber, Lyft, Via and DiDi. On these platforms, a customer requests a ride from a given location and a driver is then dispatched by the platform to pick him up and take him to his desired destination.[1] Automated warehouses powered by Kiva robots (Amazon robotics) or the Ocado smart platform provide another example. In these warehouses, products are arranged in a grid. As orders for different products arrive, robots are dispatched to collect the products and transport them to picking stations. In these spatial multi-server systems, workload is larger than in traditional systems because servers must reach customers before starting to service them, making it unclear whether the SRS rule of thumb is still valid. The central question of this chapter is the following: *How should "capacity thinking" be adapted to spatial settings, where servers need to reach customers before service can start?*

We anchor our analysis around a *spatial multi-server system* in which arrivals to a two-dimensional region follow a Poisson process. A customer draws an origin and a destination uniformly and independently in the region. From a pool of n servers, a central platform dispatches a server that must reposition to the origin of her assigned customer and then take him to his desired destination. This spatial multi-server system is different from a traditional queueing $M/M/n$ service system in at least

[1] For consistency, we refer to customers as males and servers/drivers as females throughout the chapter.

two dimensions. First, servers must "pick up" customers by repositioning to a customer location before starting service. This translates into extra workload added to the system compared to a traditional system. Second, as the imbalance of servers and customers increases, *spatial economies of scale* can make the system operate at a faster pace. For example, the larger the spatial density of idle servers, the more opportunities for better matches and the shorter the time it takes a server to pick up an arriving customer. Similarly, the larger the spatial density of waiting customers, the more opportunities for better matches and the shorter the time it takes an idling server to reach a customer. That is, in a spatial multi-server system, service rate is state-dependent and might improve with large supply-demand imbalances. This is illustrated in Figure 2.1. In order to shed light on the capacity sizing question of in-

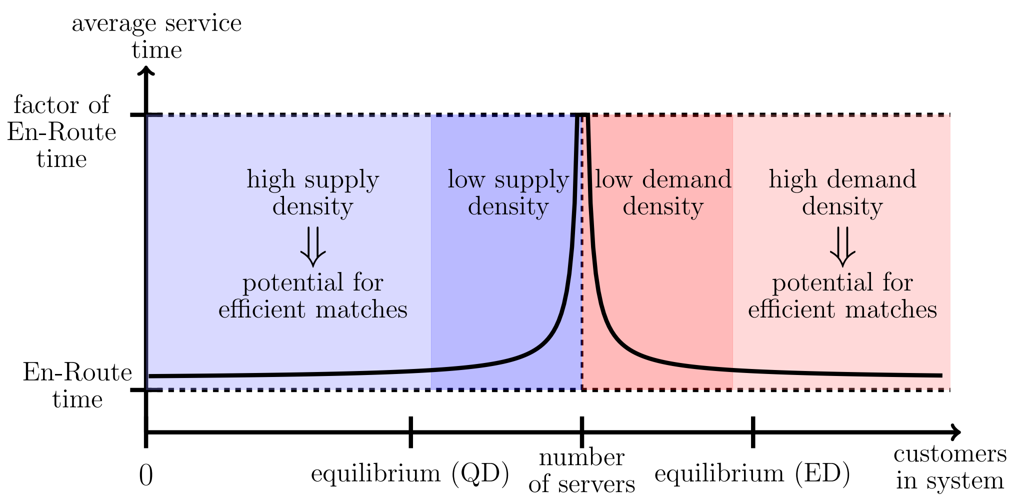


Figure 2.1: Illustration of the potential for matches and the impact on pickup times.

terest, we take a macro view of the spatial system by focusing on the key features that dictate its dynamics. More concretely, we consider a Markovian stochastic model that captures the key characteristics of input and output rates in the spatial multi-server system. Our Markovian model is a standard queueing system with n servers, but with a *state-dependent* service rate that adequately reproduces the spatial economies of scale of spatial systems. We analyze this queueing model in heavy traffic. On the one hand, the queueing setting provides guidelines for how the spatial system will behave. On the other hand, the spatial setting provides a physical interpretation of the queueing model results.

Main contributions. Our first contribution lies in the modeling domain. We develop a Markovian model that captures fundamental aspects of capacity planning in dynamic spatial environments. The system we analyze features both service speedups and service slowdowns that emerge due to the presence of spatial economies of scale. In addition, we ground our analysis on near-optimal dispatch rules derived from the vehicle routing literature.

Our second contribution lies in the set of insights and fundamental results we obtain for this class of problems. We first analyze a fluid model that highlights some of the key properties of such systems. We characterize in closed form the two possible stable equilibria of this deterministic model. These equilibria correspond to two types of potential operating regimes: the first one with a high density of waiting customers and the second one with a high density of idle servers. These equilibria are depicted in Figure 2.1. In both of these operating points, the system is able to match customers to servers efficiently since supply and demand are fairly imbalanced.

We then analyze the stochastic system in heavy traffic. In this setting we first establish that, in stark contrast with a standard multi-server system, the SRS rule will always bring the spatial multi-server system to the efficiency-driven (ED) regime, in which customers will wait for a server to be dispatched with probability approaching one. In other words, the added workload due to pickups is substantial enough and cannot be compensated by simply increasing capacity levels on the order of the square root of the offered load.

In turn, we fully characterize the asymptotic system's performance under a range of scalings. If the capacity buffer is of lower order than the offered load to the power of 2/3, then the system is in the efficiency-driven (ED) regime. The system operates around the ED equilibrium depicted in Figure 2.1. If the capacity buffer is of higher order than the offered load to the power of 2/3, then the system is in the quality-driven (QD) regime. The system operates around the QD equilibrium depicted in Figure 2.1. Hence, in a spatial environment, the QED regime may only emerge if the safety capacity is of order the offered load to the power of 2/3. We furthermore establish that the QED regime can indeed be achieved. The QED regime does not correspond to a new stable operating point of the system, but to a system that oscillates stochastically between the ED and QD equilibrium points. Reaching the QED regime is more subtle in a spatial environment, as now it does not only depend on the order of the safety capacity but also on second order terms. Furthermore, as a by-product of this analysis, we can approximate the system cost and establish that the power of 2/3 scaling is optimal in the sense that it minimizes a sum of server costs and waiting costs, which is a natural social planner's objective.

We show that the approximation method used, which greatly simplifies the analysis of an otherwise highly non-tractable system, captures the fundamental features of the true system. We validate our approach via a series of numerical simulations that show that the heavy-traffic behavior of our Markovian system closely captures that of a simulated spatial multi-server system.

In sum, our model and results imply that common rules of thumb such as the SRS rule will no longer be valid for spatial operations and, therefore, new staffing rules of thumb are necessary. This has implications for how to think about such trade offs in automated warehouses and, with the advent of fleets of self-driving cars, in ride-hailing platforms. Our results derive new rules of thumb for the implications of capacity levels on the type of service regime they induce.

2.2 Related Literature

This chapter relates to several streams of literature.

Staffing. Our goal is to analyze the performance of a system with customers arriving and being served in a spatial setting as measured by the steady-state probability of waiting in heavy traffic. The seminal work of [36] introduces the so-called Halfin-Whitt regime in which the system is taken to heavy traffic by scaling the number of servers as $R + \beta \cdot \sqrt{R}$ where R is the offered load. This is also known as the square root staffing (SRS) rule. Under this regime, the authors show that in an $M/M/n$ or $GI/M/n$, the system the probability of delay is strictly between zero and one–the QED regime. [34] and [68] study the Erlang-A case. For more on the QED regime with applications to call centers, we refer the reader to the survey papers by [33] and [3]. We also refer the reader to [69] for related work, and [56] for the more general case of the $G/GI/n$ system. [13] study the capacity sizing problem in an environment in which there is also parameter uncertainty for mean arrival rate, deriving new prescriptions for such settings and articulating how to operate depending on whether one is in an uncertainty-dominated or a variability-dominated regime. Our work is complementary to this literature in the sense that we also analyze the performance of the system as measured by the probability of delay. In our model, however, the presence of spatial frictions affects dynamics and introduces state-dependencies, leading to fundamental changes in how capacity should be scaled to achieve QED performance. For an in depth discussion about limiting regimes (ranging from the conventional heavy traffic regime to the Halfin-Whitt regime and passing through the slowdown regime) and their implications for diffusion approximations in non-spatial environments we refer the reader to [67] and [8].

State-dependent service rate. The general spatial system we aim to understand is complex and generally intractable. To gain insight we consider a simpler Markovian version of it that can be regarded as an $M/M_Q/n$ system. Our work is thus related to the broad literature on Markovian system and birth and death processes, and in particular to the works that study service systems with state-dependent processing rates; for some examples we refer the reader to [45], [46] and [55]. [25] study an Erlang-R service system in which the service rate can be sped up whenever congestion is above a certain threshold . Using a fluid analysis they show that, depending on system parameters, speeding up service can lead to both desirable and undesirable system congestion levels. In related work, [30] study a service system in which agents are sensitive to individual future work load and reduce their service rate as the system's workload increases. They show that depending on load sensitivity, the system's slowdowns can take it from moderate to substantial deterioration. Our work can be considered as a combination of both speedups and slowdowns, and the exact form of these in our context is driven by *spatial economies of scale*. As the number customers in our system increases beyond n, the density of waiting customers increases and, therefore, the next idling servers can spend less time picking up customers, i.e., service rate speeds up. Similarly, when the number of customer is increasing but below n, the density of idle cars decreases and, therefore, arriving customers may experience larger pickup times, i.e., service rate is slowed down. These effects are a result of the physical nature of our system. Related to the above papers, and in particular [30], our system features some form of bi-stability in an underlying tightly related deterministic model. In contrast, however, the equilibria emerge on different scales in our setting and asymptotically in the stochastic system, these can survive jointly.

Stochastic vehicle routing. Another related stream of related work is that of dynamic routing problems. Routing is a highly complex class of problems and measuring the performance of routing algorithms is challenging. [14, 15] show that the scaling of queues in space is fundamentally different to the one when space is ignored. In particular, [14, 15] obtain a lower bound for the minimum expected total time in the system under any dispatching policy given by $\Theta\left(\lambda/(n^2(1-\rho)^2)\right)+\Theta(1)$ in heavy traffic, as the offered load converges to 1. This is a remarkable result that provides a lower bound for all dispatching policies and in turn sets a target for the optimal expected time in the system which can be used as a guideline to measure the performance of policies. Interestingly, the size of the system scales with $1/(1-\rho)^2$ and not with $1/(1-\rho)$ as happens in the standard $M/M/n$ system. Thus the fact that we are taking into account space fundamentally changes how the system scales with traffic intensity.

Ride-hailing. In the young but quickly growing literature on ride-hailing systems, customers arrive in a spatial region and a platform matches them to drivers who, in turn, take the customers to their desired destinations. For the important problem of spatial incentives in ride-hailing system we refer the reader to [12], [17], [23], [1] and [16].

Closer to this work are the studies that investigate the problem of matching to optimize certain performance metrics. Using a fluid approach in a closed queueing network, [19] study how to route empty cars in order to maximize network utility. In related work, [54] use a fluid approach to derive policies that maximize the number of matches. [11] study matching in a closed queueing network, and show that for a Scaled MaxWeight policy, the proportion of dropped demand in steady state decays exponentially fast as the number of servers in the system grows large. In a circular city framework, [31] analyze the waiting time performance of different matching mechanisms. The focus of this chapter, in contrast, is to understand how to think about capacity planning in spatial environments. Rather than optimizing over the space of dispatching policies, we anchor the analysis around a near-optimal dispatching policy.

Closest to the present setting is [23]. There, the authors also analyze inefficiencies stemming from additional workload in a spatial system, and study the possible use of surge pricing to alleviate these. This study focuses on a different question, that of capacity planning. The two papers utilize different dispatch policies. The framework in this chapter can be used to analyze the type of dispatching considered there, in which the additional capacity needed would be of order the offered load. In contrast, in this chapter, we focus on a class of provably near-optimal dispatch rules based on the vehicle routing literature mentioned above, which, as we establish, enables one to only need a safety capacity of the order the offered load to the power of 2/3.

2.3 Spatial Queueing Model

We introduce a stochastic model for spatial capacity planning within a bounded region of a plane. Our model is an $M/M_Q/n$ queueing system (in Kendall's notation M_Q stands for state-dependent service time) that captures the fundamental aspects of a spatial system that experiences arrivals and dispatches servers to attend to those arrivals.

2.3.1 Model

Motivation. We consider two models in this chapter. The first is what we call the *general system*, where spatial elements such as origin-destination pairs of customers are explicitly modeled. The second is a *Markovian system*, which is a queueing system that approximates the general system. In the Markovian system, the spatial frictions are captured in reduced form via a state-dependent service time. All of the mathematical results in the chapter establish properties of the Markovian system that can be regarded as qualitative prescriptions for the general system. Indeed, in Section 6, we use simulation to demonstrate that the Markovian system approximates the behavior of the general system quite well.

We are interested in gaining insights on the following general system. There is a central platform, customers and servers that interact in a bounded connected subset

$\mathcal{C}$ of $\mathbb{R}^2$ (the city). Customers arrive according to a Poisson process in the city at uniformly distributed locations in the city. Each customer wishes to travel from the point they arrive to some other point also drawn uniformly at random among all locations in $\mathcal{C}$. Customers are patient and will remain in the system until served.

There is a fixed number of servers in the system, and each one can serve one customer at a time at a constant velocity. A server first repositions to the arrival location of a customer, and then she transports that customer to his destination. Upon arrival to his final destination, the customer leaves the system and the server becomes idle and waits until the platform relocates her. The repositioning of servers occurs according to some state-dependent dispatching algorithm and is controlled by the platform.

Any given customer experiences a total time in the system that is composed of three components: *waiting time*, *pickup time* and *en-route time*:

$$\text{Time in the system} = \text{Waiting} + \text{Pickup} + \text{En-route}. \tag{2.1}$$

The waiting time corresponds to the time a customer spends in the system before he is assigned a server to pick him up. The pickup time represents the time it takes for the server to relocate from where she currently is to the customer's origin location. The en-route time is the time it takes to transport a customer from his origin to his destination.

The system described at a high level above is complex and intractable to analyze in its full generality, given the stochasticity of the system, the high-dimensional state-space, and the complexity of the space of possible dispatching policies.

Queueing model. In this chapter we study what we call the Markovian system, which is a simpler queueing model that still captures the spatial features of the general system. In setting up our model, we deliberately forego the complex interactions among agents that make the general system intractable, and focus on the overall physical dynamics that dictate the processing performance of the system. We further discuss our modeling assumptions in Section 2.3.2.

We focus on a model in which customers arrive to the system according to a Poisson process with rate λ, and stay until served. There is a total of n identical servers that provide service to one customer at a time in a first come, first serve fashion. We assume that the time between the assignment of a server to a customer and the end of the service is exponentially distributed with state-dependent rate $\mu(\cdot)$. Upon arrival, if a customer finds a server idle, he is immediately assigned a server; otherwise, he waits in line. This leads to an $M/M_Q/n$ queueing system. We use $Q(t)$ to denote the total number of customers in the system at time t, which includes both customers waiting and in service.

The distinctive feature of the system we analyze and what makes it depart from a traditional multi-server queue is that *servers must be repositioned to serve customers.* As a result, the total time a server spends on a single customer corresponds to pickup time plus en-route time as opposed to just en-route time—the analogue to service time in a traditional queueing system. In turn, in order to capture the overall processing performance of the general system, the key is to select an appropriate function $\mu(\cdot)$ that isolates spatial frictions through the combination of both pickup and en-route times as highlighted in Eq. (2.1).

Any sensible choice of the service rate must be such that its inverse, $1/\mu(\cdot)$, has two components: one reflecting pickup times and the other en-route times. En-route times are simple. They correspond to the distance between two random locations in $\mathcal{C}$ (properly scaled by the velocity) and do not depend on the state of the system. If we let $\bar{s}_t$ to denote the expected time to move between two random points in $\mathcal{C}$ (for some nominal velocity), then it follows that one of the components of $1/\mu(\cdot)$ will be equal to $\bar{s}_t$. The remaining component has to relate to pickup times. These are more involved as they depend on how, based on the state of the system, the platform decides to do the assignment of servers to customers—the dispatching algorithm. To overcome this difficulty it is convenient to look at the physics of the spatial system under a particular dispatching algorithm. In the present study, we anchor the analysis around the asymptotic behavior of one notable dispatching algorithm: *nearest-neighbors* dispatch (`NN`). This algorithm is simple, intuitively appealing, and it is also near-optimal.[2] If there are more servers than customers, `NN` assigns the next arriving customer to its closest available servers. If there are less servers than customers, `NN` assigns the next idling server to its closest waiting customer.

The asymptotic behavior of `NN`, which we discuss in Section 2.3.2, leads to a particular form of the expected service time which, in turn, motivates the following expression for the state-depend rate of our queueing system when its state is q

$$\frac{1}{\mu(q)} \triangleq \frac{\bar{s}_p}{\sqrt{|q-n| \vee 1}} \;+\; \bar{s}_t, \quad q \geq 0, \tag{2.2}$$

for two given positive constants $\bar{s}_p$ and $\bar{s}_t$, where $\bar{s}_p$ represents the average pickup time when there is one server available and one passenger request. The form in Eq. (2.2) captures spatial frictions in the following way. Consider the queueing system. If $Q(t) \ll n$, then $|Q(t) - n|$ is large and many servers are available, and thus, $1/\mu(Q(t))$ is close to $\bar{s}_t$, the expected en-route time. The pickup time should be negligible given the high density of free servers in space. Similarly, if $Q(t) \gg n$, then many customers are waiting and a match with a low pickup time could be found given the high density of customers in space. Indeed, we also have that $1/\mu(Q(t))$ is close to $\bar{s}_t$ in this scenario. The important point is that whenever there is a critical idle/waiting mass at either side of the market, the physical nature of the system allows it to process customers efficiently. When $Q(t) \approx n$, we expect the match between server and customer to lead to a significantly higher pickup time. In our model, a customer's total expected service time will be close to $\bar{s}_p + \bar{s}_t$ when $Q(t) \approx n$. For

[2] Among the policies that minimize customers' expected total system time, `NN` achieves near optimal performance (see e.g., [14]) .

notational simplicity, we assume $\bar{s}_t = \bar{s}_p$ throughout the next few sections, and denote this quantity simply by $\bar{s}$. When we simulate the system in Section 2.6, we allow $\bar{s}_t$ and $\bar{s}_p$ to take distinct values.

Performance Metrics. The main objective of this chapter is to understand the implications of spatial frictions on performance metrics of the service system. In particular, we analyze these in an asymptotic regime in which the number of servers and the arrival rate grow large. We analyze the system in heavy traffic and consider a sequence of $M/M_Q/n$ queues indexed by n, with arrival rate λ_n such that

$$\rho_n < 1, \quad \lim_{n\to\infty} \lambda_n = \infty, \quad \lim_{n\to\infty} (1-\rho_n)n^{\alpha} = \beta, \quad \text{for some } \beta \in \mathbb{R}_+,\ \alpha \in (0,1), \tag{2.3}$$

where ρ_n equals $\bar{s} \cdot \lambda_n/n$. Thus, ρ_n approaches 1 from below at rate $1/n^{\alpha}$. Under these different scalings (as α varies), our goal is to study key performance metrics associated with the system. We let $\{Q_n(t)\}_{t\geq 0}$ denote the number of customers in the n-system. The dynamics of $Q_n(t)$ can be written as follows. Let $A = \{A(t) : t \geq 0\}$ and $S = \{S(t) : t \geq 0\}$ be two independent unit rate Poisson processes. The path-wise construction of Q_n is

$$Q_n(t) = Q_n(0) + A(\lambda t) - S\left(\int_0^t \mu_n(Q_n(u)) \cdot \min(n, Q_n(u))du\right), \quad Q_n(0) = Q_0. \tag{2.4}$$

The term Q_0 corresponds to the initial state of the system, the second term captures the cumulative arrivals up to time t, and the third term refers to the cumulative departures up to t. In the latter, $\mu_n(Q_n(t)) \cdot \min(n, Q_n(t))$ corresponds to the service rate of the system, with $\mu_n(Q_n(t))$ representing the service rate per server at time t and $\min(n, Q_n(t))$ the number of non-idle servers at time t.

We use $Q_n(\infty)$ to denote a random variable representing the number of customers in the system in steady-state. One key central metric we are interested in quantifying is the steady-state limiting delay probability

$$P_\infty(W) \triangleq \lim_{n\to\infty} \mathbf{P}[Q_n(\infty) \geq n],$$

in order to assess the system performance. As in classical multi-server queues (see, e.g., [36]), if $P_\infty(W) = 1$, the system is said to be operating in the *efficiency-driven* (ED) regime, if $P_\infty(W) = 0$ the system is said to be operating in the *quality-driven* (QD) regime, and if $P_\infty(W) \in (0, 1)$, the system is said to be in the *quality- and efficiency-driven* (QED) regime. In the coming sections, we characterize how $P_\infty(W)$ changes as the values of α and β change. In turn, we will also analyze implications on various other metrics such as, e.g., total system cost.

2.3.2 Discussion of the Modeling Assumptions

We now provide an asymptotic grounding for Eq. (2.2), based on the NN dispatching algorithm that is studied in the vehicle routing literature ([14]). Recall that for this policy, when there are more servers than customers, the closest idle server is assigned to a new arrival (see Figure 2.2 (a)). In the case when there are more customers than servers, as soon as a server becomes idle, we assign her to the closest customer (see Figure 2.2 (b)).

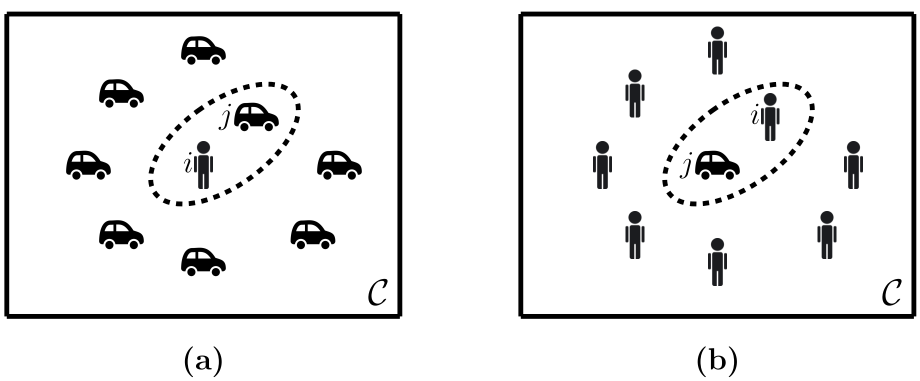


Figure 2.2: Nearest neighbor policy (NN). In (a) we have $Q(t) < n$, in (b) we have $Q(t) > n$.

The connection between $\mu(\cdot)$ and NN comes from the following argument. Consider a general system operating under NN. Suppose that at time t there is a total of $Q(t)$ customers, and that server j was matched to customer i. Depending on the state of the system, the assignment could have happened in two different ways. If $Q(t) < n$, server j must be the closest idle server to customer i among $n - Q(t)$ idle servers (see Figure 2.2 (a)). If $Q(t) \geq n$, customer i must be the closest waiting customer to server j among $Q(t) - n$ waiting customers (see Figure 2.2 (b)). In either case customer i's pickup time can be computed by comparing the distance of the closest of $|Q(t) - n| \vee 1$ random variables uniformly distributed in $\mathcal{C}$ to a single point. We can then use the following standard result from probability to obtain an asymptotic approximation for a customer's expected pickup time under `NN`.

Lemma 2.1 *Let $X_1, X_2, \ldots$ be a sequence of independent uniformly distributed random points in $\mathcal{C}$. Then, the expected minimum distance to any x_0 in the interior of $\mathcal{C}$ satisfies*

$$\mathbf{E}\left[\min_{i=1,\ldots,k} \|X_i - x_0\|\right] = \Theta\left(\frac{1}{\sqrt{k}}\right), \quad \textit{as } k \uparrow \infty.$$

Conditioning on $Q(t)$ and ignoring any dependencies among the involved random variables, Lemma 2.1 suggests the following approximation for a customer's expected pickup time

$$\mathbf{E}[\text{Pickup}|Q(t)] \approx \frac{\bar{s}_p}{\sqrt{|Q(t) - n| \vee 1}},$$

for some positive $\bar{s}_p$. The first term in Eq. (2.2) incorporates this approximation.

We note that the particular approximation we use in $\mu(\cdot)$ discussed above is not the only simplifying assumption we use in the Markovian system. We also assume that server travel times, including both pickup and en-route times, are exponentially distributed. We argue in Section 2.6 using simulation that our approximations are reasonable, in the sense that the Markovian system approximates well the behavior of the general system.

First-dispatch. Another dispatching protocol that has received attention in the literature is *first-dispatch* (`FD`). Under `FD`, an arriving customer is assigned as soon as possible to the closest idle server. Consider again Figure 2.2. In the situation depicted on the left panel (a), `NN` and `FD` operate according to the same rules. However, in the situation represented by the right panel (b) of Figure 2.2, the two dispatch rules operate quite differently. In this case, the `FD` dispatching algorithm assigns the next idling server to the longest waiting customer. As pointed out by [23], the `FD` dispatch rule can lead the system to a bad equilibrium they call the *Wild Goose Chase* in which servers spend long times picking up customers. Our framework can be used to analyze the systems' performance under the `FD` dispatch policy. Using Lemma 2.1 we can derive the following expression for an approximate service rate under `FD`:

$$\frac{1}{\mu_{\texttt{FD}}(q)} = \frac{\bar{s}_p}{\sqrt{(n-q)^+ \vee 1}} \ + \ \bar{s}_t, \quad q \geq 0.$$

Unlike the `NN` policy, the `FD` policy does not make use of *spatial economies of scale* when the system is heavily loaded with customers ($q > n$); instead, it serves customers on a first come first serve basis. This gives rise to the *Wild Goose Chase* phenomenon. Under this inefficient dispatching protocol, the number of servers required to escape ED performance equals the offered load plus a buffer term that is of the same order of the offered load, as opposed to a buffer of the order of the offered load to the power of 2/3 under `NN`. The `NN` dispatching protocol avoids this bad equilibrium outcome by exploiting spatial economies of scale even when the system is heavily loaded with customers.

2.4 Dynamics of a Related Deterministic System

Before we study the stochastic limiting properties of the Markovian system in Section 2.5, we analyze the properties of a deterministic version of it that will provide natural candidate focal points for the former system and initial insights on its behavior. In particular, we focus on a natural deterministic counterpart of Eq. (2.4).

Deterministic dynamics. Consider the dynamics of $\tilde{Q}_n(\cdot)$ described by

$$\tilde{Q}_n(t) = \tilde{Q}_n(0) + \lambda_n t - \int_0^t \mu_n\left(\tilde{Q}_n(u)\right) \cdot \min\left(n, \tilde{Q}_n(u)\right) du, \quad \tilde{Q}_n(0) = \tilde{Q}_0,$$

where $\tilde{Q}_0$ is a non-negative constant. This dynamical system has a simple interpretation. A fluid of customers joins the system at rate λ_n and departs at state-dependent rate $\mu_n(\tilde{Q}_n(t)) \cdot \min(n, \tilde{Q}_n(t))$. This dynamical system is a deterministic version of the one presented in Eq. (2.4). From the equation above, we can write $\tilde{Q}_n$ as the solution of the ordinary differential equation

$$\frac{d\tilde{Q}_n(t)}{dt} = f_n(\tilde{Q}_n(t)), \quad \tilde{Q}_n(0) = \tilde{Q}_0, \tag{2.5}$$

where

$$f_n(q) \triangleq \lambda_n - \mu_n(q) \cdot \min\{n, q\}.$$

Since $\mu_n(\cdot)$ is a Lipschitz continuous function, so is $f_n(\cdot)$. Therefore, by the Picard-Lindelof theorem, the ODE in Eq. (2.5) has a unique solution, which we denote by $\Phi(q_0, t)$ for a given $\tilde{Q}_n(0) = q_0$. In what follows, we study the equilibrium points of this solution.

Definition 2.1 (Equilibria) *We say that a point q^* is an equilibrium point of the dynamic system presented in Eq.* (2.5) *if*

$$\Phi(q^*, t) = q^*, \quad \textit{for all } t \geq 0.$$

An equilibrium point q^* is such that if the systems starts at q^*, then the systems remains at q^* for all $t \geq 0$. Observe that we can compute an equilibrium by solving $f_n(q^*) = 0$. In general, a dynamical system can have multiple equilibria but these may have different properties. We classify the equilibria according to the following definition.

Definition 2.2 (Stability of Equilibria) *An equilibrium q^* of Eq.* (2.5) *is said to be stable if for any $\epsilon > 0$, there exists $\delta > 0$ such that if $|q - q^*| < \delta$, then $|\Phi(q, t) - q^*| < \epsilon$ for all $t \geq 0$. Otherwise, q^* is unstable. If q^* is stable and there exists $\delta > 0$ such that if $|q - q^*| < \delta$, then $\lim_{t\to\infty} \Phi(q, t) = q^*$, we say that q^* is* *locally asymptotically stable. If* $\lim_{t\to\infty} \Phi(q,t) = q^*$ *for any* $q \geq 0$*, we say that* q^* *is globally asymptotically stable.*

Informally, an equilibrium q^* is *stable* if whenever the system is slightly perturbed from q^*, it remains near q^*. An equilibrium q^* is *unstable* if small perturbations of the system around q^* take the system away from q^*. If for any starting point q, the dynamic $\Phi(q,t)$ converges to q^* then q^* is *globally asymptotically stable.* If the latter is true but only in a neighborhood of q^* then q^* is *locally asymptotically stable.* Next we study the equilibria of the dynamical system from Eq. (2.5).

Equilibria characterization. Recall that the equilibrium points of Eq. (2.5) can be found by solving $f_n(q^*) = 0$. The next theorem provides a complete description of the solutions to this equation for n large.

Theorem 2.1 (Equilibrium Points) *Suppose* $\lim_{n\to\infty}(1-\rho_n)n^{\alpha} = \beta$ *and* $\rho_n \uparrow 1$*, and let* $\beta_1^* = 3/(4^{1/3})$*.*

(i) Then, there exists n_0 *such that for all* $n \geq n_0$*, the system from Eq.* (2.5) *admits an equilibrium given by*

$$\overline{q}_n = n + \frac{\rho_n^2}{(1-\rho_n)^2}.$$

Furthermore, this equilibrium is unique and globally asymptotically stable if $\alpha > 1/3$ *or if* $\alpha = 1/3$ *and* $\beta < \beta_1^*$*.*

(ii) Suppose $\alpha < 1/3$ *or* $\alpha = 1/3$ *and* $\beta > \beta_1^*$*. Then, there exists* n_0 *such that for all* $n \geq n_0$*, the system from Eq.* (2.5) *admits three equilibria given by*

$$\overline{q}_n = n + \frac{\rho_n^2}{(1-\rho_n)^2}, \tag{2.6}$$

$$\underline{q} = n - n\cdot(1-\rho_n)\cdot r_{0,n}(\rho_n), \tag{2.7}$$

$$\tilde{q}_n = n - n\cdot(1-\rho_n)\cdot r_{1,n}(\rho_n), \tag{2.8}$$

where

$$r_{i,n}(\rho_n) = \frac{4}{3}\cdot\cos\left(\frac{1}{3}\arccos\left(-\sqrt{\frac{27\rho_n^2}{4n\cdot(1-\rho_n)^3}}\right)-\frac{2\pi i}{3}\right)^2, \quad i\in\{0,1\}.$$

Furthermore, $\overline{q}_n$ and $\underline{q}$ are locally asymptotically stable and $\tilde{q}_n$ is an unstable equilibrium.

The result establishes that there are two fundamentally different regimes where the system from Eq. (2.5) can operate. When the system is heavily loaded, in the sense that $\alpha > 1/3$ or $\alpha = 1/3$ and $\beta < \beta_1^*$, then the queue length converges to a point $\overline{q}_n > n$ as t grows to ∞, independently of the initial condition. Furthermore the exact characterization of $\overline{q}_n$ provides additional insights. We have

$$\overline{q}_n \;=\; n + \frac{\rho_n^2}{(1-\rho_n)^2} \;\approx\; n + \frac{1}{\beta^2}\,n^{2\alpha}.$$

Hence, in such a system, asymptotically, there are always order $n^{2\alpha}$ customers waiting in the system to be served.

As the load decreases (α decreases) and when the system is such that $\alpha < 1/3$ or $\alpha = 1/3$ and $\beta > \beta_1^*$, then the behavior of the system is more subtle. There are two locally stable equilibria and one unstable equilibrium. Now the same equilibrium $\overline{q}_n$ still exists and is locally stable, but a new locally stable equilibrium emerges, $\underline{q}$. It is possible to show that this new equilibrium is such that[3]

$$\underline{q} \;\approx\; n - c\,n^{1-\alpha},$$

for an appropriate constant c. In other words, in such an equilibrium, there are always idle servers, and there is order $n^{1-\alpha}$ such idle servers. Hence, there are two locally stable equilibria, one with all servers busy and customers waiting ($\overline{q}_n$) and one with idle servers and no customers waiting ($\underline{q}$).

[3]This can be seeing by analyzing the Taylor expansion of the term $r_{0,n}(\rho_n)$.

Proof sketch and intuition. The proof of Theorem 2.1 relies on analyzing both equilibrium points and their stability properties. To establish the equilibria, we determine the zero crossings of $f_n(\cdot)$. With some slight rewriting,

$$
\begin{aligned}
f_n(q) &= \lambda_n - \mu_n(q) \cdot \min\{n, q\} = \lambda_n \left[1 - \left(\frac{1}{\sqrt{|q-n| \vee 1}} + 1 \right)^{-1} \cdot \frac{\min\{n, q\}}{\lambda_n \bar{s}} \right] \\
&= \lambda_n \left[1 - \frac{g_{2,n}(q)}{g_{1,n}(q)} \right],
\end{aligned}
$$

with $\quad g_{1,n}(q) = 1 + \dfrac{1}{\sqrt{|n-q| \vee 1}}, \qquad g_{2,n}(q) = \dfrac{\min(n, q)}{\lambda_n \bar{s}}.$

The function $g_{1,n}(q)$ is proportional to the amount of work a system with n servers needs to do per customer when there are q customers in the system. Analogously, $g_{2,n}(q)$ is proportional to the amount of work the system with n servers is capable of doing per customer when there are q customers in the system. Hence, determining the sign of $f_n(q)$ amounts to comparing the sizes of $g_{1,n}(q)$ and $g_{2,n}(q)$. When the former is larger than the latter, we have $f_n(q) > 0$ and the queue size grows. When the inverse is true, $f_n(q) < 0$, the queue size shrinks. When they are equal, we obtain an equilibrium point by solving for q. Figure 2.3 depicts the two functions for the two different regimes.

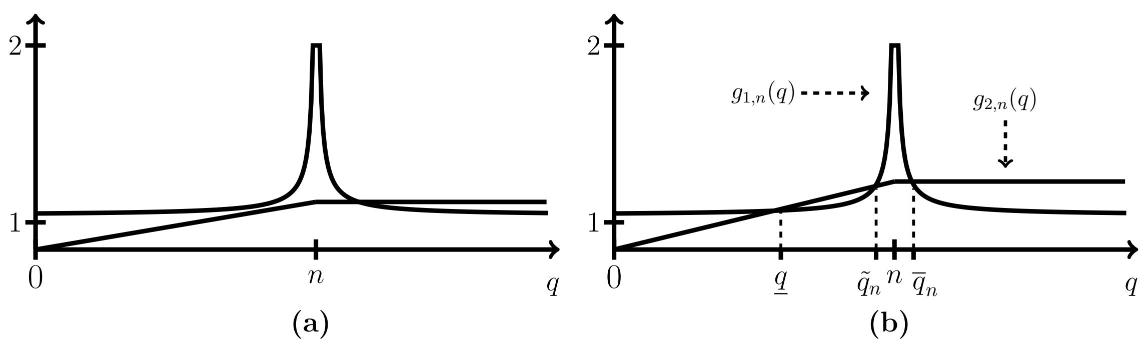


Figure 2.3: Equilibria points for system from Eq. (2.5). Plots (a) and (b) correspond to regimes (i) and (ii) from Theorem 2.1, respectively. The points where the functions $g_{1,n}(q)$ and $g_{2,n}(q)$ cross correspond to equilibria points.

As for stability, the queue length tends to grow when $g_{1,n}(q) > g_{2,n}(q)$ since the amount of work the system needs to perform per customer is greater than its ability to do work per customer. Similarly, $g_{1,n}(q) < g_{2,n}(q)$ implies the system can handle the current workload and that the queue size is decreasing. Therefore, the two equilibrium points in regime (ii) where $g_{1,n}(q) > g_{2,n}(q)$ to their left and $g_{1,n}(q) < g_{2,n}(q)$ to their right, $\underline{q}$ and $\overline{q}_n$, are stable, while $\tilde{q}_n$ is not.

An important observation is about what drives the differences between the regimes. From the heavy traffic scaling (see Eq. (2.3)) we have that $g_{2,n}(q) \approx q/(n - \beta \cdot n^{1-\alpha})$ for all $q < n$. It follows that for $q < n$ the slope of $g_{2,n}(q)$ is determined by both α and β. The theorem establishes that when α is large enough the slope of $g_{2,n}(q)$ is not steep enough to cross $g_{1,n}(q)$ and, therefore, the only possible equilibrium is $\overline{q}_n$ (See Figure 2.3 (a)). Similarly, if α is small enough then $g_{2,n}(q)$ is steep enough to cross $g_{1,n}(q)$; thus, the two extra equilibria $\underline{q}$ and $\tilde{q}_n$ emerge (See Figure 2.3 (b)). The transition point occurs when α equals 1/3. In this case, depending on the choice of β, the two extra equilibria may or may not exist. As β increases, the slope of $g_{2,n}(q)$ increases until it reaches a point from which on $g_{2,n}(q)$ is steep enough so that the two equilibria to the left of n materialize.

Interpretation in terms of the queueing system. In terms of the queuing model, when the number of customers is much larger than n, service times become shorter. In turn, the system processes customers more efficiently, which brings the total number of customers down. In addition, when the number of customers is close to n, service times are not as short as in the previous situation. This implies that the system is not as effective in processing customer, bringing the total number of customers up. That is, the queueing system (and also the general system) has a self-regulating property that is captured by the deterministic system through the equilibrium $\overline{q}_n$. When the number of customers is low (when $q < \underline{q}$), despite the fact that each customer experiences a "short" pickup time, there are just not enough customers in the system so that the arrival rate dominates departure rate, which increases the number of customers in the system. For a medium number of customers (when $q \in (\underline{q}, \tilde{q}_n)$), there are enough idle servers so that we are processing customer efficiently, but also there are enough customers in the system so that arrivals can be dominated by departures. This brings the number of customers in the system down. For a large number of customers ($q \in (\tilde{q}_n, n)$), there are not enough idle servers. Therefore, the service time of customers becomes large and, as a consequence, so does the number of customers in the system. That is, for states below n, the queueing system also has a self-regulating property that is captured by the deterministic dynamics through the equilibrium $\underline{q}$. Therefore, one might expect $\underline{q}$ and $\overline{q}_n$ to play focal roles in the queueing system, which they indeed do when we analyze the stochastic version of the system in Section 2.5.

2.5 Limiting Regimes

In this section, we first investigate the properties of the Markovian system in steady state, where the equilibria derived in the previous section for the deterministic system from Eq. (2.5) will play a central role. We then analyze the system in the asymptotic regime from Eq. (2.3), parametrized by α and β. In turn, our results lead to a parametrization of the system's regimes: QD, ED and QED. We also discuss some managerial implications of the results.

2.5.1 Steady-State Analysis

Before we provide our main results, observe that for a given scale n, the process $Q_n(t)$ is a birth and death process with birth rate λ_n and state-dependent death rate $\mu_n(Q_n(u)) \cdot \min(n, Q_n(u))$. Letting $\pi_n(k)$ be the steady-state probability that the n-system is in state k, the detailed balance equations yield

$$\pi_n(k) \cdot \frac{f_n(k)}{\lambda_n} = \pi_n(k) - \pi_n(k-1), \quad k \geq 1. \tag{2.9}$$

We first characterize the shape of the steady-state distribution $\pi_n(\cdot)$ for systems with large scale.

Proposition 2.1 (Steady-state Probability Distribution) *Suppose that* $\lim_{n\to\infty}(1-\rho_n)n^{\alpha} = \beta$, $\rho_n \uparrow 1$, *and let* $\beta_1^* = 3/(4^{1/3})$. *Then the following holds.*

(i) If $\alpha > 1/3$ *or if* $\alpha = 1/3$ *and* $\beta < \beta_1^*$, *then for* n *sufficiently large, the steady distribution* $\pi_n(\cdot)$ *is unimodular with a mode at* $\lfloor \overline{q}_n \rfloor$.

(ii) If $\alpha < 1/3$ *or if* $\alpha = 1/3$ *and* $\beta > \beta_1^*$, *then for* n *sufficiently large, the steady distribution* $\pi_n(\cdot)$ *admits two modes, one at* $\lfloor \underline{q} \rfloor$ *and one at* $\lfloor \overline{q}_n \rfloor$.

This result leverages Eq. (2.9) and the intuition obtained from Figure 2.3 to link the equilibria of the deterministic system from Eq. (2.5) with the modes of $\pi_n(k)$. From Eq. (2.9), we note that the monotonicity of $\pi_n(\cdot)$ can be determined by looking at the sign of $f_n(\cdot)$. In turn, Proposition 2.1 establishes that $\pi_n(\cdot)$ has at most two modes and that those modes are close to the equilibrium points. There is always one at $\lfloor \overline{q}_n \rfloor$, and, depending on the scaling parameters, there may or may not be another one at $\lfloor \underline{q} \rfloor$. We represent the two possibilities in Figure 2.4.

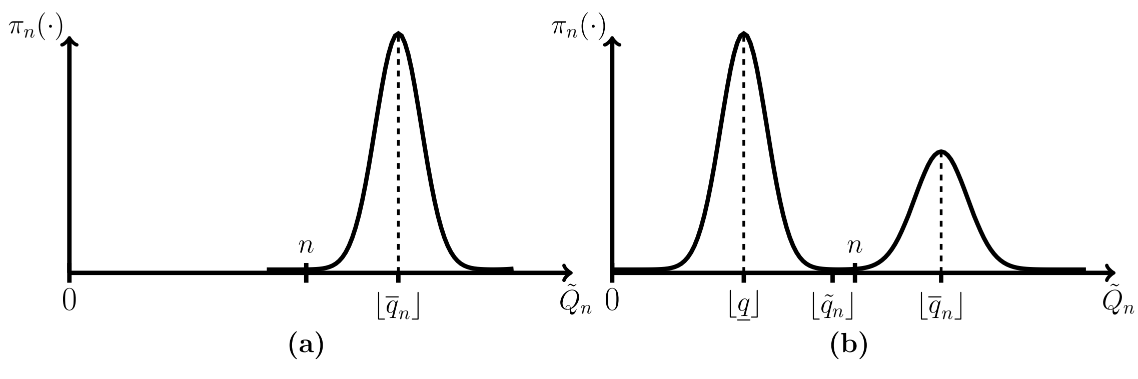


Figure 2.4: Steady-state probability $\pi_n(\cdot)$. In **(a)**, which corresponds to regime (i) in Proposition 2.1, the state distribution is unimodal with a peak at $\lfloor \overline{q}_n \rfloor$. In **(b)**, which corresponds to regime (ii) in Proposition 2.1, the state distribution is bimodal with peaks at

Whenever $\alpha > 1/3$, $\pi_n(\cdot)$ is unimodal and it peaks once to the right of n, see Figure 2.4(a). If $\alpha < 1/3$, $\pi_n(\cdot)$ is bimodal and it also peaks to the left of n, see Figure 2.4(b). If $\alpha = 1/3$ these two cases are possible depending on the parameter β. This is in line with the intuition we obtained from the deterministic analysis in Section 2.4.

In steady-state, one expects that the system spends most of the time around the modes of the distribution. However, when assessing the performance of the system in terms of probability of having to wait for a server to be assigned, one needs to analyze the steady-state distribution beyond its modes to evaluate how mass is distributed. We do so next.

2.5.2 Service Regimes

We start our analysis of service regimes by analyzing the quality-driven (QD) and effiency-driven (ED) regimes.

2.5.2.1 QD and ED regimes.

We first establish sufficient conditions for the ED and QD regimes to emerge.

Theorem 2.2 (Limiting Regimes) *Fix $\alpha \in (0,1)$ and $\beta > 0$. Suppose that $\lim_{n\to\infty} n^{\alpha}(1-\rho_n) = \beta$. Then, there exists $\beta_2^* > \beta_1^*$ such that*

(i) (ED Regime) if $\alpha \in (1/3, 1)$ or if $\alpha = 1/3$ and $\beta < \beta_2^*$, then

$$P_\infty(W) = 1,$$

(ii) (QD Regime) if $\alpha \in (0, 1/3)$ or if $\alpha = 1/3$ and $\beta > \beta_2^*$, then

$$P_\infty(W) = 0.$$

Theorem 2.2 provides a crisp characterization of the domains in which the ED and QD regimes emerge. If $\alpha \in (1/3, 1)$ or if $\alpha = 1/3$ and $\beta < \beta_2^*$, then recall from Proposition 2.1 that the steady-state probability of the number of customers in the system admits only one mode at $\lfloor \overline{q}_n \rfloor$, which is higher than n, the number of servers. Part *(i)* of Theorem 2.2 establishes that the mass is concentrated to the right of n and hence servers are almost always either en route to customers or transporting customers and almost never idle. In turn, customers, will have to wait with probability close to 1 before being assigned a server.

If $\alpha \in (0, 1/3)$ or if $\alpha = 1/3$ and $\beta > \beta_2^*$, then the the steady-state probability of the number of customers in the system admits two modes (cf. Proposition 2.1 part *(ii)*), one at $\lfloor \overline{q}_n \rfloor$ which is higher than n and one at $\lfloor \underline{q} \rfloor$ which is lower than n. Part *(ii)* of Theorem 2.2 establishes that the mass is concentrated to the left of n and hence there is almost always a fraction of servers that idle and customers almost never wait before being assigned a server. In other words, the mode to the right of n plays little role in this parameter regime.

Discussion of Capacity Planning. To further appreciate the result, recall that since $n^\alpha(1 - \rho_n) \to \beta$ we have

$$\frac{n - \lambda_n \bar{s}}{(\lambda_n \bar{s})^{1-\alpha}} \to \beta, \quad \text{that is,} \quad n \approx \lambda_n \bar{s} \ + \ \beta \cdot (\lambda_n \bar{s})^{1-\alpha}. \tag{2.10}$$

The term $\lambda_n \bar{s}$ corresponds to the standard offered load of the system as defined for standard $M/M/n$ multi-server systems. In heavy traffic, this quantity determines how the capacity of the system should be scaled with the arrival rate of customers. First, there is a nominal term, which is simply $\lambda_n \bar{s}$, that accounts for the expected amount of work requested by customers. The second term $\beta \cdot (\lambda_n \bar{s})^{1-\alpha}$ is a buffer term that accounts for stochastic variations of the system. In a classical $M/M/n$ setting, when $\alpha < 1/2$, the system is in the QD regime, when $\alpha > 1/2$, the system is in the ED regime, and when $\alpha = 1/2$ the system is in the QED regime. In contrast, in our setting when the buffer term is $\beta \cdot (\lambda_n \bar{s})^{1/2}$, the system is in the ED regime no matter the choice of β. Since our model captures spatial frictions, this result highlights that in a setting where servers need to reach customers before the start of effective service, the capacity needed to achieve QED performance is fundamentally different than in a standard setting. Moreover, spatial frictions create the need for more servers than in a standard setting for the system to operate in the QD regime. Indeed, in our model the buffer term must be $\beta \cdot (\lambda_n \bar{s})^m$ with $m \geq 2/3$. The transition between ED and QD occurs when the buffer term is $\beta \cdot (\lambda_n \bar{s})^{2/3}$, that is, the QED regime can only happen with a scaling of 2/3 which is orders of magnitude larger than the traditional SRS rule of thumb.

Proof sketch of Theorem 2.2. The proof of Theorem 2.2 consists on bounding above the terms $\mathbf{P}[Q_n(\infty) < n]$ and $\mathbf{P}[Q_n(\infty) \geq n]$, respectively, and then using asymptotic relations between the mode probabilities as established in the following result.

Proposition 2.2 *Fix $\alpha \in (0,1)$ and $\beta > 0$. Suppose that $\lim_{n\to\infty} n^\alpha(1-\rho_n) = \beta$ then*

(i)

$$\lim_{n\to\infty} \frac{1}{n^\alpha} \log\left(\frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(n)}\right) = \frac{1}{\beta},$$

(ii) if $\alpha < 1/3$ then

$$\lim_{n\to\infty} \frac{1}{n^{1-2\alpha}} \log\left(\frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)}\right) = -\frac{\beta^2}{2},$$

(iii) if $\alpha = 1/3$ then there exists a function $g(\cdot)$ such that

$$\lim_{n\to\infty} \frac{1}{n^{1/3}} \log\left(\frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)}\right) = g(\beta).$$

And there exists $\beta_2^* > \beta_1^*$ such that $g(\beta_2^*) = 0$ and if $\beta_1^* < \beta < \beta_2^*$ then $g(\beta) > 0$, whereas if $\beta > \beta_2^*$ then $g(\beta) < 0$.

Proposition 2.2 shows how the peak of the modes of $\pi_n(\cdot)$ compare to each other as n grows large. When $\alpha > 1/3$, for large n, there is only one peak given by $\lfloor \overline{q}_n \rfloor$. From part *(i)*, its steady-state probability satisfies

$$\pi_n(\lfloor \overline{q}_n \rfloor) \approx \pi_n(n) \cdot e^{n^\alpha/\beta},$$

that is, $\pi_n(\lfloor \overline{q}_n \rfloor)$ is exponentially larger than $\pi_n(n)$. Since $\pi_n(\cdot)$ is increasing to the left of $\lfloor \overline{q}_n \rfloor$ (see Proposition 2.1), this suggests that, in the limit, the number of customers in the system will be above n with high probabilty. In other words, the system will be in the ED regime.

For the case when $\alpha < 1/3$, Proposition 2.1 states that $\pi_n(\cdot)$ is bimodal and, therefore, there could be mass around both peaks. However, part *(ii)* of the proposition establishes that $\pi_n(\lfloor \underline{q} \rfloor)$ is exponentially larger than $\pi_n(\lfloor \overline{q}_n \rfloor)$,

$$\pi_n(\lfloor \underline{q} \rfloor) \approx \pi_n(\lfloor \overline{q}_n \rfloor) \cdot e^{\frac{1}{2}\beta^2 n^{1-2\alpha}}.$$

This suggests that when $\alpha > 1/3$, the tail of $\pi_n(\cdot)$ to the right of n vanishes as n becomes large. In turn, the number of customers in the system should be below n with high probability. In other words, while the distribution $\pi_n(\cdot)$, has two modes, only one mode "matters" and we expect the system to be in the QD regime.

The threshold case is $\alpha = 1/3$. In this case whether $\pi_n(\lfloor \overline{q}_n \rfloor)$ dominates $\pi_n(\lfloor \underline{q} \rfloor)$ (or vice-versa) is governed by β. When $\beta < \beta_1^*$, from Proposition 2.1, we know that $\lfloor \overline{q}_n \rfloor$ is the only mode and, therefore, $\pi_n(\lfloor \overline{q}_n \rfloor)$ dominates. If $\beta \in (\beta_1^*, \beta_2^*)$ then $\lfloor \underline{q} \rfloor$ is also a mode; however, part *(iii)* of the proposition establishes that $\pi_n(\lfloor \overline{q}_n \rfloor)$ is exponentially larger than $\pi_n(\lfloor \underline{q} \rfloor)$. That is, in this case $\lfloor \underline{q} \rfloor$ transitions into becoming a mode, but the mass it contributes is not large enough and it vanishes as n increases. Therefore, for $\beta < \beta_2^*$, the system will be in the ED regime. In contrast, when $\beta > \beta_2^*$, the roles of $\pi_n(\lfloor \overline{q}_n \rfloor)$ and $\pi_n(\lfloor \underline{q} \rfloor)$ reverse. This indicates that for $\beta > \beta_2^*$, the system will be in the QD regime.

2.5.2.2 QED regime

Theorem 2.2 implies that the QED regime, in which the asymptotic probability that customers have to wait for a server to be assigned is such that $P_\infty(W) \in (0, 1)$, may only occur if $\alpha = 1/3$ and $\beta = \beta_2^*$ as for all other values, the system is either in the ED or QD regimes. It is already apparent that the QED regime is much more subtle in our Markovian system than in classical $M/M/n$ systems as both the buffer order of magnitude (determined by α) *and* the constant in front of the buffer size (determined by β) need to be pinned down. The transition from QD to ED regimes does not occur through the constants in front of the buffer order of magnitude, leaving the question open of whether the QED regime exists at all in our Markovian system and, if so, how may it be reached. The next result establishes that there exists a QED regime and provides a characterization of it.

Theorem 2.3 (QED Regime) *Let $p_H \in (0,1)$. There exists a sequence $\{\gamma_n : n \geq 1\}$ with $\gamma_n \to 0$ as $n \uparrow \infty$ and a function $p_L(p_H) \in (0,1)$, such that if $n^{1/3}(1-\rho_n) = \beta_2^* + \gamma_n$ then*

$$p_L(p_H) \;\leq\; \liminf_{n\to\infty} \mathbf{P}[Q_n(\infty) \geq n] \leq \limsup_{n\to\infty} \mathbf{P}[Q_n(\infty) \geq n] \;\leq\; p_H,$$

with $p_L(\cdot)$ strictly increasing in p_H and such that

$$\lim_{p_H\to 1} p_L(p_H) = 1 \quad \text{and} \quad \lim_{p_H\to 0} p_L(p_H) = 0.$$

This result establishes a regime such that for n large enough the probability of waiting to be assigned a server is in $(0,1)$. In turn, the probability of not waiting also belongs to $(0,1)$. That is, the system is in the QED regime. We have not pinned down an exact expression for these probabilities but, instead, we have provided a range. As one varies $p_H \in (0,1)$, one can achieve the extreme regimes. If $p_H \approx 1$ then from the theorem we can deduce that $\mathbf{P}[Q_n(\infty) \geq n] \approx 1$; if $p_H \approx 0$ then we can deduce that $\mathbf{P}[Q_n(\infty) \geq n] \approx 0$.

Capacity Planning for the QED Regime. From a practical perspective, Theorem 2.3 provides two important insights. First, it shows that QED performance is achieved at a different scaling than in traditional multi-server systems. Typically, in those system a SRS rule can balance the trade-off between waiting times and service efficiency. In a spatial setting this is no longer enough because servers must reach their customers before starting service. Our results suggest that the right scaling is 2/3 instead of 1/2. Second, notice that since $n^{1/3}(1-\rho_n) - \gamma_n \to \beta_2^*$ we have

$$\frac{n - \lambda_n \bar{s}}{(\lambda_n \bar{s})^{1-\alpha}} - \gamma_n \to \beta_2^*, \quad \text{that is,} \quad n \approx \lambda_n \bar{s} \ + \ \beta_2^* \cdot (\lambda_n \bar{s})^{2/3} + \ \gamma_n \cdot (\lambda_n \bar{s})^{2/3}. \quad (2.11)$$

From this equation we observe that, in addition to the traditional buffer term of the form $\beta \cdot (\lambda_n \bar{s})^m$, our result establishes that an extra lower order term is needed for QED performance. In particular, in our Markovian system, it is necessary to add the term $\gamma_n \cdot (\lambda_n \bar{s})^{2/3}$. Because $\gamma_n \to 0$ this term is of lower order than the second term in Eq. (2.11). Hence, the QED regime requires a very fine balance involving second order terms compared to the buffer size in this spatial setting, in stark contrast with the classical $M/M/n$ setting.

Proof sketch of Theorem 2.3. A necessary condition to achieve the QED regime is that the peaks of $\pi_n(\cdot)$ be in a constant proportion; otherwise, one would dominate the other and the system would be in the QD or ED regime. According to Proposition 2.2 part *(iii)*, this can only happen when $\alpha = 1/3$ and $\beta = \beta_2^*$. In this case

$$\lim_{n\to\infty} \frac{1}{n^{1/3}} \log\left(\frac{\pi_n(\lfloor \bar{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)}\right) = 0,$$

that this, the $\log(\cdot)$ term is $o(n^{1/3})$. In turn, the ratio $\pi_n(\lfloor \bar{q}_n \rfloor)/\pi_n(\lfloor \underline{q} \rfloor)$ does not necessarily converge to a constant. To have it so, one would have to look at lower order terms for $\log(\pi_n(\lfloor \bar{q}_n \rfloor)/\pi_n(\lfloor \underline{q} \rfloor))$ and try to disentangle the exact rate at which $n^{1/3}(1-\rho_n)$ has to approach β_2^* so that the $\log(\cdot)$ converges to a constant. Instead of pursuing this, in the next result we show the existence of a sequence converging to zero, $\{\gamma_n^c : n \geq 1\}$, such that if $n^{1/3}(1-\rho_n)$ approaches β_2^* as $\beta_2^* + \gamma_n^c$, the peaks of $\pi_n(\cdot)$ will be in a constant proportion.

Proposition 2.3 *Fix $c \in \mathbb{R}$. Then, there exists a sequence $\{\gamma_n^c : n \geq 1\}$ with $\gamma_n^c \to 0$ such that if $n^{1/3}(1-\rho_n) = \beta_2^* + \gamma_n^c$, then*

$$\lim_{n\to\infty} \log\left(\frac{\pi_n(\lfloor \bar{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)}\right) = c.$$

In the proof of the proposition we provide a detailed explanation of how to construct the sequence $\{\gamma_n^c : n \geq 1\}$. In turn, the proposition is not just an existence result, but it also provides the exact sequence that enables us to maintain the peaks in a constant proportion. It also establishes that, for any constant $c \in \mathbb{R}$, if $n^{1/3}(1-\rho_n)$ approaches β_2^* at an appropriate rate then

$$\pi_n(\lfloor \overline{q}_n \rfloor) \approx \pi_n(\lfloor \underline{q} \rfloor) \cdot e^c.$$

In particular, as we vary c we can achieve any desired proportion. For example, if $c < 0$ then $\pi_n(\cdot)$ might look as depicted in Figure 2.4(b).

Even though there is a way to scale the system such that that the peaks are in constant proportion, this does not guarantee that the probability of being around each of them will be positive at the same time. It is possible, for example, that the dispersion of $\pi_n(\cdot)$ around $\lfloor \overline{q}_n \rfloor$ diminishes to zero while the proportion with the other peak remains constant. Therefore, we need to assess how the peaks compare to the mass around them. The next lemma provides a characterization of this.

Lemma 2.2 *Fix $\alpha \in (0,1)$ and $\beta > 0$. Suppose that $\lim_{n\to\infty} n^{\alpha}(1-\rho_n) = \beta$, then*

(i)

$$\frac{\mathbf{P}[Q_n(\infty) \geq n]}{\pi_n(\lfloor \overline{q}_n \rfloor)} = \Theta(n^{\frac{3}{2}\alpha}).$$

(ii) if $\alpha \in (0, 1/3)$, or $\alpha = 1/3$ and $\beta > \beta_1^*$, then

$$\frac{\mathbf{P}[Q_n(\infty) < n]}{\pi_n(\lfloor \underline{q} \rfloor)} = \Theta(\sqrt{n}).$$

This result establishes that the ratio of the mass to the right of n, to the peak in that region is $\Theta(n^{\frac{3}{2}\alpha})$. That is, with respect to $\pi_n(\lfloor \overline{q}_n \rfloor)$ the mass to the right of n is not negligible and, in fact, is approximately $n^{\frac{3}{2}\alpha}$ larger than $\pi_n(\lfloor \overline{q}_n \rfloor)$. Similarly, with respect to $\pi_n(\lfloor \underline{q} \rfloor)$, the mass to the left of n is non-trivial and, in fact, is approximately $\sqrt{n}$ larger than $\pi_n(\lfloor \underline{q} \rfloor)$.

Observe that in part *(i)* of the lemma, the order of the ratio depends on α. When $\alpha < 1/3$ then this ratio is not as big as the one for $\pi_n(\lfloor \underline{q} \rfloor)$ (which is $\Theta(\sqrt{n})$). This coincides with Theorem 2.2 in that for these values of α the mass to the left of n dominates the mass to the its right. Similarly, when $\alpha > 1/3$, the mass to the right of n dominates. For $\alpha = 1/3$, both ratios are of the same order. In turn, we have

$$\frac{\mathbf{P}[Q_n(\infty) \geq n]}{\mathbf{P}[Q_n(\infty) < n]} = \Theta\left(\frac{\pi_n(\lfloor \bar{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)}\right).$$

Therefore, if the ratio of the peaks is constant, then the total mass to the left *and* to the right of n can be both (asymptotically) positive and separated away from zero. That is, both sides can be "balanced" whenever the peaks are in constant proportion. We can thus combine the results in Proposition 2.3 and Lemma 2.2 to find lower and upper bounds for $\mathbf{P}[Q_n(\infty) \geq n]$. In the proof of Lemma 2.2 we find exact expressions to control for the ratios as n increases, which we then leverage to provide explicit bounds for $\mathbf{P}[Q_n(\infty) \geq n]$ that can be mapped to probability values, p_H and $p_L(p_H)$, which satisfy the properties of Theorem 2.3.

2.5.3 Orders of Magnitudes of Queues and Wait Times

The results so far provide an understanding of the different regimes the system can operate in as a function of its load. Next, we quantify queue sizes and waiting times in our system as a function the scaling parameter α. The discussion in this section underlines the differences of a spatial server system with a traditional queueing system.

Let L^s and W^s denote respectively the steady-state expected queue length (excluding customers in service) and expected wait time. Similarly, let L^c and W^c denote the corresponding quantities in the classical $M/M/n$ system. From standard queueing theory, we have that

$$L^c = \frac{\rho_n}{(1-\rho_n)} \cdot C(n, \lambda_n/\bar{s}_t),$$

where $C(n, \lambda_n/\bar{s}_t)$ satisfies the Erlang's C formula, and represents the probability of waiting (see, e.g., [4]). Assuming that $n^\alpha(1-\rho_n) \to \beta$ we have that

$$C(n, \lambda_n/\bar{s}_t) \to \begin{cases} 1 & \text{if } \alpha > 1/2, \\ \text{constant} & \text{if } \alpha = 1/2, \\ 0 & \text{if } \alpha < 1/2. \end{cases}$$

In turn, using standard arguments, one can show that for $\alpha < 1/2$, we have that L^c is $o(1)$. Meanwhile, for $\alpha \geq 1/2$, L^c is $\Theta(n^\alpha)$. This implies that for $\alpha < 1/2$, W^c is $o(1)$, while for $\alpha \geq 1/2$, W^c is $\Theta(n^{\alpha-1})$. In particular, in the Halfin-Whitt regime ($\alpha = 1/2$), we have that L^c is $\Theta(\sqrt{n})$ and W^c is $\Theta(1/\sqrt{n})$. Next, we compare these classic results with the results obtained from our Markovian system.

We first provide a rigorous statement about the order of magnitude of the size of our Markovian system around the equilibria, in the sense of deriving the subset of the real line where the queue lengths fluctuations are constrained to, assuming n is sufficiently large. We use this result to provide approximate expressions for L^s and W^s.

Proposition 2.4 *Suppose* $\lim_{n\to\infty} n^\alpha(1-\rho_n) = \beta$. *Then,*

(i) If $\alpha \in (1/3, 1)$ *or if* $\alpha = 1/3$ *and* $\beta < \beta_2^*$ *then there exists* $C > 0$ *such that*

$$\lim_{n\to\infty} \mathbf{P}\Big[-C \;\leq\; \frac{Q_n(\infty) - \lfloor \bar{q}_n \rfloor}{\sqrt{\log(n)} \cdot n^{\frac{3}{2}\alpha}} \;\leq\; C \Big] = 1.$$

(ii) If $\alpha \in (0, 1/3)$ *or if* $\alpha = 1/3$ *and* $\beta > \beta_2^*$ *then there exists* $C > 0$ *such that*

$$\lim_{n\to\infty} \mathbf{P}\Big[-C \;\leq\; \frac{Q_n(\infty) - \lfloor \underline{q} \rfloor}{\sqrt{\log(n)} \cdot \sqrt{n}} \;\leq\; C \Big] = 1.$$

Let's consider first part *(i)* of the proposition. In this case we can use Eq. (2.6) to deduce that

$$L^s \approx r^2 \frac{\rho_n^2}{(1-\rho_n)^2} \;\pm\; C \cdot n^{\frac{3}{2}\alpha} \cdot \sqrt{\log(n)} = \Theta(n^{2\alpha}),$$

Little's law delivers

$$W^s \approx r^2 \frac{\rho^2}{\lambda_n (1-\rho)^2} \pm \frac{C}{\lambda_n} \cdot n^{\frac{3}{2}\alpha} \cdot \sqrt{\log(n)} = \Theta(n^{2\alpha-1}).$$

There are several interesting observations. First, for $\alpha = 1/2$, the queue size is approximately $\Theta(n)$ and the wait time is approximately $\Theta(1)$. Note the contrast to a classical $M/M/n$ system, where $L^c = \Theta(\sqrt{n})$ and $W^c = \Theta(1/\sqrt{n})$. This makes precise how much more work we are adding to the system by including pickups. It also highlights that for $\alpha = 1/2$, the Markovian system is in the ED regime, with its long queues. Second, note that $\alpha = 1/2$ is the largest value for which W^s does not explode. In contrast, in the $M/M/n$ system, for any $\alpha \in (1/2, 1)$, the expected waiting time approaches zero.

If we focus on pickup times, we can gain further intuition about how the QED regime works in our system. Let P^s denote the expected pickup time. Then, from part *(i)* of the proposition,

$$P^s \approx \frac{\bar{s}}{\sqrt{|Q_n(\infty) - n| \vee 1}} \approx \Theta(1/n^{\alpha}).$$

For $\alpha = 1/3$, pickup times are of order $1/n^{\alpha}$ and W^s is of order $n^{2\alpha-1}$. This showcases the interplay between wait times and pickup times. When the load of the system increases (as measured by α), wait times increase because of the greater number of customers in the system, while pickup times decrease due to the increased spatial density of customers. If one attempts to minimize expected customer system times, we therefore need to balance W^s and P^s. For the regime where $\alpha \geq 1/3$, this occurs at $\alpha = 1/3$.

For the regime from part *(ii)* of the proposition, we have that $L^s \approx 0$ and $W^s \approx 0$. Moreover, we can use the fact that $\underline{q} \approx n - \Theta(n^{1-\alpha})$ to deduce that the expected number of idle server is $\Theta(n^{1-\alpha})$ and $P^s \approx \Theta(1/(n^{\frac{1-\alpha}{2}}))$. As we increase the load in the system (as measured by α), we reduce the number of idle servers. However, at the same time, pickup times increase due to the decreased spatial density of servers.

2.5.4 A Social Planner's Perspective

An alternative approach to determining the proper safety staffing level is to start from a social planner's objective, and then find the staffing level that optimizes it. A natural social planner's objective is one that incurs a cost per server of building capacity plus a waiting (and pick-up time) cost per customer. We now show that this objective function also leads us to the conclusion that a safety staffing that is equal to the offered load to the power of 2/3 is optimal.

Let us consider a service provider that pays c_s per unit of capacity and customers that incur a waiting cost of c_w per unit of waiting. That is, a social planner would like to select the level of capacity n that solves the following optimization problem

$$\min_{n} \quad c_s \cdot n + \lambda \cdot c_w \cdot \mathbf{E}[P_n + W_n]. \tag{2.12}$$

The first term in Eq. (2.12) corresponds to the cost of having n servers in the system. The second, to the cost experienced by customers while they wait to be assigned a server, W_n, and to be picked up, P_n.

Notice that from Eq. (2.3) we can write n as $\lambda \cdot \bar{s} + \beta(\lambda \cdot \bar{s})^{1-\alpha}$. Now, depending on our choice of α we can have one of two cases. When $\alpha \geq 1/3$, the average pick up times are of order $\Theta((\lambda \cdot \bar{s})^{-\alpha})$ while average waiting times are of order $\Theta((\lambda \cdot \bar{s})^{2\alpha-1})$. Replacing this in Eq. (2.12) delivers the following expression for the objective

$$c_s \cdot (\lambda \cdot \bar{s} + \beta(\lambda \cdot \bar{s})^{1-\alpha}) + c_w \cdot ((\lambda \cdot \bar{s})^{1-\alpha} + (\lambda \cdot \bar{s})^{2\alpha}).$$

Among all values $\alpha > 1/3$ the term that dominates in the expression above is the total waiting times, that is, $(\lambda \cdot \bar{s})^{2\alpha}$. This is increasing in α. Hence, $\alpha = 1/3$ leads to lower (asymptotic) costs compared to all values of $\alpha > 1/3$.

For the case $\alpha \leq 1/3$, let $\underline{\pi}_\lambda$ be the steady state probability the number of customers being below n and let $\overline{\pi}_\lambda$ be $1 - \underline{\pi}_\lambda$. Similar to the case when $\alpha > 1/3$, we can rewrite the objective in Eq. (2.12) to obtain

$$c_s \cdot (\lambda \cdot \bar{s} + \beta(\lambda \cdot \bar{s})^{1-\alpha}) + c_w \cdot \Big(\{\underline{\pi}_\lambda \cdot (\lambda \cdot \bar{s})^{\frac{1+\alpha}{2}} + \overline{\pi}_\lambda \cdot (\lambda \cdot \bar{s})^{1-\alpha}\} + \{\underline{\pi}_\lambda \cdot 0 + \overline{\pi}_\lambda (\lambda \cdot \bar{s})^{2\alpha}\} \Big).$$

When $\alpha < 1/3$ the term that dominates is of order $(\lambda \cdot \bar{s})^{1-\alpha}$. This term is decreasing in α. In this case, $\alpha = 1/3$ leads to lower (asymptotic) costs compared to all values of $\alpha \leq 1/3$.

In conclusion, in a large system, the system total social cost measured by capacity cost and waiting cost will be minimized by selecting the number of servers n according to $\lambda \cdot \bar{s} + \beta(\lambda \cdot \bar{s})^{2/3}$, where β should be tuned.

2.6 Numerical Experiments and General Simulation

In this section, we aim at (i) illustrating the results in the Markovian system (§2.6.1), and also to (ii) compare the behavior obtained in the Markovian system to that of the actual physical system that motivated the Markovian system (§2.6.2).

Simulation setup. We consider a square city $\mathcal{C} = [0, 2] \times [0, 2]$ and assume $v = 1$, implying that $\bar{s}_t \cdot v \approx 1.0428$. The time horizon will be $T = 4,000$. We simulate the general system introduced in Section 3.3 and the Markovian system under several different conditions, starting from $Q_n(0) = 0$, in order to capture the ED, QD and QED regimes. We scale the number of servers in the system according to

$$n = \lceil \lambda \bar{s}_t \ + \ \beta \cdot (\lambda \bar{s}_t)^{1-\alpha} \rceil. \tag{2.13}$$

For $\alpha \in \{1/4, 1/2\}$, we consider $\beta = 2.1$. For $\alpha = 1/3$, we vary $\beta \in \{2.1, 2.4, 2.7\}$.

2.6.1 Markovian System

We begin by numerically illustrating our theoretical results for the Markovian system. We consider the rate

$$\frac{1}{\mu(q)} = \frac{\bar{s}_p}{\sqrt{|q - n| \vee 1}} \ + \ \bar{s}_t, \quad q \geq 0, \tag{2.14}$$

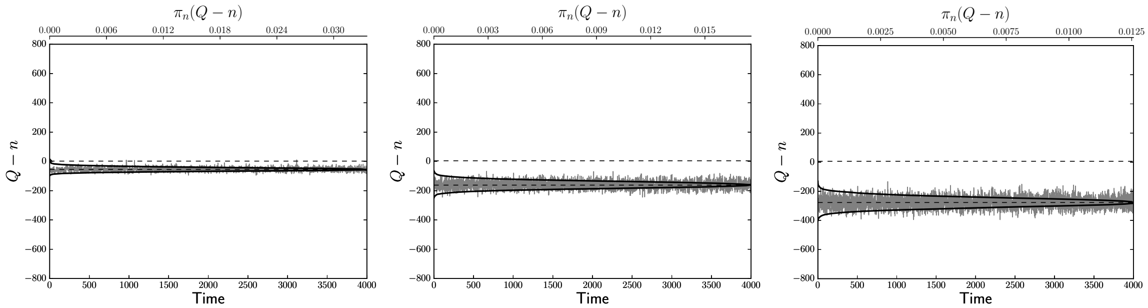
with $\bar{s}_p = \bar{s}_t = 1.0428$, that is, the coefficient in front of the pickup times coincides with the expected travel time between two points. Recall from §2.3.1 that these two parameters need not to be the same because $\bar{s}_p$ comes from an asymptotic approximation. In the next section we consider more realistic values for $\bar{s}_p$ that we estimate from simulation.

In Figures 2.5-2.6, we depict sample paths of the the number of customers in the system minus the number of servers for the various parameters and superimpose a corresponding histogram (taken from the path between periods 500 and 4,000). Furthermore, the two modes $\lfloor \overline{q}_n \rfloor$ and $\lfloor \underline{q} \rfloor$ (when they exist) minus n are depicted.

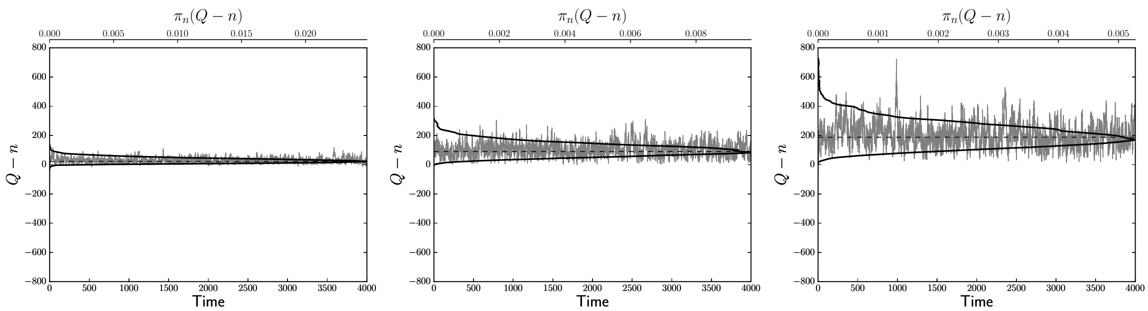
In Figure 2.5(a), $\alpha = 0.25$ and we depict the system for three different scales. In line with Theorem 2.2, one observes that the system spends almost all its time around $\lfloor \underline{q} \rfloor$ and as the scale increases, the probability of wait approaches zero. The system is in the QD regime.

In Figure 2.5(b), $\alpha = 0.5$ and we depict the system for three different scales. Note that in this case, there is only one mode, $\lfloor \overline{q}_n \rfloor$. In line with Theorem 2.2, one observes that the system spends almost all its time around $\lfloor \overline{q}_n \rfloor$ and as the scale increases, the probability of wait approaches 1. The system is in the ED regime.

In Figure 2.6, $\alpha = 1/3$ and we depict the system for three values of β. This is the only setting where, asymptotically and depending on β, the system can oscillate between the two equilibria and asymptotically, a positive fraction of the customers (separated from 0 and 1) will wait before being assigned a server. Indeed, we observe that for small values of β, the system operates most often with $Q > n$, as in the ED regime. As β increases (center plot), the fraction of time the system spends in states such that $Q < n$ increases, in which case, the system is in the QED regime. When β increases further, the system enters the QD regime.



(a) $\alpha = 0.25$



(b) $\alpha = 0.5$

Figure 2.5: Simulation of the Markovian system. We consider $\beta = 2.1$ and from left to right $\lambda \in \{100, 400, 800\}$. The bottom $x-$axis corresponds to the simulation time, while the top $x-$axis corresponds to probabilities. In the figure we observe both a sample path and $\pi_n(\cdot)$. The dashed lines correspond to the modes $\lfloor \underline{q} \rfloor$ and $\lfloor \overline{q}_n \rfloor$ as given by Theorem 2.1.

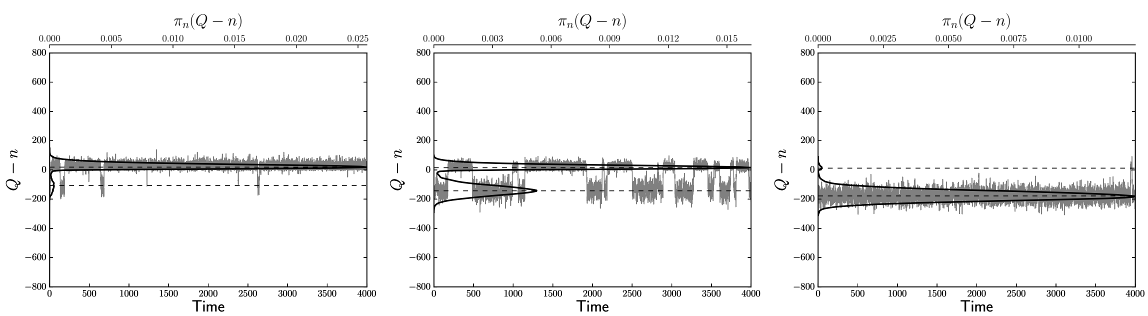


Figure 2.6: Simulation of the Markovian system. We consider $\alpha = 1/3$ and $\lambda = 800$ and from left to right $\beta \in \{2.1, 2.4, 2.7\}$. The bottom $x-$axis corresponds to the simulation time, while the top $x-$axis corresponds to probabilities. In the figure we observe both a sample path and $\pi_n(\cdot)$. The dashed lines correspond to the modes $\lfloor \underline{q} \rfloor$ and $\lfloor \overline{q}_n \rfloor$ as given by Theorem 2.1.

2.6.2 Comparing the General and Markovian Systems

Next we simulate the general system and compare it the Markovian system. Our purpose in this section is two-fold. First, we illustrate the system's behavior under the different scalings. In particular, we test whether for $\alpha < 1/3$ and $\alpha > 1/3$ the general system oscillates around the equilibria to the left and right of n, respectively. For $\alpha = 1/3$, we also test how by varying β the general system can, as predicted by the Markovian system, oscillate around both equilibria.

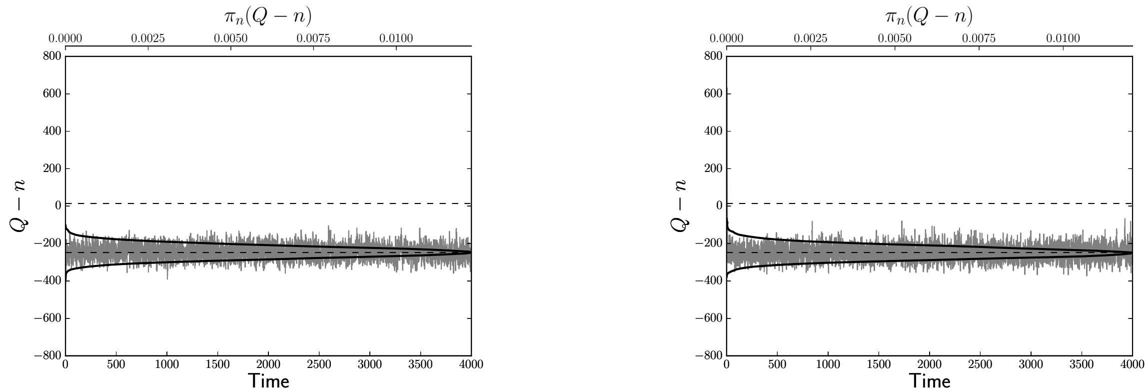
Second, we provide numerical evidence for the quality of the Markovian system as an approximation to the general system. To ensure an appropriate comparison, we proceed as follows:

- We fix λ, α and β, and use Eq. (2.13) to obtain the number of servers.
- We simulate the general system for the computed value of n.
- We estimate $\bar{s}_p$, see Eq. (2.14). Then we simulate the Markovian system with rate given by Eq. (2.14), and compute the theoretical modes/equilibria.
- We compare the system behavior for both the Markovian and general systems.

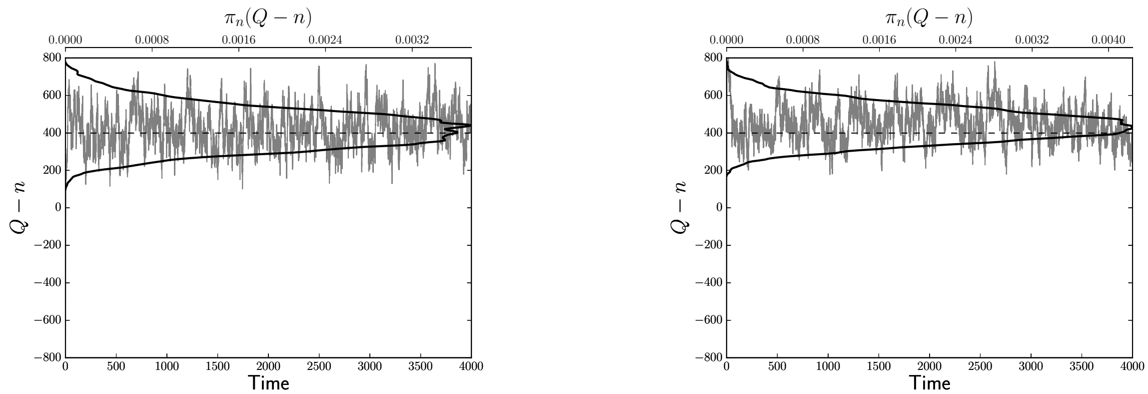
In Figures 2.7-2.8, we depict sample paths of the queue lengths in the general system (right column) and compare it to the Markovian system (left column). For the sake of exposition we fix $\lambda = 800$ throughout, but all the simulations are consistent for large values of λ.

We observe that for low α ($\alpha = 0.25$, Figure 2.7(a)), the general system queue admits a behavior very similar to the proposed Markovian approximation. In particular, the general system also admits a mode exactly around $\lfloor \underline{q} \rfloor$ (as predicted by the theory for the Markovian system) and this behavior is consistent across different scales.

For high α ($\alpha = 0.5$, Figure 2.7(b)), the general system queue admits again a behavior very similar to the proposed Markovian approximation. Again, the general system also admits a mode exactly around $\lfloor \overline{q}_n \rfloor$ (as predicted by the theory for the Markovian system).



(a) $\alpha = 0.25$, $n = 1161$, $\bar{s}_p = 1.531$, $\lfloor \underline{q} \rfloor - n = -248$, $\lfloor \overline{q}_n \rfloor - n = 14$

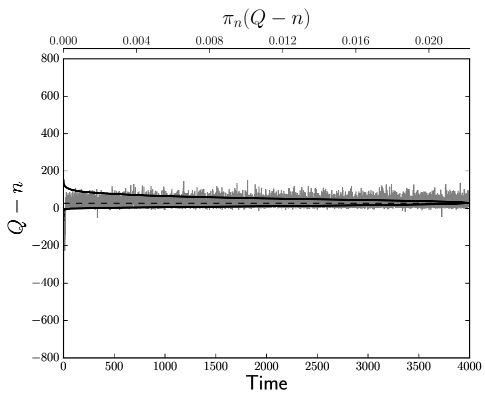


(b) $\alpha = 0.5$, $n = 895$, $\bar{s}_p = 1.518$, $\lfloor \overline{q}_n \rfloor - n = 399$

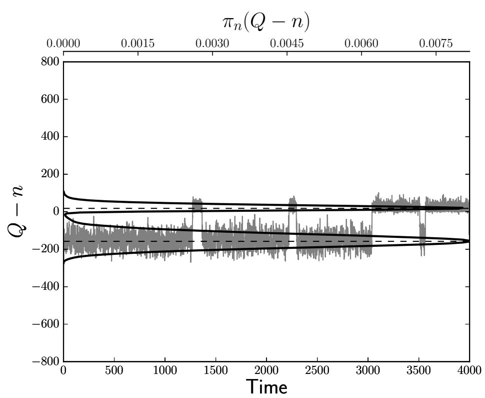
Figure 2.7: Simulation for Markovian (left) and General (right) systems. We consider $\beta = 2.1$. The bottom $x-$axis corresponds to the simulation time, while the top $x-$axis corresponds to probabilities. In the figure we observe both a sample path and $\pi_n(\cdot)$. The dashed lines correspond to the modes $\lfloor \underline{q} \rfloor$ and $\lfloor \overline{q}_n \rfloor$ as given by Theorem 2.1.

For the critical value of α ($\alpha = 1/3$, Figures 2.7(a) and 2.8(b)), the general system queue admits again a behavior very similar to the proposed Markovian approximation. For low values of β (Figure 2.8(a)), both systems operate in the ED regime. As β increases (Figure 2.8(b)), both systems move into the QED regime, as the queue oscillates between the two equilibria.

Across values of α and β and across scales, this simulation highlights the usefulness of the Markovian system in capturing some of the key features and predicting some of the behavior of the general system.



(a) $\alpha = 1/3$, $\beta = 2.1$, $n = 1021$, $\bar{s}_p = 1.256$, $\lfloor \bar{q}_n \rfloor - n = 28$



(b) $\alpha = 1/3$, $\beta = 2.7$, $n = 1074$, $\bar{s}_p = 1.286$, $\lfloor \underline{q} \rfloor - n = -158$, $\lfloor \bar{q}_n \rfloor - n = 18$

Figure 2.8: Simulation for Markovian (left) and General (right) systems. We consider $\alpha = 1/3$. The bottom $x-$axis corresponds to the simulation time, while the top $x-$axis corresponds to probabilities. In the figure we observe both a sample path and $\pi_n(\cdot)$. The dashed lines correspond to the modes $\lfloor \underline{q} \rfloor$ and $\lfloor \bar{q}_n \rfloor$ as given by Theorem 2.1.

2.7 Conclusion

In the present chapter, we have proposed a framework for studying how spatial frictions affect capacity planning. In particular, we propose a reduced-form Markovian system that captures spatial economies of scale, leading to a crisp characterization of the trade-offs at play in such environments.

We have established a mapping from load to types of regimes in heavy traffic. In particular, recalling Eq.(2.3), we have focused on regimes parametrized by α and β, where

$$\lim_{n\to\infty}(1-\rho_n)n^{\alpha}=\beta, \qquad \text{for some } \beta\in\mathbb{R}_+,\ \alpha\in(0,1).$$

Figure 2.9 summarizes some of the main findings. The ED regime emerges whenever

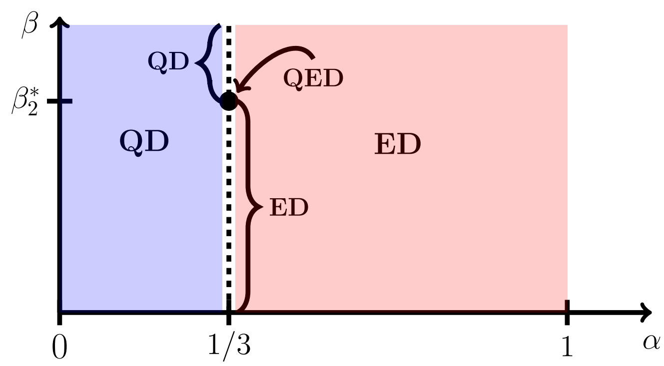


Figure 2.9: Regimes for different values of α and β.

$\alpha > 1/3$ and the QD regime emerges whenever $\alpha < 1/3$. When $\alpha = 1/3$, the three regimes QD, ED and QED can emerge and the latter can only emerge for one critical value of β, which we label β_2^*. We have further demonstrated through simulations that the Markovian approximation provides a reliable guideline for the behavior of a general system.

This chapter opens up various avenues of potential research, from both methodological and modeling perspectives. Analyzing the case when customers are impatient and might abandon the system if not served after some time is a natural extension. On the one hand, abandonment decreases the workload of the system as fewer customer have to be processed; on the other hand, it increases the system's workload as having fewer customers implies that *spatial economies of scale* become less advantageous. The important question in this case is whether, in order to achieve QED performance, abandonment necessitates just a change in β or more fundamental change in α. Another interesting extension is to study how the results in this study can be generalized to cases where origin-destination demand patterns generate imbalances in the system. In this case, the additional workload stemming from pickups might be even larger. How would this impact capacity sizing? An additional important practical question is to consider time-varying demand patterns that might require alternatives to steady-state analysis.

From a methodological perspective, an interesting extension would be to establish some of form of convergence of the processes in the general system to those in the Markovian approximation. More generally, there is potential to generalize the main result of this chapter to any near optimal dispatching protocol by directly studying the spatial system. A simple back-of-the-envelope calculation serves to enlighten the latter claim. From [15] we can deduce that the expected number of customers in the system in steady state is bound below by

$$\frac{n}{2} - n \cdot (1 - \rho_n) + C \cdot \frac{\rho_n^2}{(1 - \rho_n)^2}.$$

The second term in this expression represents the number of idle server in the system, $n - n\rho_n$; while the third term maps to the number of customers waiting or being picked up. These two terms are opposing forces that push the system to have less and more customers, respectively. Using the heavy traffic scaling from Eq. (2.3), we can deduce that the second term scales as $\Theta(n^{1-\alpha})$, while the third term as $\Theta(n^{2\alpha})$. Observe that these scalings relate to those of the equilibria in Theorem 2.1. Intuitively, quality and efficiency should balance when the two opposing forces balance each other. This occurs when $1 - \alpha$ equals 2α or, equivalently, when α equals $1/3$, as our results prescribe. Note that this derivation does not rely on a specific dispatching protocol, but only on one that is optimal or "near optimal" compared to the lower bound. Deriving the 2/3-scaling result at a full level of generality is an exciting direction for future work.

Chapter 3

The Scope of Sequential Screening With Ex-Post Participation Constraints

3.1 Motivation and Overview of Results

Sequential screening models have been used extensively in economics and revenue management to study optimal contract design when buyers learn their valuations over time. In the classic formulation of sequential screening pioneered by [27], a profit-maximizing seller faces buyers that have initial partial-private information about their valuation, for example the mean, and privately learn their full valuation after some time. In the classic setting, buyers are required to participate from an *interim* perspective: their *expected gains* at the time of contracting have to offset their outside option. A salient example discussed by [27] is the airline industry in which, for example, travelers purchase tickets in advance, but may only realize their actual valuation once the date of the trip approaches.

Even though the optimal contracts that arise may offer partial refunds, the initial advanced price is large enough such that some travelers experience negative *ex-post* utility while still being willing to participate interim. This situation arises in other industries as well, such as hotels, theaters or even railroads where advanced pricing/refunds type contracts are also offered.

In many new markets, however, sellers are constrained to sell products in such a way that buyers obtain a non-negative net utility once they have realized their valuation, that is from an *ex-post* perspective. For example, in online shopping buyers may have the chance to return a purchased item after delivery, usually at no or low cost ([42]). In the online display advertising market typical business constraints impose that publishers cannot use up-front fees ([9]) and instead run auctions, for example second-price. Thus, the seller needs to guarantee participation not only initially – at the interim level – but also after the buyers have completely learned their valuation – at the ex-post level.

Motivated by these new markets, we study the sequential screening problem as described by [27] and in order to match our previous narrative we incorporate *ex-post participation constraints*. Ex-post participation constraints rule out the optimal contracts derived by [27] with up-front fees. As pointed out by [42] because different up-front fees cannot be used to price discriminate the different buyers, it may be that a *static* contract, one that does not screen the buyers interim, becomes optimal under ex-post participation constraints. Building on the work by [42], our objective is to understand when in fact the optimal selling mechanism is *static* (buyers are not screened interim) or *sequential* (buyers are screened interim) and obtain a full characterization of such contracts. Our work highlights the significant revenue improvements that can be attained by using a sequential contract relative to a static one, even in the presence of ex-post participation constraints.

Our model considers a seller who is selling one unit of an object at zero marginal cost to a buyer who has an outside option of zero. The sequence of events unfolds in two periods. In the first, the buyer privately learns her *interim* type, for example the mean of her valuation distribution, and the parties contract—important parts of our analysis are done for binary interim types of buyers, low and high. The high type has a distribution of *ex-post* values that dominates the distribution of the low type in some stochastic order. The contract specifies allocation and payment functions. In the second period, the buyer privately learns her valuation, and allocations and transfers are realized. At this point, the buyer only accepts the contracting terms if her realized net utility is weakly larger than her outside option. This model aligns with our aforementioned examples. In online shopping, the first period corresponds to the purchasing time. At this time the buyer possesses private information about her valuation but can only know her valuation with certainty after inspecting the purchased item. In the second period, the buyer is delivered the item and has the option to return it, at low or no cost. In the case of display advertising, some publishers use a sequence of auctions known as "waterfall auctions" that implicitly impose different priorities over participants.[1] Commonly, higher-priority auctions have higher reserve prices. The first period can be thought of as the time at which the buyer decides in which auction (priority/reserve) to participate in. The second period is when the auctions are actually run.

Main contributions. One of our main contributions is to characterize when a static contract—that is, a contract that does not sequentially screen buyers—is optimal. We provide a necessary and sufficient condition for the optimality of the aforementioned contract, we refer to it as the average *profit-to-rent condition*. The characterization we provide is intuitive. At the static contract the seller offers a single price to both low and high type buyers. This price is too large for low types and too low for high types relative to what the seller would set if he were to know the types. To increase his revenue with respect to the static contract, the seller could try to increase the price for high type buyers, however, this would incentivize them to imitate the low types. Another option the seller has is to decrease the price for low type buyers, but this would again incentivize the high types to mimic the low types. In order to increase revenue and to deter high type buyers from imitating the low types, the seller can reduce the price for a portion of low types thus serving more of them and, at the same time, randomize their allocation so that high types

[1]See, for example, https://adexchanger.com/the-sell-sider/the-programmatic-waterfall-mystery. A similar dynamic occurs when sellers offer "preferred deals" to advertisers (see, for example, [50]).

do not take the low types' contract. The *profit-to-rent* condition establishes that this deviation is not profitable for the seller; hence, the profit-to-rent condition is necessary for the optimality of the static contract. Notably, we also show that it is sufficient. Our characterization is a weighted average monotonicity condition of the virtual valuations around the optimal static threshold that in some settings encodes information about the similarity of the interim types. For example, in the case of exponential valuations, the static contract is optimal if and only if the means of the distributions of the low and high type are appropriately close.

Our second main contribution characterizes the optimal mechanism when the condition mentioned above does not hold and a static contract is no longer optimal. We prove that the optimal *sequential* contract randomizes the low type and gives a deterministic allocation to the high type. Randomization occurs to prevent the high type buyer from taking the low type's contract. More specifically, the optimal contract is characterized by an allocation probability $x \in (0, 1)$, and three thresholds θ_1, θ_2, and θ_H with $\theta_1 \leq \theta_H \leq \theta_2$. In this contract, the seller allocates the object to a low type buyer with probability x whenever her valuation is between θ_1 and θ_2, and asks for a payment of $\theta_1 \cdot x$. When the valuation of this type is above θ_2, the object is always allocated to her and the seller demands a payment of $\theta_2 - (\theta_2 - \theta_1) \cdot x$. The high type buyer gets the object with certainty and only when her valuation is above θ_H, at which point the payment she has to make to the seller is θ_H. These parameters are set in such a way that the interim incentive compatibility constraints are satisfied.

A salient feature of this type of contract is that it discriminates the low type in *two dimensions*. First, we establish that θ_1 is above the optimal threshold a seller would set if she was selling exclusively to low type buyers. That is, the low type buyer is being allocated the object less often in the presence of high type buyers. The opposite holds for high type buyers, they are being allocated the object more often than if they were alone. Second, there is a range of values for which the object is sold to the low type with some probability strictly below one, which further reduces the chances of a low type to receive the object compared to a case in which there are no high type buyers. We illustrate these results with the example of the exponential distribution for which we have explicit solutions. We find that for exponential valuations the sequential contract can exhibit revenue improvements of up to 16-27% with respect to the static contract.

Towards the end of the chapter, we consider the case of many interim types. We generalize the *profit-to-rent* condition to a setting with an arbitrary number of interim types. We also discuss directions on how to expand our analysis and results to this setting, as well as the challenges that arise.

3.2 Related Literature

Our model builds on the sequential screening literature as pioneered by [27], with an *interim participation constraint*.[2] In contrast, in this chapter we impose an *ex-post participation constraint*. The closest paper to ours that studies sequential screening with ex-post participation constraints is [42]. They establish that the static contract is optimal under a monotonicity condition regarding the cross-hazard rate functions. This condition rules out some common distributions for values such as the exponential distribution. Furthermore, the condition is only sufficient, and therefore, does not provide a complete characterization of the space of primitives for which the static contract is optimal. We close this gap by providing a necessary and sufficient condition under which the static contract is optimal. Our condition leverages the economic intuition that lies behind a potential profitable deviation from the optimal static contract. Further and importantly, when the condition fails we characterize the optimal

[2] See [2] for a recent adaptation of the [27] formulation to study advanced purchase contracts in revenue management settings.

sequential mechanism and show that randomization of one of the interim types is required for optimality.[3]

In terms of approaches, [42] relax both the low to high incentive constraint and monotonicity constraint and then show that, under their condition, the contract that maximizes the Lagrangian is deterministic and that as a result the static contract is optimal. In contrast, we also relax the incentive constraint but maintain the monotonicity constraint. For the relaxed problem, we perform a first-principle analysis, in the style of [59] and [32] that leads us to identify the right structure of the optimal contract. In turn, this permits us to characterize the optimal sequential contract when our condition breaks. In related recent work, [37] considers a setting in which a seller can design the screening mechanism as well as the information disclosure mechanism with ex-post participation constraints.

The sequential nature of our model and the presence of ex-post participation constraints is related to the work of [7] and [9]. These authors consider a model in which a seller, constrained by ex-post participation (also motivated by the display advertising market), repeatedly sells objects to a buyer whose valuations are independent across periods. Both papers provide characterizations for a nearly optimal mechanism. They are different from ours because we consider a single sale and construct the exactly optimal mechanism in a sequential screening model.

Our optimal mechanism is related to the BIN-TAC auction derived in the context of online display advertising by [24]. This is a *static* auction that offers two options to advertisers: a buy-it-now (BIN) option in which buyers can purchase the impression at a posted high price, and a take-a-chance (TAC) option in which the highest bidders are randomly allocated the impression (if no bidder went for the BIN). This auction

[3]See also [47] and [28] for examples of multi-good environments in which stochastic allocations can improve over deterministic ones. In a related note, [43] establish that with multiple, as opposed to a single good, generically, the static contract is not optimal for the sequential screening problem with ex-post participation constraints.

is tailored to approximate ironing in the classic static Myerson setting for non-regular distributions that commonly arise in display advertising settings. This mechanism is similar in spirit to ours as it randomizes low valuation buyers to separate them from high valuations ones. However, with one bidder the BIN-TAC auction reduces to a posted price which corresponds to the static contract in our setting. In contrast to their static setting, we study a two-period model in which the buyer is sequentially screened and randomization occurs even with one bidder.

3.3 Model

3.3.1 Payoffs and Private Information

We consider a seller (he) who is selling one unit of an object at zero cost to a buyer (she) with an outside option of zero value. Both parties are risk-neutral and have quasilinear utility functions. The sequence of events unfolds in two periods.

In the first period, the buyer privately learns her *type* and then the parties contract. The type provides information about the distribution of the *ex-post values* of the buyer, her true willingness-to-pay for the object. The contract specifies allocation and payment functions.

In the second period, the buyer privately learns her valuation, and allocations and transfers are realized. We refer to the type realized in period 1 as the *interim type* and the valuation realized in period 2 as the *ex-post type*.

There are finitely many types, denoted $k \in \{1, \ldots, K\}$, and the prior probability of type k is given by α_k with $\alpha_k > 0$ and $\sum_{k=1}^{K} \alpha_k = 1$. In the second period, a buyer of type k privately learns her valuation θ which we assume to have a continuously differentiable c.d.f. $F_k(\cdot)$ and pdf $f_k(\cdot)$, with full support in $\Theta \subseteq [0, \infty]$. We assume that Θ is a connected interval of the form $[0, \theta_{max}]$. It will be convenient to denote the upper c.d.f. by

$$\overline{F}_k(\cdot) \triangleq 1 - F_k(\cdot).$$

All the distributions are common knowledge. We denote the virtual valuation $\mu_k(\cdot)$ of interim type k by

$$\mu_k(\theta) \triangleq \theta - \frac{1 - F_k(\theta)}{f_k(\theta)}, \quad \forall k \in \{1, \ldots, K\}, \quad \forall \theta \in \Theta.$$

For the rest of the chapter we make the standard assumption that:

$$\frac{1 - F_k(\theta)}{f_k(\theta)}, \quad \text{is non-increasing in } \theta, \forall k \in \{1, \ldots, K\}. \tag{DHR}$$

This assumption facilitates our discussions. However, for our formal results we will need a weaker assumption that we introduce later.

The terms of trade are specified in the first period by the seller. For a payment $t \in \mathbb{R}$ and a probability of receiving the object $x \in [0, 1]$, a buyer with valuation θ receives a utility of $\theta \cdot x - t$, while the seller gets paid t.

We assume that the buyer agrees to purchase the object only if she is guaranteed a non-negative net utility for any possible valuation of the object she might have. That is, we require $\theta \cdot x - t$ to be non-negative for all θ. The seller's problem is to design a contract that maximizes his expected payment, satisfying the ex-post participation constraint together with incentive compatibility.

3.3.2 Mechanism Design Formulation

By means of the revelation principle (see, e.g., [51]) we can focus on incentive compatible direct revelation mechanisms, with allocations $x_k : \Theta \to [0, 1]$ and transfers $t_k : \Theta \to \mathbb{R}$, that depend on the types (k, θ) reported to the mechanism. Then, for a buyer reporting an interim type k' and an ex-post type θ' the mechanism allocates the object with probability $x_{k'}(\theta')$ and charges the buyer $t_{k'}(\theta')$.

We define the ex-post utility of a buyer who reported k in the first period and θ' in the second period while her true valuation is θ as

$$u_k(\theta;\theta') \triangleq \theta \cdot x_k(\theta') - t_k(\theta'),$$

with the understanding that $u_k(\theta)$ equals $u_k(\theta;\theta)$. Similarly, we define the interim expected utility of a buyer whose true interim type is k but reported to the mechanism k' as

$$U_{kk'} \triangleq \int_{\Theta} \max_{\theta'\in\Theta}\{u_{k'}(z;\theta')\} \cdot f_k(z)dz,$$

where the maximum is included because double deviations are in principle allowed. Note, however, that with distributions with common support and under ex-post incentive compatibility, the maximum will always be achieved at θ' equal to z, and we can restrict attention to single deviations.

There are two kinds of incentive compatibility constraints that must be satisfied by our mechanism. The first one is the ex-post incentive compatibility or (IC^{xp}) constraint which requires that for any report in the first period, truth-telling is optimal in the second period, that is,

$$u_k(\theta) \geq u_k(\theta;\theta') \quad \forall k \in \{1,\dots,K\}, \forall\theta\in\Theta. \tag{IC^{xp}}$$

The second one is the interim incentive compatibility or (IC^{i}) constraint which requires that truth-telling is optimal in the first period, that is,

$$U_{kk} \geq U_{kk'} \quad \forall k,k' \in \{1,\dots,K\}. \tag{IC^{i}}$$

Also, we require the mechanism to satisfy an ex-post individual rationality constraint or (IR^{xp})

$$u_k(\theta) \geq 0, \quad \forall k \in \{1,\dots,K\}, \quad \forall\theta\in\Theta. \tag{IR^{xp}}$$

Then, the seller's problem is

$$\max \quad \sum_{k=1}^{K} \alpha_k \cdot \int_{\Theta} t_k(z) \cdot f_k(z) dz \qquad (\mathcal{P})$$

$$\text{s.t} \quad (IC^i), (IC^{xp}), (IR^{xp})$$

$$0 \leq \mathbf{x} \leq 1 \ ,$$

where we use boldfaces to denote vectors. Observe that (IR^{xp}) implies interim individual rationality. In fact, if we were to relax $(\mathcal{P})$ by considering only interim individual rationality we would be in the setting of [27] for discrete interim types.

In general, two types of contract can arise as a solution to the seller's problem $(\mathcal{P})$: *static* and *sequential*. A static solution to problem $(\mathcal{P})$ corresponds to the case when the allocations and transfers (x_k, t_k) do not depend on the interim type k. In this case we have a unique menu (x, t) that is offered to the buyer and the contract does not screen among interim types. We use $(\mathcal{P}^s)$ to denote the constrained version of $(\mathcal{P})$ to static contracts, which we refer to as *the static program*. In contrast, a sequential solution allows for different menus that depend on the interim type k, and each type of buyer self-selects into one of the menus. The problem $(\mathcal{P})$, referred as *the sequential program*, allows for such solutions.

The main focus of this chapter is two-fold. First, to study when the optimal solutions to the static and sequential programs, $(\mathcal{P}^s)$ and $(\mathcal{P})$, coincide. Second, when they do not coincide, we aim to characterize the optimal solution to $(\mathcal{P})$.

3.4 A Classic Example of Sequential Screening

We use the motivating example of [27] to illustrate the power of sequential screening in the presence of an ex-post participation constraint. We show that a sequential contract outperforms the static contract.

There are two types of potential buyers, low type and high type. One-third of potential buyers are low type whose valuation is uniformly distributed in $[1, 2]$, two-thirds are high type buyers with valuation uniformly distributed in $[0, 1] \cup [2, 3]$. [27] think of the low type as a leisure traveler and of the high type as a business traveler with the same mean but larger variability in her valuation. The seller has a production cost equal to 1.

The optimal static contract sets the optimal monopoly price, $\widehat{p}$, equal to 2, which yields a profit of 1/3. The static contract only serves the high types with high realized valuations. [27] in their setting with an *interim participation constraint* show that the seller can significantly increase its profits with sequential screening by offering a menu of advanced payments/partial refund contracts. They establish that the optimal contract for their setting offers an advanced payment of 1.5 and no refund to the leisure traveler, and an advanced payment of 1.75 and 1 of refund to the business traveler. In this contract a buyer can have a negative realized net utility. For example, the leisure traveler initially pays 1.5 but her actual valuation can be any value within $[1, 2]$ and, therefore, half of the time she will obtain negative net utility after learning her valuation.

Because of the advanced payments these contracts typically will not satisfy an *ex-post participation constraint*, which we study next.

Let us consider the following sequential contract as a simple deviation from the optimal static contract. The seller offers a menu of two quantities and prices, (x_L, p_L) and (x_H, p_H). The second contract is set equal to the optimal static contract, that is, $(x_H, p_H) = (1, 2)$. Hence, the selling price for the high type is 2 and high types that buy receive the full quantity.

Now, we find the optimal quantity and price for the low type buyer. Given the contract for the high type, the seller's profit is given by:

$$\frac{1}{3} \times x_L \times (p_L - 1) \times (2 - p_L) + \frac{2}{3} \times \frac{1}{2} \times (2 - 1),$$

where $x_L \in [0,1]$ and $p_L \in [1,2]$. We need to ensure that the menus are interim incentive compatible. The low to high incentive constraint is always satisfied (p_H equals 2), and the high to low incentive constraint is given by:

$$\frac{1}{2} \times \left(\frac{5}{2} - 2\right) \geq \frac{1}{2} \times x_L \times \left(\frac{5}{2} - p_L\right).$$

Profit maximization implies that this constraint must be binding, and therefore, the seller's profit becomes:

$$\frac{1}{3} \times \frac{(p_L - 1) \times (2 - p_L)}{5 - 2p_L} + \frac{1}{3}.$$

The first order condition yields an optimal price equal to $\left(5 - \sqrt{3}\right)/2$ which, in turn, delivers a profit of $2/3 - 1/(2\sqrt{3})$. The improvement of the sequential contract versus the optimal static contract is then $1 - \sqrt{3}/2 \approx 13\%$.

From this simple exercise we learn an important lesson: even in a simple setting a sequential contract can have substantial benefits over a static contract. In this chapter we study more generally when a sequential contract outperforms a static contract and what drives this revenue improvement.

3.5 Optimality of Static Contract

First, we start by characterizing conditions under which it is optimal not to screen the interim types. In the main theorem of this section we provide a necessary and sufficient condition for the static contract to be optimal. We begin with a reformulation of the problem based on standard techniques that use the envelope theorem, and enables us to solve for the allocation and utilities of the lowest ex-post types instead of both allocations and transfers. Using the reformulation we characterize the optimal static contract. In Section 3.5.2, we use the optimal static contract together with a simple deviation analysis to obtain an intuitive necessary condition for its optimality. In Section 3.5.3, we show that this condition is both necessary and sufficient.

3.5.1 Problem Reformulation and Static Solution

We obtain a more amenable characterization of the constraints by eliminating the transfers from the them as in the classical Myersonian analysis.

Lemma 3.1 (Necessary and Sufficient Conditions for Implementation) *The mechanism* $(\mathbf{x}, \mathbf{t})$ *satisfies* (IC^i),(IC^{xp}) *and* (IR^{xp}) *if and only if*

1. $x_k(\cdot)$ *is a non-decreasing function for all* k *in* $\{1, \ldots, K\}$ *and*

$$u_k(\theta) = u_k(0) + \int_0^\theta x_k(z)dz, \quad \forall k \in \{1, \ldots, K\}, \forall \theta \in \Theta. \tag{3.1}$$

2. $u_k(0) \geq 0$ *for all* k *in* $\{1, \ldots, K\}$.

3. $u_k(0) + \int_\Theta x_k(z)\overline{F}_k(z)dz \geq u_{k'}(0) + \int_\Theta x_{k'}(z)\overline{F}_k(z)dz$ *for all* k, k' *in* $\{1, \ldots, K\}$.

All proofs are provided in the Appendix. The first condition in the lemma is the standard envelope condition and it comes from the ex-post incentive compatibility constraint. The second condition is derived from the ex-post individual rationality constraint and the fact that $u_k(\theta)$ is non-decreasing. The third condition is the envelope formula inserted into the interim incentive compatibility constraint.

Lemma 3.1 enables us to obtain a more compact formulation for the seller's problem. Specifically, we can use equation (3.1) and integration by parts to write down the objective of $(\mathcal{P})$ in terms of the allocation rule $\mathbf{x}$ and the indirect utilities $\{u_k(0)\}_{k=1}^K$ of the lowest ex-post types. To this end, we denote each $u_k(0)$ as a new variable by u_k. The new formulation is then:

$$\begin{aligned}
\max_{0 \leq \mathbf{x} \leq 1} \quad & -\sum_{k=1}^K \alpha_k u_k + \sum_{k=1}^K \alpha_k \int_\Theta x_k(z)\mu_k(z)f_k(z)dz \qquad (\mathcal{P}) \\
\text{s.t} \quad & x_k(\theta) \quad \text{non-decreasing}, \quad \forall k \in \{1, \ldots, K\} \\
& u_k \geq 0, \quad \forall k \in \{1, \ldots, K\} \\
& u_k + \int_\Theta x_k(z)\overline{F}_k(z)dz \geq u_{k'} + \int_\Theta x_{k'}(z)\overline{F}_k(z)dz, \quad \forall k, k' \in \{1, \ldots, K\},
\end{aligned}$$

Note that in ($\mathcal{P}$) the variables are the allocation rule $\mathbf{x}$ and the vector of the indirect utilities of the lowest ex-post types $\boldsymbol{u}$. Once we solve for these variables the transfers are determined by equation (3.1).

As we mentioned before, a solution to ($\mathcal{P}$) that screens the interim types is a sequential contract. In contrast, a static solution to ($\mathcal{P}$) pools the interim types. Formally, we say that a solution to ($\mathcal{P}$) or contract is *static* when $x_k(\cdot) \equiv x(\cdot)$ and $u_k \equiv u$ for all k in $\{1, \dots, K\}$.

We earlier defined the virtual valuation $\mu_k(\cdot)$ of interim type k. Given (DHR) the virtual valuation for each type k has exactly one zero which we denote by $\widehat{\theta}_k$. Without loss of generality we assume *for the remainder of the chapter* that we can order the interim types:

$$\widehat{\theta}_1 \leq \cdots \leq \widehat{\theta}_K.$$

It turns out that solving ($\mathcal{P}$) over the space of static contracts is a simpler problem. The (IC^{xa}) constraints disappear from the problem because in this case there is effectively only one interim type. Also, it is clear that any optimal solution sets $u_k = 0$ for all k in $\{1, \dots, K\}$. So, the static version of the seller's problem is given by

$$\max_{0 \leq x \leq 1} \quad \int_{\Theta} x(z) \cdot \Big(\sum_{k=1}^{K} \alpha_k \mu_k(z) f_k(z) \Big) dz \qquad (\mathcal{P}^s)$$

$$\text{s.t} \quad x(\theta) \quad \text{non-decreasing,}$$

where a simple calculation shows that the term in parenthesis is equal to the virtual value function of the mixture distribution times the density function of the mixture. Hence, this problem corresponds to the classic optimal mechanism design problem applied to the mixture distribution over types.

From this formulation we see that the relevant quantity that shapes the allocation $x(\cdot)$ is $\bar{\mu}(\theta) \triangleq \sum_{k=1}^{K} \alpha_k \mu_k(\theta) f_k(\theta)$. In general, because there is only one buyer, independent of any regularity assumptions imposed over $\bar{\mu}(\theta)$, one can show that an optimal way to choose a non-decreasing allocation $x(\cdot)$ that maximizes

$$\int_{\Theta} x(z)\bar{\mu}(z)dz, \tag{3.2}$$

is a threshold allocation, that is, a single posted price (see, e.g., [52] or [57]). We summarize this in the following lemma.

Lemma 3.2 (Threshold Allocation)

A solution to $(\mathcal{P}^s)$ *is a threshold allocation characterized by* $\widehat{\theta}$ *in* $[\widehat{\theta}_1, \widehat{\theta}_K]$ *that maximizes* (3.2).

3.5.2 A Necessary Condition

In the rest of this Section and the next Section 3.6 we provide our results for the setting with binary interim types. We denote the low type by L and the high type by H. In Section 3.7 we return to the general setting with finitely many interim types.

The static optimal solution is characterized by a threshold allocation $\widehat{\theta}$. In this section, we leverage this characterization, and perform an analysis in the style of [21], to deduce an intuitive necessary condition for the optimality of the static contract. As we will show later in Section 3.5.3 this condition turns out to be not only necessary but also sufficient.

For ease of exposition, we assume that the high type dominates the low type in the hazard rate order sense:

$$\frac{1-F_H(\theta)}{f_H(\theta)} \geq \frac{1-F_L(\theta)}{f_L(\theta)}, \quad \forall \theta \in \Theta. \tag{3.3}$$

We note that we do not need this assumption for the formal arguments.

Suppose now that a static contract is optimal, that is, setting a single posted price equal to $\widehat{\theta}$ for both types solves $(\mathcal{P})$. Consider Figure 3.1, where we have plotted the virtual value function weighted by the density function for each type.[4] If the types

[4]We needly represent the virtual valuation weighted by $f_k(\cdot)$. This does not change the signs in the figure but gives a convenient geometric representation.

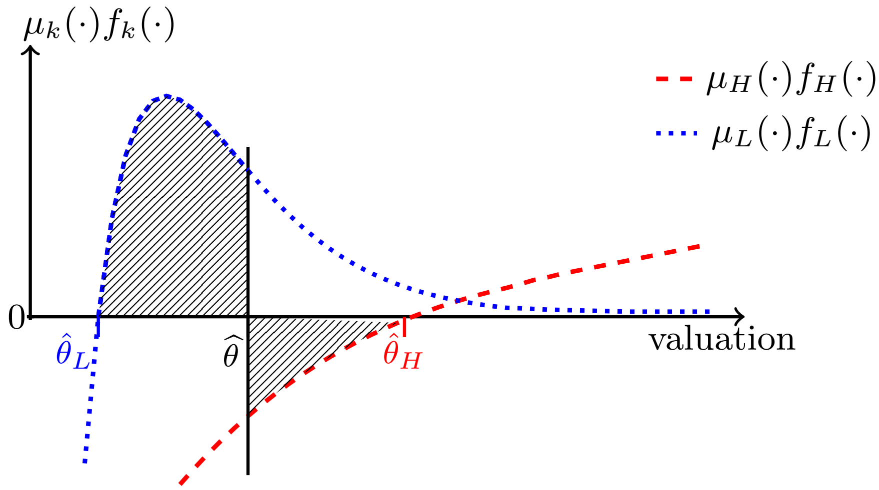


Figure 3.1: Weighted virtual valuations for low type (dotted line) and high type (dashed line) buyer around $\widehat{\theta}$. The shaded areas correspond to the virtual revenue that the seller leaves on the table when using a static contract with respect to the case in which the interim types are public information.

were public, the seller would optimally set posted prices equal to $\widehat{\theta}_L$ and $\widehat{\theta}_H$ for types L and H, respectively. In this way, the seller would serve buyers if and only if they have positive virtual values. In contrast, when selecting a single posted price $\widehat{\theta}$, there is surplus that the seller is not extracting; the shaded area shows the regions of the virtual valuations for each type that the static contract is not capturing. For the high type, the static contract serves too many buyers, some of them with negative virtual values; hence, the seller would be better off by offering a higher price. For the low type, the static contract serves too few buyers, leaving positive virtual value buyers unserved; hence, the seller would prefer to choose a lower price. A challenge, though, is that the seller faces incentive compatibility constraints that restrict this type of possible deviations/improvements:

1. Selling to fewer high types implies increasing the price for high types; but then the high types have an incentive to accept the low type contract and such a deviation is not feasible.

2. Selling to more low types amounts to reducing the price from $\widehat{\theta}$ to some value θ_1.

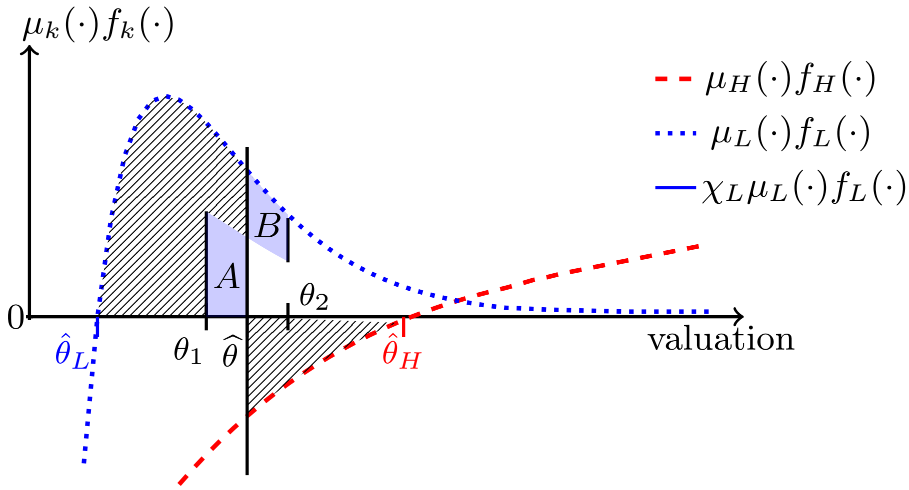


Figure 3.2: Weighted virtual valuations for low type (dotted line) and high type (dashed line) buyer around $\widehat{\theta}$. The shaded areas correspond to the virtual revenue that the seller leaves on the table when using a static contract with respect to the case in which the interim types are public information. We show deviation from the static contract for the low type (solid line). If $A - B \geq 0$ the deviation is profitable.

> However, to prevent the high types from taking the low type contract the seller must decrease the quantity offered to the low types (or equivalently, randomize their allocation).

This second improvement is feasible by choosing a quantity (probability) $0 < x_L < 1$ to all low types inside an interval $[\theta_1, \theta_2]$ with $\theta_1 \leq \widehat{\theta} \leq \theta_2$, see Figure 3.2.

Formally these allocations correspond to the following menu:

$$x_L(\theta) = \begin{cases} 0 & \text{if } \theta < \theta_1, \\ x_L & \text{if } \theta_1 \leq \theta \leq \theta_2, \\ 1 & \text{if } \theta_2 < \theta; \end{cases} \quad x_H(\theta) = \begin{cases} 0 & \text{if } \theta < \widehat{\theta}, \\ 1 & \text{if } \widehat{\theta} \leq \theta; \end{cases} \tag{3.4}$$

with $u_L = u_H = 0$. We refer to this deviation as *an interior variation or improvement.*

The interior improvement is feasible only if it satisfies both incentive compatibility constraints. Inserting the menu (3.4) into the incentive constraints in ($\mathcal{P}$) we obtain for the low type:

$$x_L \int_{\theta_1}^{\theta_2} (1 - F_L(\theta))d\theta + \int_{\theta_2}^{\theta_{max}} (1 - F_L(\theta))d\theta \geq \int_{\widehat{\theta}}^{\theta_{max}} (1 - F_L(\theta))d\theta,$$

and for the high type:

$$\int_{\widehat{\theta}}^{\theta_{max}} (1 - F_H(\theta))d\theta \geq x_L \int_{\theta_1}^{\theta_2} (1 - F_H(\theta))d\theta + \int_{\theta_2}^{\theta_{max}} (1 - F_H(\theta))d\theta,$$

and/or in a more compact form as a bracketing inequality:

$$\frac{\int_{\widehat{\theta}}^{\theta_2}(1 - F_L(\theta))d\theta}{\int_{\theta_1}^{\theta_2}(1 - F_L(\theta))d\theta} \leq x_L \leq \frac{\int_{\widehat{\theta}}^{\theta_2}(1 - F_H(\theta))d\theta}{\int_{\theta_1}^{\theta_2}(1 - F_H(\theta))d\theta}, \tag{3.5}$$

which contains both incentive compatibility constraints. The monotone hazard rate condition (3.3) guarantees that x_L as in given by (3.5) always exists. The interior variation is thus feasible and we can select x_L so as to maximize the seller's revenue.

Indeed, evaluating the interior variation in the seller's objective yields:

$$x_L \cdot \int_{\theta_1}^{\theta_2} \mu_L(\theta) f_L(\theta) d\theta + \int_{\theta_2}^{\theta_{max}} \mu_L(\theta) f_L(\theta) d\theta,$$

and since $\mu_L(\theta) \geq 0$ in $[\theta_1, \theta_2]$ (c.f Figure 3.2) the right hand side inequality in (3.5) must be tight.

With the interior variation, the seller serves more low-value buyers in $[\theta_1, \widehat{\theta}]$ at the level of x_L. This comes at the expense of offering a lower quantity, a loss of $1 - x_L$ to buyers with values in $[\widehat{\theta}, \theta_2]$. In Figure 3.2 the area A corresponds to the additional revenue the seller can make due to the variation because he is serving more low type buyers, and region B is the efficiency loss due to the incentive constraints.

If the static contract is optimal then this variation cannot be profitable. In terms of Figure 3.2 this means the areas must satisfy $A \leq B$. Hence, if the static contract is optimal then

$$A = x_L \cdot \int_{\theta_1}^{\widehat{\theta}} \mu_L(\theta) f_L(\theta) d\theta \leq (1 - x_L) \cdot \int_{\widehat{\theta}}^{\theta_2} \mu_L(\theta) f_L(\theta) d\theta = B.$$

In turn, since the optimal choice of x_L always equals the right hand side of (3.5), we can insert x_L in terms of the ratio, and after some re-arranging we get

$$\frac{\int_{\theta_1}^{\widehat{\theta}} \mu_L(\theta) f_L(\theta) d\theta}{\int_{\theta_1}^{\widehat{\theta}}(1 - F_H(\theta))d\theta} \leq \frac{\int_{\widehat{\theta}}^{\theta_2} \mu_L(\theta) f_L(\theta) d\theta}{\int_{\widehat{\theta}}^{\theta_2}(1 - F_H(\theta))d\theta}. \tag{3.6}$$

To better understand this inequality consider a monopolist who faces a consumer with valuation distributed according to $F_k(\cdot)$. Observe that at some price θ_b the expected profit $\Pi_k(\theta_b)$ the monopolist makes and the expected consumer's *informational rents* $I_k(\theta_b)$ are given by

$$\Pi_k(\theta_b) \triangleq \theta_b \cdot (1 - F_k(\theta_b)) = \int_{\theta_b}^{\theta_{max}} \mu_k(\theta) f_k(\theta) d\theta \quad \text{and} \quad I_k(\theta_b) \triangleq \int_{\theta_b}^{\theta_{max}} (1 - F_k(\theta))\, d\theta.$$

If the monopolist considers lowering the price from θ_b to θ_a then the change in profit is $\Pi_k(\theta_a) - \Pi_k(\theta_b)$. The lower price positively impacts the information rents which increase by $I_k(\theta_a) - I_k(\theta_b)$. The ratio $(\Pi_k(\theta_a) - \Pi_k(\theta_b))/(I_k(\theta_a) - I_k(\theta_b))$ then is a measure of the average impact in profits per unit of consumer rents the seller experiences due to the price variation. In condition (3.6) we have a cross version of this ratio. In the numerator we take $k = L$ and in the denominator $k = H$. In light of this observation condition (3.6) suggests the following definition.

Definition 3.1 (Average Profit-to-Rent Ratio)

The average profit-to-rent ratio is defined by:

$$R^{jk}(\theta_a, \theta_b) \triangleq \frac{\Pi_j(\theta_a) - \Pi_j(\theta_b)}{I_k(\theta_a) - I_k(\theta_b)}, \quad \forall j, k \in \{L, H\}, \quad 0 \leq \theta_a \leq \theta_b \leq \theta_{max}.$$

The average profit-to-rent ratios measure changes in the seller's profit normalized by the information rents he gives away to the consumer due to a price deviation. The ratio R^{jk} compares the impact on profit for type j with the increase in the information rent for type k. This cross ratio arises as the incentive compatibility constraint for type k implies that a modification in the contract for type j affects type k as well. This was clear from our discussion regarding the internal variation above. There, a price θ_1 (smaller than $\widehat{\theta}$) for the type L creates a profit improvement for the seller measured by the numerator of R. Since the seller has to make sure that type H does not take the type L contract (by reducing quantity), this price decrease generates a loss to the seller quantified by the denominator of R.

Back to (3.6) we notice that the numerator in either ratio refers to the revenue that the seller is making from the low type over some interval, and the denominator refers to the information rent of the high type over the same interval. Now, since the choice of θ_1, θ_2 was arbitrary, we obtain the following necessary condition by taking minimum and maximum at both sides of the inequality in (3.6). If the static contract is optimal then

$$\max_{\theta_1 \leq \widehat{\theta}} R^{LH}(\theta_1, \widehat{\theta}) \leq \min_{\widehat{\theta} \leq \theta_2} R^{LH}(\widehat{\theta}, \theta_2), \tag{3.7}$$

The above condition establishes that if the static contract is optimal then any extra revenue the seller can garner from low type buyers is offset by the efficiency loss due to the incentive compatibility constraints: $A - B \leq 0$ for any possible choice of θ_1 and θ_2.

3.5.3 A Necessary and Sufficient Condition

We now establish that condition (3.7) is in fact a necessary and sufficient condition for the optimal static solution to coincide with the optimal solution to $(\mathcal{P})$. Before we provide the main theorem, we introduce some notation for the quantities of interest that will help us to further refine our intuition. While we maintain the binary type framework here; we note that all definitions naturally extend to finitely many types as we will see in Section 6.

The local version of the average profit-to-rent ratio, when $\theta_a < \widehat{\theta} < \theta_b$ are close to $\widehat{\theta}$, gives raise to the *profit-to-rent ratio*.

Definition 3.2 (Profit-to-Rent Ratio)

The profit-to-rent ratio between type j and k is defined by:

$$r^{jk}(\theta) \triangleq \frac{\mu_j(\theta) f_j(\theta)}{1 - F_k(\theta)}, \quad \forall j, k \in \{L, H\}, \forall \theta \in \Theta.$$

The ratio $r^{jk}(\theta_b)$ is obtained by $\lim_{\theta_a \uparrow \theta_b} R^{jk}(\theta_a, \theta_b)$. Observe that condition (DHR) is stronger than and implies that $r^{kk}(\theta)$ is non-decreasing for each $k \in \{L, H\}$. The latter is the condition we use for our formal results.

Now, we are ready to state and discuss the main result of this section.

Theorem 3.1 (Optimality of Static Contract)

Suppose $r^{kk}(\theta)$ is non-decreasing for each $k \in \{L, H\}$. The static contract is optimal if and only if

$$\max_{\theta \leq \widehat{\theta}} R^{LH}(\theta, \widehat{\theta}) \leq \min_{\widehat{\theta} \leq \theta} R^{LH}(\widehat{\theta}, \theta). \tag{APR}$$

This results completes the necessity condition given in Section 3.5.2 by showing that it is also sufficient. We showed in Section 3.5.2 that condition (APR) established that the specific deviation that increases the sales to the lower type with a lower quantity is not profitable relative to the static contract.

Theorem 3.1 now establishes that in fact this is not only a necessary but in fact a sufficient condition. The sufficiency condition is noteworthy as it arises from "simple" deviations, namely, those that assign the low type an interior allocation in a small interval around the static optimal price. In particular, we do not need to be concerned either with more elaborate deviations which offers the low type several options in his menu, nor do we need to trace simultaneous changes to the offers to the high type. The present theorem confirms that this type of interior improvement for the low type is sufficient to study changes in the seller's revenue. In fact, we will establish in Section 3.6 that the family of allocations suggested by the interior variation completely describes the optimal sequential mechanism as well.

To prove the sufficiency in Theorem 3.1 we rely a on dualization-type of argument. For the necessity, we assume that condition (APR) is not satisfied and show that in that case there is a profitable deviation as given by the following proposition.

Proposition 3.1 (Revenue Improvement)

Suppose $r^{LL}(\theta)$ *is non-decreasing. Assume condition* (APR) *does not hold. Then there exists* θ_1, θ_2 *such that* $\theta_1 < \widehat{\theta} < \theta_2$ *and* $R^{LH}(\theta_1, \widehat{\theta}) > R^{LH}(\widehat{\theta}, \theta_2)$*, for which the allocation in* (3.4) *with*

$$x_L = \frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z)dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz},$$

yields a strict improvement in $(\mathcal{P})$ *over the static contract.*

In the proof of Proposition 3.1 we can see that as soon as condition (APR) breaks two things happen. First, a non-static contract becomes feasible as it does not violate the incentive compatibility constraints. Note that the proposition is similar to the discussion in Section 3.5.2; however, it is more general because it does not assume hazard rate order to guarantee feasibility. The mere fact that (APR) breaks implies the feasibility of the new allocation. Second, the same contract obtains a larger expected revenue than the static one. So, from this we see that (APR) is preventing both the feasibility and optimality of a sequential contract.

3.5.4 The Exponential Example

Before we move to the study of the optimal sequential contract it might be helpful to build some more intuition for the results. We shall consider the case of exponentially distributed values. The main result of this section establishes that the static contract is optimal if and only if the mean of the interim types are sufficiently close.

We consider the exponential density functions

$$f_k(\theta) = \lambda_k e^{-\lambda_k \theta}, \quad k = \{L, H\} \quad \theta \geq 0.$$

We assume $\lambda_L > \lambda_H$, so L and H stand for low and high type respectively. Note that H has a higher mean $(1/\lambda_H)$ than L $(1/\lambda_L)$ and that H dominates L in the sense of the hazard rate stochastic order and in first order stochastic dominance. In addition, for the interim probabilities we have $\alpha_L + \alpha_H = 1$ with $\alpha_L, \alpha_H > 0$.

We begin by studying the optimal solution to the static formulation. The optimal static contract is given by a threshold allocation. Thus, in the exponential case the seller's expected revenue for any given threshold θ is

$$\Pi^{\text{static}}(\theta) \triangleq \int_{\theta}^{1} (\alpha_L \mu_L(z) f_L(z) + \alpha_H \mu_H(z) f_H(z)) dz = \alpha_L \theta e^{-\lambda_L \theta} + \alpha_H \theta e^{-\lambda_H \theta}.$$

In order to find the optimal threshold we just need to maximize the expression above. The first order condition yields

$$\alpha_L(\theta - \frac{1}{\lambda_L})\lambda_L e^{-\lambda_L \theta} + \alpha_H(\theta - \frac{1}{\lambda_H})\lambda_H e^{-\lambda_H \theta} = 0, \tag{3.8}$$

that is, the optimal threshold is a zero of the mixture virtual valuation. Notice that equation (3.8) cannot be explicitly solved; however, we can (as we do in the forthcoming results) provide comparative statics. Interestingly, in Proposition 3.3 below, we show that we can obtain explicit expressions for the thresholds characterizing the optimal sequential contract. The following lemma provides some initial properties of the optimal static contract.

Lemma 3.3

The optimal solution to $(\mathcal{P}^s)$ is a threshold allocation characterized by $\widehat{\theta}$ in $[\frac{1}{\lambda_L}, \frac{1}{\lambda_H}]$, solving (3.8)*. Also, $\widehat{\theta}$ is a non-increasing function of α_L with $\widehat{\theta}(0) = \frac{1}{\lambda_H}$ and $\widehat{\theta}(1) = \frac{1}{\lambda_L}$.*

Next, we state a necessary and sufficient condition for the static contract to be optimal.

Proposition 3.2 (Necessity and Sufficiency for the Exponential Model)

The static contract is optimal if and only if

$$\lambda_L - \lambda_H \leq \frac{1}{\widehat{\theta}} \tag{3.9}$$

The result follows from Theorem 3.1, but it requires some effort to determine the max and min in (APR) in closed form. We note that in the right hand side, $\widehat{\theta}$, is a solution to equation (3.8) and, therefore, it also depends on the parameters λ_L and λ_H. Subsequent corollaries provide sharper characterizations that only depend on model primitives. We highlight that (3.9) corresponds to a particular case of condition (APR).

Proposition 3.2 provides an intuitive characterization for when the seller is better-off screening the interim types than not. In terms of equation (3.9), when λ_L and λ_H are sufficiently close, then equation (3.9) should hold, in which case the static contract is optimal. Conversely, when λ_L and λ_H are sufficiently apart from each other, the static contract may not be optimal.

Intuitively, when the interim types are similar any contract that screens the types would be close in terms of expected revenue to the static contract because for each type it could get at most what it would get by setting thresholds $1/\lambda_L$ and $1/\lambda_H$ respectively, but $\widehat{\theta} \in [\frac{1}{\lambda_L}, \frac{1}{\lambda_H}]$. However, when screening, the seller has to pay an extra cost to prevent the types from mimicking each other and, since the contracts' revenue will be similar, it is likely that this cost offsets the earnings from screening. On the other hand, when interim types are sufficiently apart in their mean valuation then the seller can tailor the contract to each type and in this way extract more from them than in the static contract.

Corollary 3.1 *Assume $\lambda_L \in (\lambda_H, 2\lambda_H]$, then for any $\alpha_L \in [0,1]$ the static contract is optimal.*

This result establishes that when the distributions of the low and high type buyers are sufficiently close to each other then no matter in which proportion the types are, the static contract is always optimal.

Corollary 3.2 *Assume $\lambda_L > 2\lambda_H$, then there exists $\bar{\alpha} \in (0,1)$ such that for all $\alpha_L \in (0,\bar{\alpha})$ the sequential contract is strictly optimal and for all $\alpha_L \in [\bar{\alpha},1]$ the static contract is optimal.*

Corollary 3.2 asserts that when the mean of the low and high type buyers are sufficiently different then both contracts can be optimal. If the proportion of low type is low enough (but not zero) then the seller is better-off screening the types. On the other hand, if there is a very large proportion of low type buyers then the static contract is optimal. This follows because as α_L increases, one can show that $\widehat{\theta}$ decreases, and at some point condition (3.9) holds. This discussion suggests our final corollary.

Corollary 3.3 *For λ_H and α_H fixed, there exists $\bar{\lambda}_L$ larger than $2\lambda_H$ such that for all $\lambda_L \in [\bar{\lambda}_L, \infty)$ the sequential contract is strictly optimal.*

3.5.5 Discussion

We introduced earlier the condition (DHR) which establishes that the hazard rates

$$h^{jk}(\theta) = \frac{1 - F_k(\theta)}{f_j(\theta)}$$

are non-increasing when j equals k. A related condition is about the cross-hazard rate functions,

$$h^{jk}(\theta) \quad \text{are non-increasing in } \theta, \quad \forall j,k \in \{L,H\}. \tag{R}$$

To the best of our knowledge condition (R) was first introduced in the context of sequential screening by [42]. In that paper the authors show that under condition (R) the optimal solution to $(\mathcal{P})$ and to $(\mathcal{P}^s)$ coincide, that is, the static contract is optimal. In fact, they show this result for multiple interim types. We discuss our generalization of condition (APR) to multiple types in Section 3.7. However,

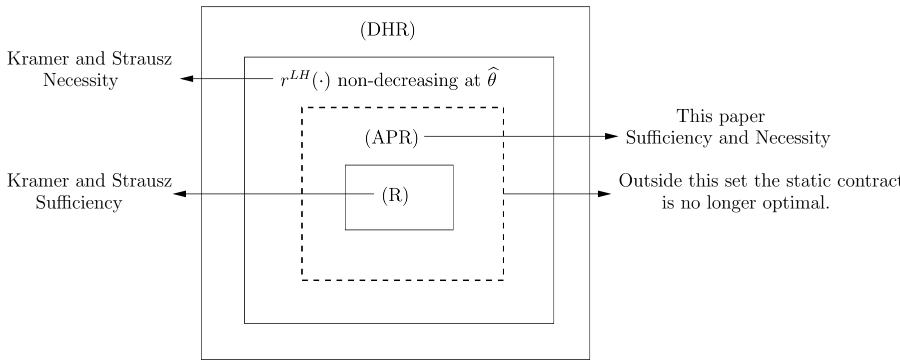


Figure 3.3: Optimality of the static contract for (DHR) distributions, with $K = 2$ and a single buyer.

condition (R) is rather restrictive and is not satisfied by some common distributions. For example, the condition is not satisfied by any pair of exponential distributions, because in this case the cross-hazard rate is given by:

$$h^{jk}(\theta) = \frac{e^{-(\lambda_k - \lambda_j)\theta}}{\lambda_j}, \quad j, k = L, H.$$

If, without loss of generality, we consider $\lambda_L > \lambda_H$ then $h^{LH}(\theta)$ is an increasing function and, therefore, it violates conditions (R). However, notice (DHR) is satisfied because the simple hazard rate functions are constant and equal to $1/\lambda_k$.

We can also compare Theorem 3.1 with Lemma 12 in [41]. In that Lemma they assume $h^{HH}(\theta) > h^{LL}(\theta)$, which implies $\widehat{\theta}_L < \widehat{\theta}_H$, and establish that a necessary condition for the static contract to be optimal is to have the profit- to-rent ratio $r^{LH}(\theta)$ being increasing at $\widehat{\theta}$. Our result also contains this lemma, because if $r^{LH}(\cdot)$ was decreasing at $\widehat{\theta}$ we can always find $\theta_1 < \widehat{\theta}$ and $\theta_2 > \widehat{\theta}$ such that

$$R^{LH}(\theta_1, \widehat{\theta}) > R^{LH}(\widehat{\theta}, \theta_2),$$

so (APR) does not hold and, therefore, the static contract would not be optimal. Figure 3.3 illustrates how our condition (APR) closes the gap between the ones by Krähmer and Strausz.

We can compare condition (R) and (APR). Note that condition (R) implies the monotonicity of the profit-to-rent ratios, and therefore condition (APR) holds as

$$R^{LH}(\theta,\widehat{\theta}) = \frac{\int_{\theta}^{\widehat{\theta}} \overline{F}_H(z) r^{LH}(z)dz}{\int_{\theta}^{\widehat{\theta}} \overline{F}_H(z)dz} \leq r^{LH}(\widehat{\theta}), \quad \forall \theta \leq \widehat{\theta},$$

and

$$R^{LH}(\widehat{\theta},\theta) = \frac{\int_{\widehat{\theta}}^{\theta} \overline{F}_H(z) r^{LH}(z)dz}{\int_{\widehat{\theta}}^{\theta} \overline{F}_H(z)dz} \geq r^{LH}(\widehat{\theta}), \quad \forall \theta \geq \widehat{\theta}.$$

Hence, the result by [42] that if condition (R) holds then the static contract is optimal follows as a corollary of Theorem 3.1. We highlight that while condition (R) implies the profit- to-rent ratios are increasing, our condition (APR) only implies a type of monotonicity over an appropriate weighted average of the profit-to-rent ratios. This is sensible as we are dealing with interim expected seller's revenues and interim incentive compatibility constraints.

In terms of methodology, our approach differs from that of [42]. Their approach consists of relaxing the low to high interim IC constraint and then – by using their condition (R) – they relax the monotonicity constraint and prove that the solution must be a threshold schedule for each type. From there, they show that the threshold for both types must be equal and, therefore, the static contract is optimal.

In our approach we do not use a relaxation of the general formulation nor do we impose conditions on the primitives besides that $r^{kk}(\theta)$ are non-decreasing. For the sufficiency we construct a Lagrangian relaxation with multipliers for the incentive compatibility constraints, but we do not relax the monotonicity constraints. The multipliers relate to the profit-to-rent ratios at the static threshold $\widehat{\theta}$; they measure the change in the objective per unit of change in the constraints. Then by leveraging a result from [57] that the optimal contract must involve a threshold allocation we prove that under (APR) the solution to the relaxation is the static contract.

3.6 Sequential Contract

We now proceed to provide the complete characterization of the optimal sequential contract when the necessary and sufficient condition associated with the static contract fails. As hinted in Section 3.5.2 and by Proposition 3.1 the optimal sequential contract gives a deterministic allocation to the high type and, for mid-range values, it randomizes the low type buyer (or equivalently reduces the quantity allocated).

3.6.1 The Structure of the Sequential Contract

Our analysis consists in studying the following relaxation to $(\mathcal{P})$

$$\max_{0\leq \mathbf{x}\leq 1} \quad -\sum_{k\in\{L,H\}} \alpha_k u_k + \sum_{k\in\{L,H\}} \alpha_k \int_\Theta x_k(z)\mu_k(z)f_k(z)dz \tag{$\mathcal{P}_R$}$$

$$\begin{aligned} \text{s.t} \quad & x_k(\theta) \quad \text{non-decreasing}, \quad \forall k\in\{L,H\} \\ & u_k \geq 0, \forall k\in\{L,H\} \\ & u_H + \int_\Theta x_H(z)\overline{F}_H(z)dz \geq u_L + \int_\Theta x_L(z)\overline{F}_H(z)dz. \end{aligned}$$

The difference between $(\mathcal{P}_R)$ and the original problem $(\mathcal{P})$ is the omission of the incentive constraint for the low type to report truthfully. Importantly, we do not relax the monotonicity constraint. We obtain a characterization of the optimal solution to $(\mathcal{P}_R)$ as stated by the following theorem.

Theorem 3.2 (Relaxed Solution)

Suppose $r^{kk}(\theta)$ is non-decreasing for each $k\in\{L,H\}$. Consider problem $(\mathcal{P}_R)$, the optimal solution has allocations

$$x_L^\star(\theta) = \begin{cases} 0 & \text{if } \theta < \theta_1, \\ x_L & \text{if } \theta_1 \leq \theta \leq \theta_2, \\ 1 & \text{if } \theta_2 < \theta; \end{cases} \quad x_H^\star(\theta) = \begin{cases} 0 & \text{if } \theta < \theta_H, \\ 1 & \text{if } \theta_H \leq \theta. \end{cases}$$

The threshold values $\theta_1, \theta_H, \theta_2$ *satisfy* $\widehat{\theta}_L \leq \theta_1 \leq \theta_H \leq \theta_2$, $\theta_H \leq \widehat{\theta}_H$ *and*

$$x_L = \frac{\int_{\theta_H}^{\theta_2} \overline{F}_H(z)dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz}.$$

Note that if $\theta_1 = \theta_H$ we recover the static contract. Importantly, the optimal contract of $(\mathcal{P}_R)$ has the same structure as the profitable deviation to the static contract presented in Proposition 3.1. The only difference is that in the former the threshold for the high type may not necessarily equal to $\widehat{\theta}$ as in the latter. With this generalization one can show that the proposed profitable deviation is indeed optimal for $(\mathcal{P}_R)$. The associated transfers are given by:

$$t_L^\star(\theta) = \begin{cases} 0 & \text{if } \theta < \theta_1, \\ \theta_1 \cdot x_L & \text{if } \theta_1 \leq \theta \leq \theta_2, \\ \theta_2 - (\theta_2 - \theta_1) \cdot x_L & \text{if } \theta_2 < \theta; \end{cases} \quad t_H^\star(\theta) = \begin{cases} 0 & \text{if } \theta < \theta_H, \\ \theta_H & \text{if } \theta_H \leq \theta. \end{cases}$$

Our optimality proof adapts arguments by [32] to our setting. We use an improvement argument to show that the optimal contract of $(\mathcal{P}_R)$ only requires a simple threshold allocation without randomization for the high type. Finally, we use another improvement argument to show that the low type allocation only requires a single interval of randomization.

More specifically, consider a low type allocation that randomizes between an interval $[\theta_a, \theta_b]$. Recall the argument in Section 3.5.3 where we found a revenue improvement while keeping feasibility, in particular, while maintaining the high to low IC constraint. Using a similar reasoning, we can show that feasibly improving upon the random allocation requires the following condition to hold for some $\tilde{\theta}$:

$$R^{LH}(\theta_a, \tilde{\theta}) = \frac{\int_{\theta_a}^{\tilde{\theta}} \overline{F}_H(z) r^{LH}(z)dz}{\int_{\theta_a}^{\tilde{\theta}} \overline{F}_H(z)dz} \leq \frac{\int_{\tilde{\theta}}^{\theta_b} \overline{F}_H(z) r^{LH}(z)dz}{\int_{\tilde{\theta}}^{\theta_b} \overline{F}_H(z)dz} = R^{LH}(\tilde{\theta}, \theta_b). \quad (3.10)$$

In general this condition is not satisfied, because the profit-to-rent ratio $r^{LH}(\cdot)$ does not need to be a non-decreasing function. Therefore, we cannot find a feasible improvement over the random allocation contract, and hence, we cannot restrict attention to deterministic contracts for the low type. In contrast, a similar argument for the high type yields the expression $R^{HH}(\theta_a, \tilde{\theta}) \leq R^{HH}(\tilde{\theta}, \theta_b)$, which always holds when $r^{HH}(\cdot)$ is non-decreasing. Hence, we can restrict attention to a deterministic threshold contract for the high type.

In addition, the low type allocation only requires a single interval of randomization. To see this, suppose for example that $x_L^\star(\theta)$ equals x_a in $(\theta_a, \tilde{\theta})$ and x_b in $(\tilde{\theta}, \theta_b)$ with $0 < x_a < x_b < 1$, and also assume (3.10) does not hold. Then, it is possible to show that we can increase x_a and decrease x_b (maintaining feasibility) and obtain an improvement to the objective function. We can do this until x_a and x_b collapse into a single value.

The discussion above highlights again the importance of the average profit-to-rent ratios in our analysis, as they quantify revenue improvements while maintaining incentive compatibility. Now, the next result characterizes the optimal sequential contract and it also provides conditions that allow to compute the optimal thresholds.

Theorem 3.3 (Optimal Sequential Contract)

Suppose $r^{kk}(\theta)$ is non-decreasing for each $k \in \{L, H\}$. The optimal sequential contract coincides with the optimal solution of $(\mathcal{P}_R)$ as given by Theorem 3.2.

In Theorem 3.2 we provided the characterization of the optimal solution to $(\mathcal{P}_R)$. In the proof of Theorem 3.3 we argue that the optimal solution to $(\mathcal{P}_R)$ is feasible for $(\mathcal{P})$ and thus optimal. In turn, we obtain a full characterization of the optimal sequential contract up to three parameters.

In terms of solving for the optimal sequential contract, Theorems 3.2 and 3.3 imply that we can ignore the IC constraints and do a search over three parameters to maximize seller's revenues over the proposed contract structure, θ_1, θ_2 and θ_H. In the proof of Theorem of 3.3 we show that the optimality conditions for the thresholds $\theta_1 \leq \theta_H \leq \theta_2$ are:

1. $R^{LH}(\theta_1, \theta_2) \leq \min_{\theta_2 \leq \theta} R^{LH}(\theta_2, \theta)$;

2. $\max_{\theta \leq \theta_2} R^{LH}(\theta, \theta_2) \leq R^{LH}(\theta_1, \theta_2)$;

3. $\alpha_L \cdot R^{LH}(\theta_1, \theta_2) + \alpha_H r^{HH}(\theta_H) = 0$.

Conditions (1) and (2) put together are similar to (APR) where θ_2 plays the role of $\widehat{\theta}$. Similarly to the case of the static contract, one can show that any allocation that randomizes beyond θ_2 is never profitable. In turn, randomization should only occur for valuations below θ_2. Condition (2) by itself also implies that among all the intervals that can be randomized, the interval (θ_1, θ_2) is the most profitable. To see this let us compare the seller's revenue when it randomizes the low type buyer over some interval (θ, θ_2) and (θ_1, θ_2) (and it gives a deterministic allocation to the high type). Using Theorem 3.2 the allocation x_L that satisfies incentive compatibility in each case is:

$$x_L(\theta) = \frac{\int_{\theta_H}^{\theta_2} \overline{F}_H(z)dz}{\int_{\theta}^{\theta_2} \overline{F}_H(z)dz} \quad \text{and} \quad x_L(\theta_1) = \frac{\int_{\theta_H}^{\theta_2} \overline{F}_H(z)dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz}.$$

Hence, doing a revenue comparison, we conclude that randomizing the low type buyer over (θ_1, θ_2) is better than over (θ, θ_2) if and only if

$$\frac{\int_{\theta_H}^{\theta_2} \overline{F}_H(z)dz}{\int_{\theta}^{\theta_2} \overline{F}_H(z)dz} \cdot \int_{\theta}^{\theta_2} \mu_L(z) f_L(z)dz \leq \frac{\int_{\theta_H}^{\theta_2} \overline{F}_H(z)dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz} \cdot \int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z)dz,$$

equivalently, $R^{LH}(\theta, \theta_2) \leq R^{LH}(\theta_1, \theta_2)$ for arbitrary $\theta \leq \theta_2$ which is exactly condition (2). Finally, condition (3) is simple a first order optimality condition on θ_H.

It is interesting to note that in the optimal solution the low type buyers are allocated the object over a larger interval ($\theta_1 \leq \theta_H$) but they are randomized. This is done as a way to prevent the buyers from mimicking each other. Specifically, we must have $\theta_1 \leq \theta_H$; otherwise, the low type buyers would have an incentive to pretend being the high type since that would get them allocated the object more often and at a lower price. Similarly, $\theta_H \leq \theta_2$ otherwise high type buyers would choose the low type contract for a better allocation and a lower price.

It is worth noting that the sequential contract makes the low type worse-off and the high type better-off with respect to the contract the seller would offer if he could perfectly screen each type. For the low type, that contract would set a threshold equal to $\widehat{\theta}_L$ and would always allocate the object when her value is above the threshold. However, the sequential contract allocates the object to the low type whenever her valuation is above $\theta_1 \geq \widehat{\theta}_L$ with positive probability. So the low type is worse-off in two dimensions, it is allocated the object less often and with less probability. On the other hand, the high type buyer gets allocated the object more often and with certainty since $\theta_H \leq \widehat{\theta}_H$.

3.6.2 The Exponential Example Continued

In Section 3.5.4 we studied the properties and structure of the optimal static contract for exponential valuations. In particular, we applied our necessary and sufficient condition to this family of distributions and obtained an intuitive characterization.

Proposition 3.3

Assume condition (3.9) does not hold, then the optimal allocation is

$$x^\star_L(\theta) = \begin{cases} 0 & \text{if } \theta < \theta_1, \\ x & \text{if } \theta_1 \leq \theta; \end{cases} \quad \text{and} \quad x^\star_H(\theta) = \begin{cases} 0 & \text{if } \theta < \theta_H, \\ 1 & \text{if } \theta_H \leq \theta; \end{cases}$$

The thresholds are given by:

$$\theta_1 = \frac{1}{\lambda_L - \lambda_H} \quad \text{and} \quad \theta_H = \frac{1}{\lambda_H} - \frac{\alpha_L}{\alpha_H}\frac{e^{-1}}{\lambda_L - \lambda_H},$$

with $\theta_1 \leq \theta_H$. The probability of receiving the object for the low type is:

$$x = \exp\Big(- \lambda_H \Big[\frac{1}{\lambda_H} - \frac{\alpha_L}{\alpha_H}\frac{e^{-1}}{\lambda_L - \lambda_H} - \frac{1}{\lambda_L - \lambda_H}\Big]\Big). \tag{3.11}$$

This result follows from Theorem 3.3 and the characterization of the three free parameters that follow. We note that in the exponential case we only have two intervals for the low type's allocation as we can show that $\theta_2 = \infty$.

We now illustrate our findings with numerical results where we vary the difference in the mean between the low and the high type. Specifically, we fix α_L to be 0.7 and λ_H to be 0.5, that is, the high type has mean 2. Since we are assuming $\lambda_L > \lambda_H$, we consider $\lambda_L = \lambda_H + \delta$ with $\delta > 0$. Figure 3.4 shows how the different thresholds vary as δ increases or, equivalently, as the mean of the low type decreases to zero. As we can see, there is a value of δ (δ =0.93) to the left of which the static contract is optimal and to its right the sequential contract is optimal. This aligns with Proposition 3.2 because as δ increases, $(\lambda_L - \lambda_H)$ increases, and therefore, we expect it to be larger than $1/\widehat{\theta}$ (see Corollary 3.2 and Corollary 3.3). At a more intuitive level as δ increases both distribution become more and more different from each other with one of them having a larger average value than the other. Thus, there is gain in screening the types.

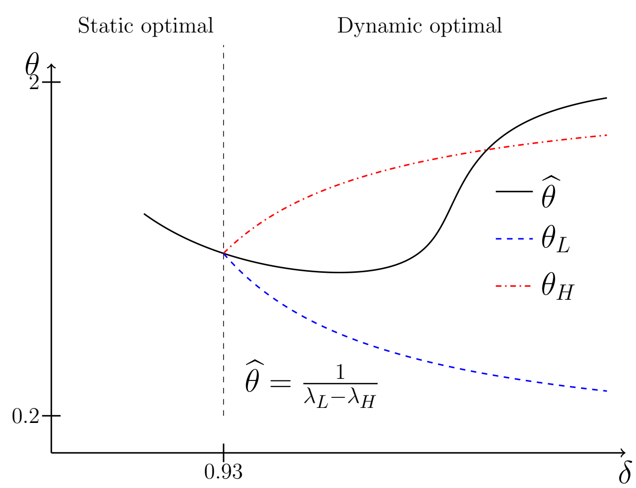


Figure 3.4: Optimal thresholds for static and sequential contracts when setting $\lambda_L = \lambda_H + \delta$, with $\alpha_L = 0.7$ and $\lambda_H = 0.5$.

In terms of thresholds, for the static contract we observe that $\widehat{\theta}$ is decreasing at the beginning and then it increases getting closer to $1/\lambda_H = 2$. This happens because as we increase δ we are making $1/\lambda_L$ smaller; however, at some point this value is too small and, therefore, the probability of allocating the object to a low type, $P(\text{value low type} > \widehat{\theta}) = e^{-\lambda_L \widehat{\theta}}$, is going to be so low that the seller will be better off by choosing a threshold tailored for the high type, that is, close to $1/\lambda_H = 2$. For the sequential thresholds, the one for the low type is decreasing while the one for the high type is increasing. This makes sense because in the sequential case the seller can adjust the threshold for each type; hence, as δ increases the distributions become more and more different and, therefore, is optimal to set thresholds closer and closer to the threshold a seller would set if he knew the types in advance, that is, $1/\lambda_L$ and $1/\lambda_H$. Also, note that from equation (3.11) we see that x is a decreasing function of δ because as the mean of the low type goes to zero we are less and less constrained to offer a high probability of allocation; however, in the limit $x(\delta) \approx e^{-1}$, hence even though the low type buyers will have values concentrated at zero we still need to reduce their quantity so that high types do not take their low price contract.

We can also compare the different mechanism in terms of revenue. Note that with the contracts from Proposition 3.3, the optimal revenue for the sequential contract Π^{seq} can be shown to be equal to:

$$\Pi^{\text{seq}} = \alpha_L \cdot x \cdot \theta_1 \cdot e^{-\lambda_L \theta_1} + \alpha_H \cdot \theta_H \cdot e^{-\lambda_H \theta_H}.$$

Then, we can plot the different revenues as we vary δ. Figure 3.5 (left panel and thick line in right panel) depicts the results. For values of δ above 0.93 the sequential contract dominates the static. Further the sequential contract can achieve a significant improvement over the static contract, getting as high as 16.5%. Note that when δ grows large the improvement of the sequential over the static decreases because both contracts set the thresholds to maximize what they can extract from the high type buyer. Actually, with some abuse of notation, we have that

$$\lim_{\delta \to \infty} \Pi^{\text{seq}}(\delta) = \lim_{\delta \to \infty} \Pi^{\text{static}}(\delta) = \alpha_H \frac{e^{-1}}{\lambda_H},$$

which equals the optimal revenue a seller could make if he was only selling to the high type buyer. The right panel in Figure 3.5 shows the revenue improvement for different instances as we vary α_L. Consistent with Corollary 3.3, given α_H and λ_H, there exists λ_L large enough such that the sequential contract is strictly better than the static one. The figure also shows that the larger α_L the larger has to be the difference between the types for the sequential and static contracts to differ. When α_L is large $\widehat{\theta}$ is tailored for low types and so (3.9) holds for more values of λ_L. However, screening occurs when the mean of the low type is sufficiently small (δ large) in which case, due to the low values and high fraction of the low type, the revenue improvement can achieve better percentage performance (e.g., 27% for $\alpha_L = 0.9$).

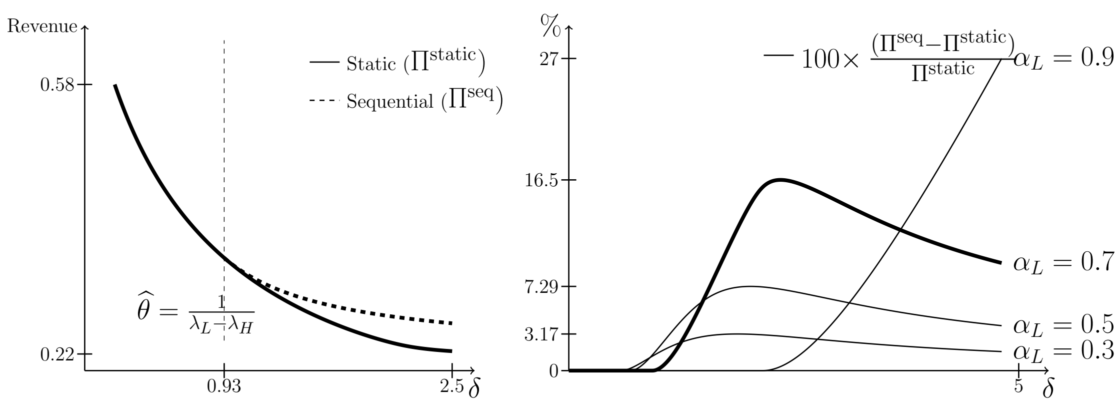


Figure 3.5: *Left:* Optimal expected revenue for static and sequential. *Right:* Percentage improvement of the sequential over the static contract. In both figures we set set $\lambda_L = \lambda_H + \delta$ with $\lambda_H = 0.5$. In the *left* figure we set $\alpha_L = 0.7$ while in the *right* figure α_L takes values in $\{0.3, 0.5, 0.7, 0.9\}$.

3.6.3 Menu Implementation

Next, we discuss how the optimal sequential contract can be implemented in practice. By means of the taxation principle we can verify that the following menu of contracts is an indirect implementation of our optimal mechanism:

- contract H: there is a single posted price of $p_H = \theta_H$;

- contract L: the buyer can choose between two items:

  (a) buy at a price of $p_L = \theta_1 \cdot x_L$ and be allocated with probability x_L.

  (b) buy at a price of $p_L = \theta_2 - (\theta_2 - \theta_1) \cdot x_L$ and be allocated with probability 1.

The prices in the above menu of contracts are set using the values in Theorems 3.2 and 3.3. This implementation offers a posted price to the high type buyer, and gives to the low type buyer two options. In option (a) the low type buyer can pay a low price but it can potentially not acquire the item or equivalently, get a reduced quantity; in (b), the low type buyer pays a high price and always gets the object.

An appealing feature of the implementation is that if we think of allocations as quantities, then we can order the per unit prices. In contract L, the per unit prices are θ_1 and $\theta_1 \cdot x_L + \theta_2 \cdot (1 - x_L)$ for (a) and (b), respectively. Hence, the per unit price in (a) is less than or equal to the one in (b). That is, the low type in (a) receives less of the good but at a discounted price compare to the low type in (b). For contract H, the per unit price is θ_H and, since θ_1 is less than or equal to θ_H, the low type in (a) receives less of the good at a discounted price compared to the high type buyer. Comparing the per unit prices of the low type in (b) and the high type is less straightforward. Even-though θ_H is between θ_1 and θ_2 we are not able to compare it to $\theta_1 \cdot x_L + \theta_2 \cdot (1 - x_L)$. However, intuitively, even if the high type puts a large mass in values larger than θ_2 we expect the per unit price of the high type to be below the one of the low type in (b) because, otherwise, the high type buyer would have an incentive to take contract L. Equivalently, the high type or the low type in (b) have to pay a premium for the additional quantity. We can also refer back to the exponential case of Section 3.5. From Proposition 3.3, the premium the high type has to pay is given by $\theta_H - \theta_1 = \log(1/x_L)/\lambda_H$ and, therefore, the larger the quantity the lower is the premium. Finally, note that this implementation accommodates the case in which the static contract is optimal. In that case, we have $x_L = 1$ and $\theta_1 = \theta_H = \theta_2$ thus both contracts are the same.

3.7 Multiple Types

Until now, we have studied the optimality of the static contract and the optimal sequential mechanism for two types of interim buyers. In this section, we consider an arbitrary number of interim types $\{1, \ldots, K\}$ and investigate some properties of the solution to $(\mathcal{P})$. In particular, we provide a generalized version of condition (APR). Then, we provide numerical evidence and highlight the challenges associated with the characterization of the optimal sequential mechanism when $K > 2$.

3.7.1 A Necessary and Sufficient Condition

Our generalized necessary and sufficient condition relies on a characterization of the changes in the objective around the static solution when considering allocation deviations. With this purpose, consider the following set:

$$\mathcal{A} \triangleq \Big\{(\lambda_{ij})_{i,j\in\{1,\cdots,K\}^2} \geq 0 : \sum_{j\neq k} \lambda_{jk} \cdot \overline{F}_j(\widehat{\theta}) = \alpha_k \cdot \mu_k(\widehat{\theta}) \cdot f_k(\widehat{\theta}) + \overline{F}_k(\widehat{\theta}) \cdot \sum_{j\neq k} \lambda_{kj},$$
$$\alpha_k \geq \sum_{j\neq k} \lambda_{kj} - \sum_{j\neq k} \lambda_{jk}, \quad \forall k \in \{1, \ldots, K\}\Big\}.$$

The set $\mathcal{A}$ contains the multipliers associated with the IC constraints that encode the change in the objective as we deviate from the static allocation. Roughly speaking, when the static contract is optimal, allocation perturbations in the contract of each type should equal the dualized costs associated to such perturbations in the IC constraints. In other words, the derivative of the Lagrangian with respect to allocations around the static solution equals zero. This is captured by the set of equalities in the definition of $\mathcal{A}$. In addition, the set of inequalities ensures that the optimal ex-post utilities of the lowest valuation buyers are zero. Note that multipliers being in the set

$\mathcal{A}$ are necessary for optimality. The next result provides a necessary and sufficient condition.

Theorem 3.4 (Necessary and Sufficient Conditions for Finitely Many Types)

The set $\mathcal{A}$ is non-empty. Furthermore, if there exists a feasible solution to $(\mathcal{P})$ which strictly satisfies all the IC constraints then the static contract is optimal if and only if there exist $(\lambda_{ij})_{i,j\in\{1,\cdots,K\}^2} \in \mathcal{A}$ such that

$$\max_{\theta\leq\widehat{\theta}}\left\{\alpha_k\cdot R^{kk}(\theta,\widehat{\theta})-\sum_{j\neq k}\lambda_{jk}\cdot\frac{\int_{\theta}^{\widehat{\theta}}\overline{F}_j(z)dz}{\int_{\theta}^{\widehat{\theta}}\overline{F}_k(z)dz}\right\}\leq\min_{\widehat{\theta}\leq\theta}\left\{\alpha_k\cdot R^{kk}(\widehat{\theta},\theta)-\sum_{j\neq k}\lambda_{jk}\cdot\frac{\int_{\widehat{\theta}}^{\theta}\overline{F}_j(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_k(z)dz}\right\}, \tag{APR^M}$$

for all $k\in\{1,\ldots,K\}$.

The strict feasibility to $(\mathcal{P})$ corresponds to a standard Slater condition. Condition (APR^M) is obtained by studying the Lagrangian when the static contract is optimal and disentangling the key conditions it must satisfy. To obtain a better understanding of this condition it is helpful to see how it generalizes the necessary and sufficient condition provided in Theorem 3.1 for two types. The general condition of Theorem 3.4 in the binary case becomes

$$\max_{\theta\leq\widehat{\theta}}\left\{\alpha_1\cdot R^{11}(\theta,\widehat{\theta})-\lambda_{21}\cdot\frac{\int_{\theta}^{\widehat{\theta}}\overline{F}_2(z)dz}{\int_{\theta}^{\widehat{\theta}}\overline{F}_1(z)dz}\right\}\leq\min_{\widehat{\theta}\leq\theta}\left\{\alpha_1\cdot R^{11}(\widehat{\theta},\theta)-\lambda_{21}\cdot\frac{\int_{\widehat{\theta}}^{\theta}\overline{F}_2(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_1(z)dz}\right\}, \tag{3.12}$$

for the low type, and

$$\max_{\theta\leq\widehat{\theta}}\left\{\alpha_2\cdot R^{22}(\theta,\widehat{\theta})-\lambda_{12}\cdot\frac{\int_{\theta}^{\widehat{\theta}}\overline{F}_1(z)dz}{\int_{\theta}^{\widehat{\theta}}\overline{F}_2(z)dz}\right\}\leq\min_{\widehat{\theta}\leq\theta}\left\{\alpha_2\cdot R^{22}(\widehat{\theta},\theta)-\lambda_{12}\cdot\frac{\int_{\widehat{\theta}}^{\theta}\overline{F}_1(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_2(z)dz}\right\}, \tag{3.13}$$

for the high type, where λ_{12} and λ_{21} belong to $\mathcal{A}$. We next argue that condition (APR) holds if and only if there exist $\lambda_{12},\lambda_{21}\in\mathcal{A}$ such that conditions (3.12) and (3.13) hold. Suppose (APR) holds. Since we expect the low to high IC constraint not to be binding we take λ_{12} equal to zero. Because λ must belong to $\mathcal{A}$ this necessarily implies that λ_{21} is equal to $\alpha_1 r^{12}(\widehat{\theta})$. For this choice of multipliers inequality (3.13) follows directly from r^{kk} being increasing. At the same time, the choice of multipliers together with (APR) imply that both the max and the min in (3.12) are equal to zero. To see this consider the maximum in (3.12) and take $\theta = \widehat{\theta}$, since λ_{21} equal to $\alpha_1 r^{12}(\widehat{\theta})$ the expression inside the brackets is zero. Hence, the maximum in (3.12) is bounded below by zero. It is also bounded above by zero,

$$\alpha_1 \cdot R^{11}(\theta, \widehat{\theta}) - \lambda_{21} \cdot \frac{\int_\theta^{\widehat{\theta}} \overline{F}_2(z)dz}{\int_\theta^{\widehat{\theta}} \overline{F}_1(z)dz} \leq 0 \Leftrightarrow R^{12}(\theta, \widehat{\theta}) \leq r^{12}(\widehat{\theta}), \quad \forall \theta \leq \widehat{\theta}.$$

When (APR) holds the right hand side inequality above always holds. A similar argument applies to the min. Therefore, the condition provided in Theorem 3.1 implies APR^M for the binary case. The converse implication follows from a contradiction argument which for the sake of brevity we omit.

The two type case is amenable to this simplification because one can readily solve for the multipliers: λ_{12} equal to zero is a natural choice (the low to high IC constraints can be relaxed c.f Section 3.6), and λ_{21} equal to $\alpha_1 r^{12}(\widehat{\theta})$ then follows from the definition of $\mathcal{A}$. Unfortunately, when $K > 2$ the space of deviations is richer and so is the possible selection of multipliers; in turn this precludes such a clear characterization as in the two type case.

We stress that by judiciously choosing the multipliers it is straightforward to verify that as in the two type case, condition (R) of [42] implies our condition (APR^M) also in the case of multiple types, and thus the optimality of the static contract.

By contrast, a complete characterization of the sequential contract seems substantially more complex with finitely many types. Next, in the context of exponentially distributed ex-post types, we briefly describe partial results and highlight the challenges associated with multiple types that already appear in the numerical analysis.

3.7.2 The Exponential Example Continued

Despite the challenges that we discuss below, we are able to provide the following characterization

Proposition 3.4 *For exponential valuations the optimal allocations have at most one randomized interval.*

Proposition 3.4 establishes that for exponentially distributed valuation the optimal contract is simple in the sense that each interim type's allocation is randomized at most in one interval. The proof consists on noticing that the monotonicity constraints form a cone, and then using duality and complementary slackness. It is worth mentioning that the proof method applies more generally but the structure of the contract in general depends on the values of the dual variables values corresponding to the IC constraints. In the exponential case, the argument can be simplified to show that the simple structure in the result arises independent of these variables' values.

The characterization in Proposition 3.4 only establishes the structure of the optimal allocations but it does not provide information on the number of contracts that the optimal solutions will ultimately feature. For example, if $K = 4$ Proposition 3.4 does not say whether the optimal solution will pool the interim types creating either one, two, three or four different contracts. In general, the full range of contracts from static to fully sequential (K different contracts) is possible.

To further explore the structure of optimal contracts we provide numerical results. In Figure 3.6 we show the optimal allocations when $K = 4$ and all interim types have exponentially distributed valuations. A first observation is that for different proportions α_k of interim types the optimal contract can feature different levels of separation. Panel (a) in the figure corresponds to an optimal static contract (no separation), and panel (d) in the figure corresponds to an optimal sequential contract that features a different contract for each interim type (full separation). As a second

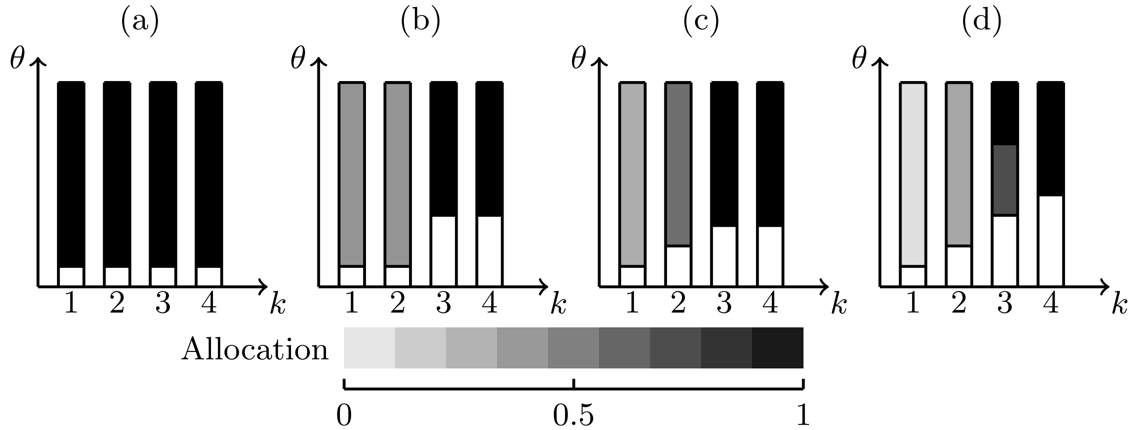


Figure 3.6: Optimal allocations for $K = 4$, types have exponential distribution with means $(2.2, 5.0, 12, 50)$ respectively (for numerical simplicity, we use truncated versions of these distributions in the interval [0,60]). In each panel the vertical axis corresponds to buyers' valuations and the horizontal axis corresponds to the interim type. Each bar represents the allocation for each type, lighter grey indicates lower probability of allocation while darker grey indicates higher probability of allocation. White represents no allocation and black full allocation. From panel (a) to (d) the fractions, α_k, for each type are: $(0.7, 0.2, 0.05, 0.05)$, $(0.4, 0.1, 0.4, 0.1)$, $(0.3, 0.2, 0.4, 0.1)$ and $(0.25, 0.25, 0.1, 0.4)$, respectively.

observation note that out of the four instances depicted in Figure 3.6 only one, (d), has four contracts in the optimal solution. Finding the minimal number of contracts that give a good approximation to the optimal multiple type sequential contract is a question outside the scope of this chapter but that may be of interest to study in the future.

Observe that across the instances in Figure 3.6 each optimal contract has at most one interval of valuation for which randomization occurs (c.f Proposition 3.4). This simple structure of the optimal contract appears however not to be robust to other specifications of the value distributions. When we consider the case of normally distributed valuations (using truncated normal random variables), the optimal contract might exhibit several different intervals of randomization for a given type. In general, richer contract features may arise when we combine exponential, normal, uniform or other distributions. As a consequence, generally speaking, it is challenging to analytically characterize the optimal solution. The challenge here is that classic relaxation approaches used in the mechanism design literature do not apply in our setting. For example, relaxing all the upward incentive constraints and leaving only the local downward incentive constraints does not work because in general global downward incentive constraints bind. Moreover, binding constraints are highly sensitive to model primitives. Improving our understanding of this setting may be an interesting avenue for future research.

3.8 Conclusion

We considered the scope of sequential screening in the presence of ex-post participation constraints. The ex-post participation constraints limit the ability of the seller to extract surplus. As the buyer has to be willing to participate in the contractual arrangement following every realization of his valuation, the surplus has to be extracted ex-post rather than at the interim level.

Despite these restrictions sequential screening generally allows the seller to increase his revenue beyond the statically optimal revenue. The gains from sequential screening become more pronounced to the extent that the interim types differ in their willingness to pay. A natural implementation of the optimal mechanism simply offers the buyer the choice among different menus in the first stage. The choice of menu in the first period merely restricts the possible choices in the second period. In particular, it is not necessary to ask the buyer for any transfer before the final transaction occurs. Moreover, the buyer only has to make a transfer if she receives the object.

In contrast to the static solution where the optimal policy is always to sell the largest possible quantity, the sequential screening policy offers intermediate quantities. This departure from the bang-bang policy in a linear utility setting arises due to the presence of the ex-post participation constraint in conjunction with the incentive compatibility constraints.

There are several natural directions to extend the present work. Our stronger results were for the case of binary interim types while allowing for a continuum of valuations for each type. We also presented an extension of Theorem 3.1 to multiple types as well as a characterization and numerical results for exponential valuations. We would like to further explore the characterization of the optimal sequential contract to multiple types and general valuation distributions. An interesting question here concerns the number of randomization intervals per type and whether the number of intermediate allocations increases with the number of interim types. Also, is there a fixed number of intermediate allocations that yield a good approximation to the optimal solution for an arbitrary number of interim types? Similarly, is there a fixed number of contracts that yield a good approximation to the optimal solution for an arbitrary number of interim types?

We might also be interested in analyzing how the number of competing buyers may affect the nature of the optimal mechanism. This has important practical consequences particularly in industries that use market mechanisms like auctions, such as display advertising alluded at the beginning of the chapter. We note that this extension is not direct, because with multiple buyers we lose the threshold structure of the optimal static allocation when the mixture distribution is not regular and ironing may be required. However, we conjecture that in this case an approximately optimal market design would consist of running a series of “waterfall auctions” with different priorities across participants.

Bibliography

[1] Philipp Afèche, Zhe Liu, and Costis Maglaras. Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. *Working paper*, 2017.

[2] Mustafa Akan, Baris Ata, and James Dana. Revenue management by sequential screening. *Journal of Economic Theory*, 159(PB):728–774, 2015.

[3] Zeynep Aksin, Mor Armony, and Vijay Mehrotra. The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management*, 16(6):665–688, 2007.

[4] Arnold O Allen. *Probability, statistics, and queueing theory*. Academic Press, 2014.

[5] Luigi Ambrosio and Aldo Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In *Optimal transportation and applications*, pages 123–160. Springer, 2003.

[6] Edward J Anderson and Peter Nash. *Linear programming in infinite-dimensional spaces: theory and applications*. John Wiley & Sons, 1987.

[7] Itai Ashlagi, Constantinos Daskalakis, and Nima Haghpanah. Sequential mechanisms with ex-post participation guarantees. 2016.

[8] Rami Atar. A diffusion regime with nondegenerate slowdown. *Operations Research*, 60(2):490–500, 2012.

[9] Santiago Balseiro, Vahab Mirrokni, and Renato Paes Leme. Dynamic mechanisms with martingale utilities. 2016.

[10] Siddhartha Banerjee, Daniel Freund, and Thodoris Lykouris. Multi-objective pricing for shared vehicle systems. *arXiv preprint arXiv:1608.06819*, 2016.

[11] Siddhartha Banerjee, Yash Kanoria, and Pengyu Qian. State dependent control of closed queueing networks. In *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*, pages 2–4. ACM, 2018.

[12] Siddhartha Banerjee, Carlos Riquelme, and Ramesh Johari. Pricing in ride-share platforms: A queueing-theoretic approach. 2015. Working paper.

[13] Achal Bassamboo, Ramandeep S Randhawa, and Assaf Zeevi. Capacity sizing under parameter uncertainty: Safety staffing principles revisited. *Management Science*, 56(10):1668–1686, 2010.

[14] Dimitris J Bertsimas and Garrett van Ryzin. A stochastic and dynamic vehicle routing problem in the euclidean plane. *Operations Research*, 39(4):601–615, 1991.

[15] Dimitris J Bertsimas and Garrett van Ryzin. Stochastic and dynamic vehicle routing in the euclidean plane with multiple capacitated vehicles. *Operations Research*, 41(1):60–76, 1993.

[16] Omar Besbes, Francisco Castro, and Ilan Lobel. Surge pricing and its spatial supply response. 2018. Working paper, Columbia University.

[17] Kostas Bimpikis, Ozan Candogan, and Daniela Saban. Spatial pricing in ride-sharing networks. *Operations Research (forthcoming)*, 2016.

[18] Adrien Blanchet and Guillaume Carlier. Optimal transport and cournot-nash equilibria. *Mathematics of Operations Research*, 41(1):125–145, 2015.

[19] Anton Braverman, Jim G Dai, Xin Liu, and Lei Ying. Empty-car routing in ridesharing systems. 2016. Working paper, Northwestern University.

[20] Nicholas Buchholz. Spatial equilibrium, search frictions and efficient regulation in the taxi industry. 2017. Working paper.

[21] Jeremy Bulow and John Roberts. The simple economics of optimal auctions. *Journal of political economy*, 97(5):1060–1090, 1989.

[22] Gerard P Cachon, Kaitlin M Daniels, and Ruben Lobel. The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management*, 2017.

[23] Juan Camilo Castillo, Dan Knoepfle, and Glen Weyl. Surge pricing solves the wild goose chase. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 241–242. ACM, 2017.

[24] L Elisa Celis, Gregory Lewis, Markus Mobius, and Hamid Nazerzadeh. Buy-it-now or take-a-chance: Price discrimination through randomized auctions. *Management Science*, 60(12):2927–2948, 2014.

[25] Carri W Chan, Galit Yom-Tov, and Gabriel Escobar. When to use speedup: An examination of service systems with returns. *Operations Research*, 62(2):462–482, 2014.

[26] Richard Cole, Yevgeniy Dodis, and Tim Roughgarden. Pricing network edges for heterogeneous selfish users. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 521–530. ACM, 2003.

[27] Pascal Courty and Hao Li. Sequential screening. *The Review of Economic Studies*, 67(4):697–717, 2000.

[28] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. Strong duality for a multiple-good monopolist. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 449–450. ACM, 2015.

[29] Raymond J Deneckere and Preston R McAfee. Damaged goods. *Journal of Economics & Management Strategy*, 5(2):149–174, 1996.

[30] Jing Dong, Pnina Feldman, and Galit B Yom-Tov. Service systems with slowdowns: Potential failures and proposed solutions. *Operations Research*, 63(2):305–324, 2015.

[31] Guiyun Feng, Guangwen Kong, and Zizhuo Wang. We are on the way: Analysis of on-demand ride-hailing systems. 2017. Working paper, University of Minnesota.

[32] William Fuchs and Andrzej Skrzypacz. Government interventions in a dynamic market with adverse selection. *Journal of Economic Theory*, 158:371–406, 2015.

[33] Noah Gans, Ger Koole, and Avishai Mandelbaum. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, 5(2):79–141, 2003.

[34] Ofer Garnett, Avishai Mandelbaum, and Martin Reiman. Designing a call center with impatient customers. *Manufacturing & Service Operations Management*, 4(3):208–227, 2002.

[35] Itai Gurvich, Martin Lariviere, and Antonio Moreno. Operations in the on-demand economy: Staffing services with self-scheduling capacity. 2016. Working paper, Kellogg School of Management, Northwestern University, Evanston, IL.

[36] Shlomo Halfin and Ward Whitt. Heavy-traffic limits for queues with many exponential servers. *Operations research*, 29(3):567–588, 1981.

[37] Tiborn Heumann. Optimal sequential screening with ex-post participation constraint. 2016.

[38] Harold Hotelling. Stability in competition. In *The Collected Economics Articles of Harold Hotelling*, pages 50–63. Springer, 1990.

[39] Ming Hu and Yun Zhou. Dynamic type matching. 2016. Working paper, Rotman School of Management, University of Toronto.

[40] Ming Hu and Yun Zhou. Price, wage and fixed commission in on-demand matching. 2017. Working paper, Rotman School of Management, University of Toronto.

[41] Daniel Krähmer and Roland Strausz. Optimal sales contracts with withdrawal rights. 2014. mimeo.

[42] Daniel Krähmer and Roland Strausz. Optimal sales contracts with withdrawal rights. *The Review of Economic Studies*, 82(2):762–790, 2015.

[43] Daniel Krähmer and Roland Strausz. Optimality of sequential screening with multiple units and ex post participation constraints. *Economics Letters*, 142:64–68, 2016.

[44] David G Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1969.

[45] Avi Mandelbaum and Gennady Pats. State-dependent queues: approximations and applications. *Stochastic networks*, 71:239–282, 1995.

[46] Avi Mandelbaum, Gennady Pats, et al. State-dependent stochastic networks. part i. approximations and applications with continuous diffusion limits. *The Annals of Applied Probability*, 8(2):569–646, 1998.

[47] Alejandro M Manelli and Daniel R Vincent. Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *Journal of Economic theory*, 137(1):153–185, 2007.

[48] Andreu Mas-Colell. On a theorem of schmeidler. *Journal of Mathematical Economics*, 13(3):201–206, 1984.

[49] Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002.

[50] Vahab Mirrokni and Hamid Nazerzadeh. Deals or no deals: Contract design for online advertising. 2015.

[51] Roger B Myerson. Incentive compatibility and the bargaining problem. *Econometrica*, pages 61–73, 1979.

[52] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

[53] Afshin Nikzad. Thickness and competition in ride-sharing markets. 2018. Working paper, Stanford University.

[54] Erhun Ozkan and Amy R Ward. Dynamic matching for real-time ridesharing. 2016. Working paper, Marshall School of Business, University of Southern California.

[55] Stephen G Powell and Kenneth L Schultz. Throughput in serial lines with state-dependent behavior. *Management Science*, 50(8):1095–1105, 2004.

[56] Josh Reed. The G/GI/N queue in the halfin–whitt regime. *The Annals of Applied Probability*, 19(6):2211–2269, 2009.

[57] John Riley and Richard Zeckhauser. Optimal selling strategies: When to haggle, when to hold firm. *The Quarterly Journal of Economics*, 98(2):267–289, 1983.

[58] Tim Roughgarden and Éva Tardos. How bad is selfish routing? *Journal of the ACM (JACM)*, 49(2):236–259, 2002.

[59] William Samuelson. Bargaining under asymmetric information. *Econometrica*, pages 995–1005, 1984.

[60] Filippo Santambrogio. Optimal transport for applied mathematicians, volume 87 of progress in nonlinear differential equations and their applications, 2015.

[61] David Schmeidler. Equilibrium points of nonatomic games. *Journal of statistical Physics*, 7(4):295–300, 1973.

[62] Samuel Shelbey. CRC standard mathematical tables. 1975.

[63] Wacław Sierpiński and C Cecilia Krieger. *General topology*. Number 7. University of Toronto Press Toronto, 1952.

[64] Kalyan T Talluri and Garrett J Van Ryzin. *The theory and practice of revenue management*, volume 68. Springer Science & Business Media, 2006.

[65] Terry Taylor. On-demand service platforms. *Manufacturing & Service Operations Management (forthcoming)*, 2017.

[66] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[67] Amy R Ward. Asymptotic analysis of queueing systems with reneging: A survey of results for fifo, single class models. *Surveys in Operations Research and Management Science*, 17(1):1–14, 2012.

[68] Ward Whitt. Efficiency-driven heavy-traffic approximations for many-server queues with abandonments. *Management Science*, 50(10):1449–1461, 2004.

[69] Ward Whitt. What you should know about queueing models to set staffing requirements in service systems. *Naval Research Logistics (NRL)*, 54(5):476–484, 2007.

Appendices

Appendix A

Surge Pricing and Its Spatial Supply Response

A.1 Proofs for Section 1.4

Proof of Lemma 1.1. Consider any $z, y \in \mathcal{C}$. Then, for essentially any $w \in \mathcal{B}$, we have

$$\begin{aligned} V_{\mathcal{B}}(y) \geq U(w) - \|w-y\| &= U(w) - \|z-w\| + \|z-w\| - \|w-y\| \\ &\geq U(w) - \|z-w\| - \|z-y\|, \end{aligned}$$

where the second inequality follows from the triangular inequality. This implies, by the definition of the essential supremum, that

$$V_{\mathcal{B}}(y) + \|z-y\| \geq V_{\mathcal{B}}(z).$$

Next, we would like to subtract $V_{\mathcal{B}}(y)$ from both sides of the previous inequality. This operation can be done only if $V_{\mathcal{B}}(y)$ is finite for any y in $\mathcal{C}$, but this is guaranteed by Lemma A.1 (stated and proved right after this proof). Hence, we obtain $V_{\mathcal{B}}(z) - V_{\mathcal{B}}(y) \leq \|z-y\|$. Since we can interchange the roles of z and y, we have proved that $|V_{\mathcal{B}}(z) - V_{\mathcal{B}}(y)| \leq \|z-y\|$, for all $z, y \in \mathcal{C}$.

□

Lemma A.1 *Consider a measurable set $\mathcal{B} \subseteq \mathcal{C}$ such that $\Gamma(\mathcal{B}) > 0$, let p be a measurable mapping $p : \mathcal{B} \to \mathbb{R}_+$, and let $\tau \in \mathcal{F}(\mu)$. Then, $V_{\mathcal{B}}(x|\, p, \tau) \in \left[-H, \alpha \cdot \overline{V}\right]$ for all $x \in \mathcal{C}$, where $H = \max_{x,y \in \mathcal{C}} \|x-y\|$. Furthermore, $V(x|\, p, \tau) \geq 0$ for all $x \in \mathrm{supp}(\Gamma)$.*

Proof of Lemma A.1. Fix $x \in \mathcal{C}$, we show that $V_{\mathcal{B}}(x|\, p, \tau) \in \left[-H, \alpha \cdot \overline{V}\right]$. For the lower bound, note that for any $y \in \mathcal{B}$, we have $U(y) - \|y-x\| \geq -H$.

Since $\Gamma(\mathcal{B}) > 0$, the definition of essential supremum implies that $V_{\mathcal{B}}(x|\, p, \tau) \geq -H$. Similarly, for the upper bound, note that for any $y \in \mathcal{B}$, $\alpha \cdot \overline{V} \geq U(y) - \|y - x\|$ and hence the definition of essential supremum yields $V_{\mathcal{B}}(x|\, p, \tau) \leq \alpha \cdot \overline{V}$.

Finally, we show that $V(x|\, p, \tau) \geq 0$ for all $x \in \text{supp}(\Gamma)$. Since $x \in \text{supp}(\Gamma)$ we have that $\Gamma(B(x,\delta)) > 0$ for all $\delta > 0$, where $B(x,\delta)$ is an open ball of radius δ. For any $y \in B(x,\delta)$ we have $U(y) - \|y - x\| > -\delta$, and since $\Gamma(B(x,\delta)) > 0$ we deduce that $V_{B(x,\delta)}(x|\, p, \tau) > -\delta$ for all $\delta > 0$. In turn, we have $V(x|\, p, \tau) \geq V_{B(x,\delta)}(x|\, p, \tau) > -\delta$ for all $\delta > 0$ and, therefore, $V(x|\, p, \tau) \geq 0$.

□

Proof of Proposition 1.1. We show how to reformulate the platform's objective as in the statement of the proposition. The key step is to establish that

$$U\left(x, p(x), s^{\tau}(x)\right) \;=\; V(x|\, p, \tau) \qquad \tau_2 - a.e.\ x \in \mathcal{C}, \tag{A.1}$$

namely, whenever there is post-relocation supply at a given location in equilibrium, the drivers originating at such a location can achieve maximum utility by staying at that location. We state and prove this result in Lemma A.2 (stated and proved following this proof). Note that this result holds $\tau_2 - a.e$ so if we want to interchange $U\left(x, p(x), s^{\tau}(x)\right)$ with $V(x|\, p, \tau)$ we have to do it under the measure τ_2. We next analyze the main term in the platform's objective function which we denote by $\mathbf{Rev}(p, \tau)$

$$\begin{aligned}
\mathbf{Rev}(p,\tau) &\overset{(a)}{=} \int_{\mathcal{C}_\lambda} p(y) \cdot \min\left\{ s^{\tau}(y), \overline{F}_y(p(y))\lambda(y) \right\} \mathbf{1}_{\{s^{\tau}(y)>0\}} d\Gamma(y) \\
&= \frac{1}{\alpha} \int_{\mathcal{C}} \alpha\ p(y) \cdot \min\left\{ 1, \frac{\overline{F}_y(p(y))\lambda(y)}{s^{\tau}(y)} \right\} \mathbf{1}_{\{s^{\tau}(y)>0\}} s^{\tau}(y) d\Gamma(y) \\
&= \frac{1}{\alpha} \int_{\mathcal{C}} U(y, p(y), s^{\tau}(y)) \mathbf{1}_{\{s^{\tau}(y)>0\}} s^{\tau}(y) d\Gamma(y) \\
&\overset{(b)}{=} \frac{1}{\alpha} \int_{\mathcal{C}} U(y, p(y), s^{\tau}(y)) \mathbf{1}_{\{s^{\tau}(y)>0\}} d\tau_2(y) \\
&\overset{(c)}{=} \frac{1}{\alpha} \int_{\mathcal{C}} V(y) \mathbf{1}_{\{s^{\tau}(y)>0\}} d\tau_2(y),
\end{aligned}$$

where (a) holds because whenever $\lambda(y) = 0$ or $s^\tau(y) = 0$, the minimum term in the integral becomes zero; (b) follows from the fact that $U(y, p(y), s^\tau(y))\mathbf{1}_{\{s^\tau(y)>0\}}$ is a measurable function with values in $[0, \alpha \cdot \overline{V}]$ and from recalling that $s^\tau = d\tau_2/d\Gamma$; and (c) is a consequence of Eq. (A.1) since we are integrating over the measure τ_2. In turn, focusing on the platform's objective function, this yields

$$\begin{aligned}(1-\alpha) \cdot \mathbf{Rev}(p, \tau) &= \gamma \int_{\mathcal{C}_\lambda} V(y)\mathbf{1}_{\{s^\tau(y)>0\}} d\tau_2(y) \\ &\overset{(a)}{=} \gamma \int_{\mathcal{C}_\lambda} V(y)\mathbf{1}_{\{s^\tau(y)>0\}} s^\tau(y) d\Gamma(y) \\ &= \gamma \int_{\mathcal{C}_\lambda} V(y) s^\tau(y) d\Gamma(y),\end{aligned}$$

where (a) holds because $V(y)\mathbf{1}_{\{s^\tau(y)>0\}}$ is measurable with values in $[0, \alpha \cdot \overline{V}]$ and we recall again that $s^\tau = d\tau_2/d\Gamma$. This completes the proof.

□

Lemma A.2 (Equilibrium Utilities) *For any price mapping p and corresponding equilibrium τ, let $\mathcal{B} \subseteq \mathcal{C}$ such that $\Gamma(\mathcal{B}) > 0$, then*

$$U(y, p(y), s^\tau(y)) \;=\; V_{\mathcal{B}}(y|\, p, \tau) = V(y|\, p, \tau) \qquad \tau_2 - a.e.\ y \in \mathcal{B}.$$

Furthermore,

$$U(y, p(y), s^\tau(y)) \;\leq\; V_{\mathcal{B}}(y|\, p, \tau) \qquad \Gamma - a.e.\ y \in \mathcal{B}.$$

Proof of Lemma A.2. We prove that

$$U(y, p(y), s^\tau(y)) \;=\; V_{\mathcal{B}}(y|\, p, \tau) \qquad \tau_2 - a.e.\ y \in \mathcal{B}.$$

The proof for $V(y|\, p, \tau)$ instead of $V_{\mathcal{B}}(y|\, p, \tau)$ follows the same steps and is, thus, omitted. Let $A \subseteq \mathcal{B}$ be a set defined by

$$A \triangleq \{y \in \mathcal{B} : U(y) = V_{\mathcal{B}}(y)\}. \tag{A.2}$$

We want to prove $\tau_2(A^c) = 0$, where the complement is taken with respect to $\mathcal{B}$. Consider the sets

$$A^- \triangleq \{y \in \mathcal{B} : U(y) < V_{\mathcal{B}}(y)\}, \quad A^+ \triangleq \{y \in \mathcal{B} : U(y) > V_{\mathcal{B}}(y)\}.$$

We will establish that $\tau_2(A^-) = 0$ and $\tau_2(A^+) = 0$. We begin with A^- and note that

$$\begin{aligned}
\tau_2(A^-) &= \tau(\mathcal{C} \times A^-) \\
&\overset{(a)}{=} \tau(\{(x,y) \in \mathcal{C} \times A^- : U(y) - \|y - x\| = V(x)) \\
&\overset{(b)}{\leq} \tau(\{(x,y) \in \mathcal{C} \times A^- : U(y) \geq V(y)) \\
&\overset{(c)}{\leq} \tau(\{(x,y) \in \mathcal{C} \times A^- : U(y) \geq V_{\mathcal{B}}(y)) \\
&\overset{(d)}{\leq} \tau(\{(x,y) \in \mathcal{C} \times \mathcal{B} : V_{\mathcal{B}}(y) > U(y) \geq V_{\mathcal{B}}(y)) \\
&= 0,
\end{aligned}$$

where (a) follows from the equilibrium definition, and (b) from the fact that $V(x) + \|x - y\| \geq V(y)$ (see Lemma 1.1). In (c) we have used $V(y) \geq V_{\mathcal{B}}(y)$, while (d) follows from $y \in A^-$ and $A^- \subseteq \mathcal{B}$.

To show that $\tau_2(A^+) = 0$, it suffices to show that show $\Gamma(A^+) = 0$ (this will also show the last statement of the lemma). For any $n \in \mathbb{N}$ define the set $A_n^+ \triangleq \{y \in \mathcal{B} : U(y) \geq V_{\mathcal{B}}(y) + \frac{1}{n}\}$, and note that $A^+ = \bigcup_{n\in\mathbb{N}} A_n^+$. It is enough to show that $\Gamma(A_n^+) = 0$ for all $n \in \mathbb{N}$. We proceed by contradiction. Suppose there exists $n \in \mathbb{N}$ such that $\Gamma(A_n^+) > 0$. Let $\epsilon > 0$ be such that $\epsilon < \frac{1}{2n}$, and consider a finite partition $\{I_i^\epsilon\}_{i=1}^{K(\epsilon)}$ of $\mathcal{C}$, where for any $x, y \in I_i^\epsilon$ we have $\|x - y\| \leq \epsilon$. Observe that

$$0 < \Gamma(A_n^+) = \Gamma(A_n^+ \cap \bigcup_{i=1}^{K(\epsilon)} I_i^\epsilon) = \sum_{i=1}^{K(\epsilon)} \Gamma(A_n^+ \cap I_i^\epsilon),$$

therefore, there exists $i \in \{1, \ldots, K(\epsilon)\}$ such that $\Gamma(A_n^+ \cap I_i^\epsilon) > 0$. Take $x \in I_i^\epsilon$, then for any $y \in A_n^+ \cap I_i^\epsilon$

$$U(y) \geq V_{\mathcal{B}}(y) + \frac{1}{n} \geq V_{\mathcal{B}}(x) - \|y - x\| + \frac{1}{n} > V_{\mathcal{B}}(x) - \|y - x\| + 2\epsilon \geq V_{\mathcal{B}}(x) + \|y - x\|,$$

where the second inequality comes from the Lipschitz property (see Lemma 1.1). The last two inequalities hold because of our choice of ϵ and $x, y \in I_i^\epsilon$. We conclude that

$$A_n^+ \cap I_i^\epsilon \subseteq \{y \in \mathcal{B} : \Pi(x,y) > V_{\mathcal{B}}(x)\}.$$

This would therefore imply that $\Gamma(\{y \in \mathcal{B} : \Pi(x,y) > V_{\mathcal{B}}(x)\}) > 0$. However, this contradicts the definition of $V_{\mathcal{B}}(x)$. Hence we must have $\Gamma(A_n^+) = 0$ for all $n \in \mathbb{N}$, and in turn $\Gamma(A^+) = 0$.

□

Proof of Lemma 1.2. For ease of notation let us use x_a to denote $X_a(z|\,p,\tau)$. We also denote $A(z|p,\tau)$ by $A(z)$.

Closure: Let the sequence $\{x^n\}_{n\in\mathbb{N}} \subset A(z)$ be such that $x^n \to x$. We show that $x \in A(z)$, that is,

$$\lim_{\delta\downarrow 0} V_{B(z,\delta)}(x) = V(x). \tag{A.3}$$

Since $x^n \in A(z)$, $z \in \mathcal{IR}(x^n)$ for all $n \in \mathbb{N}$, i.e.,

$$\lim_{\delta\downarrow 0} V_{B(z,\delta)}(x^n) = V(x^n), \quad \forall n \in \mathbb{N}. \tag{A.4}$$

Note that Eq. (A.4) implies that $\Gamma(B(z,\delta)) > 0$ for all $\delta > 0$; otherwise, $V_{B(z,\delta)}(\cdot)$ would be $-\infty$ and so the limit would not be well defined. We next establish Eq. (A.3) from first principles. Fix $\epsilon > 0$. Since x^n converges to x we can find $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have $\|x^n - x\| \leq \frac{\epsilon}{3}$. In particular, from Eq. (A.4) applied to n_0 we deduce that

$$\exists \delta_0 > 0, \text{ such that } \forall \delta \leq \delta_0,\ \frac{\epsilon}{3} + V_{B(z,\delta)}(x^{n_0}) \geq V(x^{n_0}). \tag{A.5}$$

Using the Lipchitz property of $V_{B(z,\delta)}(\cdot)$ and $V(\cdot)$ (see Lemma 1.1), and that $\|x^{n_0} - x\| \leq \frac{\epsilon}{3}$ yields

$$V_{B(z,\delta)}(x^{n_0}) - V_{B(z,\delta)}(x) \leq \|x^{n_0} - x\| \leq \frac{\epsilon}{3} \text{ and } V(x^{n_0}) - V(x) \geq -\|x^{n_0} - x\| \geq -\frac{\epsilon}{3}.$$

In turn, using Eq. (A.5), we have that for all $\delta \leq \delta_0$,

$$V(x) - \epsilon \leq V(x^{n_0}) - \frac{2\epsilon}{3} \leq V_{B(z,\delta)}(x) \leq V(x).$$

Since ϵ was arbitrary, we deduce that Eq. (A.3) holds, and therefore, $x \in A(z)$.

Interval: We show that $A_a(z) = [z, x_a]$. The definition of x_a immediately implies that $A_a(z) \subseteq [z, x_a]$, so we only need to prove the reverse inclusion. First, since we can always construct a sequence $\{x^n\}_{n\in\mathbb{N}} \subset A(z)$, with $x^n \to x_a$, the closure property implies that $x_a \in A(z)$. Second, we make use of Lemma A.3 (stated and proved right after this proof). Consider $x \in [z, x_a]$ then Lemma A.3 implies that $z \in \mathcal{IR}(x|\, p, \tau)$ or, equivalently, $x \in A_a(z)$.

Union: Since for every $a \in R_z$ we have $A_a(z) \subset A(z)$, the same is true for the union. In the opposite direction, if we take $x \in A(z)$ then there exists $a \in R_z$ such that $x \in [z, x_\theta] = A_a(z)$.

□

Lemma A.3 *For any price mapping p and corresponding equilibrium τ, if $y \in \mathcal{IR}(x|\, p, \tau)$ then $y \in \mathcal{IR}(z|\, p, \tau)$ for all $z \in [x \wedge y, x \vee y]$.*

Proof of Lemma A.3. Let $y \in \mathcal{IR}(x|\, p, \tau)$. If $x = y$ there is nothing to prove. Without loss of generality, suppose $x < y$. Since $y \in \mathcal{IR}(x|\, p, \tau)$ we have that $\lim_{\delta\downarrow 0} V_{B(y,\delta)}(x) = V(x)$. Observe that this implies that $\Gamma(B(y,\delta)) > 0$ for all $\delta > 0$; if this is not true, $V_{B(y,\delta)}(x) = -\infty$ and the limit would not be well defined. Next, fix $z \in [x, y]$, we want to prove that $y \in \mathcal{IR}(z|\, p, \tau)$, i.e., $\lim_{\delta\downarrow 0} V_{B(y,\delta)}(z) = V(z)$ or equivalently we need to show that

$$\forall \epsilon > 0,\ \exists \delta_0 > 0 \text{ such that } \forall \delta \le \delta_0,\ V_{B(y,\delta)}(z) + \epsilon \ge V(z). \tag{A.6}$$

Consider $\epsilon > 0$ and $\delta_1 > 0$ such that $x \notin B(y, \delta_1)$ $(x < y)$, and note that since $y \in \mathcal{IR}(x|\, p, \tau)$ we can find $\delta_0 > 0$ such that

$$V(x) \le V_{B(y,\delta)}(x) + \frac{\epsilon}{3}, \quad \forall \delta \le \delta_0.$$

Consider $\delta \le \min\{\delta_1, \delta_0, \frac{\epsilon}{6}\}$, then the previous inequality implies

$$U(w) - \|w - x\| \le V_{B(y,\delta)}(x) + \frac{\epsilon}{3}, \qquad \Gamma - a.e.\ w \text{ in } \mathcal{C}. \tag{A.7}$$

Note that since $z \in [x, y]$, for any $y' \in B(y, \delta)$ we have

$$\|y' - x\| - \|y' - z\| \ge -2\delta + \|z - x\|,$$

and, therefore,

$$\min_{y' \in B(y,\delta)} \{\|y' - x\| - \|y' - z\|\} \geq -2\delta + \|z - x\|.$$

This and Lemma A.4 (which we state and prove after the present proof) deliver

$$V_{B(y,\delta)}(z) \geq V_{B(y,\delta)}(x) - \frac{\epsilon}{3} - 2\delta + \|z - x\|.$$

This inequality together with Eq. (A.7) yield:

$$V_{B(y,\delta)}(z) + \frac{\epsilon}{3} + 2\delta - \|z - x\| \geq U(w) - \|w - x\| - \frac{\epsilon}{3}, \qquad \Gamma - a.e. \ w \ \text{ in } \mathcal{C}.$$

Then, $\Gamma - a.e. \ w$ in $\mathcal{C}$ we have

$$\begin{aligned} V_{B(y,\delta)}(z) + \frac{2}{3}\epsilon + 2\delta &\geq U(w) - \|w - x\| + \|z - x\| \\ &= U(w) - \|w - z\| + \|w - z\| - \|w - x\| + \|z - x\| \\ &\geq U(w) - \|w - z\| - \|z - x\| + \|z - x\| \\ &= U(w) - \|w - z\|, \end{aligned}$$

implying that $V_{B(y,\delta)}(z) + \frac{2}{3}\epsilon + 2\delta \geq V(z)$. Since $2\delta \leq \frac{\epsilon}{3}$ we conclude that $V_{B(y,\delta)}(z) + \epsilon \geq V(z)$.

□

Lemma A.4 *Let* $\epsilon, \delta > 0$ *and* $x, y, z \in \mathcal{C}$. *If* $\Gamma(B(y,\delta)) > 0$ *then*

$$V_{B(y,\delta)}(z) \geq V_{B(y,\delta)}(x) - \epsilon + \min_{y' \in B(y,\delta)} \{\|y' - x\| - \|y' - z\|\},$$

Proof of Lemma A.4. Define the following set

$$R \triangleq \Big\{ y' \in B(y,\delta) : \Pi(x, y') \geq V_{B(y,\delta)}(x) - \epsilon \Big\}.$$

Note that $\Gamma(R) > 0$. If not, we could find a lower essential upper bound in $B(y,\delta)$.

Let $y' \in R$

$$\begin{aligned}\Pi(z,y') &= U(y') - \|y'-z\| - \|y'-x\| + \|y'-x\| \\ &= \Pi(x,y') - \|y'-z\| + \|y'-x\| \\ &\geq V_{B(y,\delta)}(x) - \epsilon - \|y'-z\| + \|y'-x\| \\ &\geq V_{B(y,\delta)}(x) - \epsilon + \min_{y'\in B(y,\delta)} \{\|y'-x\| - \|y'-z\|\}.\end{aligned}$$

Since $\Gamma(R) > 0$ we must have that

$$V_{B(y,\delta)}(z) \geq \Pi(z,y') \qquad \Gamma-\ a.e\ y' \in R.$$

Putting the last two inequalities together yields the desired result.

□

Proof of Lemma 1.3. Consider $x \in A(z)$, that is, $z \in \mathcal{IR}(x|\, p,\tau)$. We next establish that $V(x) = V(z) - \|z-x\|$. First, by the Lipschitz property of V we have $V(z) \leq V(x) + \|z-x\|$. So we only need to prove the opposite inequality. Fix $\epsilon > 0$. Since $z \in \mathcal{IR}(x)$ we can find $\delta_1(\epsilon) > 0$ such that

$$\frac{\epsilon}{2} + V_{B(z,\delta)}(x) \geq V(x), \quad \forall \delta \leq \delta_1(\epsilon), \tag{A.8}$$

and $V_{B(z,\delta)}(\cdot)$ takes finite values. Define the set

$$R^{x,\delta,\epsilon} \triangleq \{y' \in B(z,\delta) : U(y') - \|y'-x\| > V(x) - \epsilon\}.$$

Suppose there exists $\delta \leq \delta_1(\epsilon)$ such that $\Gamma(R^{x,\delta,\epsilon}) = 0$. This would imply that $V(x) - \epsilon \geq V_{B(z,\delta)}(x)$, which together with Eq. (A.8) yields a contradiction. Hence for all $\delta \leq \delta_1(\epsilon)$, $\Gamma(R^{x,\delta,\epsilon}) > 0$.

Fix $\delta \leq \delta_1(\epsilon)$. Next we verify that $V_{R^{x,\delta,\epsilon}}(z) \geq V(x) - \epsilon + \|z-x\| - 2\delta$. For any $y' \in R^{x,\delta,\epsilon}$

$$\begin{aligned}U(y') - \|y'-z\| &= U(y') - \|y'-x\| + \|y'-x\| - \|y'-z\| \\ &\geq V(x) - \epsilon + \|y'-x\| - \|y'-z\| \\ &\geq V(x) - \epsilon + \|z-x\| - 2\delta,\end{aligned}$$

were the last inequality follows from the triangular inequality. From the definition of the essential supremum we deduce that $V_{R^{x,\delta,\epsilon}}(z) \geq V(x) - \epsilon + \|z - x\| - 2\delta$. Because $V(z) \geq V_{R^{x,\delta,\epsilon}}(z)$ we must have $V(z) \geq V(x) + \|z - x\| - \epsilon - 2\delta$. We selected $\delta \leq \delta_1(\epsilon)$, and ϵ arbitrarily. So we can let $\delta \downarrow 0$ and then $\epsilon \downarrow 0$ to obtain that $V(z) \geq V(x) + \|z - x\|$. This concludes the proof.

□

Proof of Proposition 1.2. Consider the segment $[x, y]$ and define the set

$$L \triangleq \{y' \in \mathcal{C} : \exists t \geq 0 \text{ such that } y' = x + t \cdot (y - x)\},$$

that is L is the set of point along the ray that starts at x and contains the segment $[x, y]$. Since $y \in L$ and $y \in \mathcal{IR}(x|\, p, \tau)$ the following quantity is well defined

$$z \triangleq \sup\{y' \in L : y' \in \mathcal{IR}(x|\, p, \tau)\}.$$

We prove that z is a sink location such that $x, y \in A(z|\, p, \tau)$. First, we show that $z \in \mathcal{IR}(x|p, \tau)$. Consider a sequence $\{z_n\} \subset L$ such that $z_n \in \mathcal{IR}(x|p, \tau)$ and $z_n \to z$. Fix $\epsilon > 0$, $\hat{\delta} > 0$ and choose n such that $\|z_n - z\| < \hat{\delta}/2$. Since $z_n \in \mathcal{IR}(x|p, \tau)$ then there exists $\delta_0(n, \epsilon) > 0$ such that for all $\delta \leq \delta_0(n, \epsilon)$ we have $V_{B(z_n,\delta)}(x) \geq V(x) - \epsilon$. In particular, for any $\delta \leq \min\{\delta_0(n, \epsilon), \hat{\delta}/2\}$ we have $B(z_n, \delta) \subseteq B(z, \hat{\delta})$ and, therefore,

$$V_{B(z,\hat{\delta})}(x) \geq V_{B(z_n,\delta)}(x) \geq V(x) - \epsilon.$$

Since the choice of ϵ and $\hat{\delta}$ was arbitrary we conclude that $\lim_{\hat{\delta}\downarrow 0} V_{B(z,\hat{\delta})}(x) = V(x)$. That is, $z \in \mathcal{IR}(x|\, p, \tau)$ which also shows that $A(z) \neq \emptyset$.

Next, to show that z is a sink location we argue that we cannot have $z \in A(z')$ for some $z' \neq z$. If we did then $z' \in \mathcal{IR}(z|\, p, \tau)$ for some $z' \neq z$. First suppose that $z' \in L$. If $z' > z$ this would contradict the definition of z as being maximal. If $z' < z$ then by Lemma 1.3 the function $V(\cdot)$ would be decreasing in (z', z), and by the same lemma it would be increasing in (x, z). This is a contradiction.

Second, suppose that $z' \notin L$. That is the vectors $z'-x$ and $z-x$ are not collinear. Fix $\epsilon > 0$, since $z' \in \mathcal{IR}(z|\, p, \tau)$ we can find $\delta(\epsilon) > 0$ such that $V_{B(z',\delta)}(z) \geq V(z) - \epsilon$ for all $\delta \leq \delta(\epsilon)$. Moreover, from $z' \neq z$ and the no collinearity property we have that $\|x - z'\| + \gamma \leq \|x - z\| + \|z - z'\|$ for some $\gamma > 0$ sufficiently small. Hence, if we take $\delta \leq \min\{\delta(\epsilon), \gamma/3\}$ we deduce that

$$\begin{aligned}
V_{B(z',\delta)}(x) &\geq U(w) - \|w - x\| \\
&= U(w) - \|w - z\| + \|w - z\| - \|w - x\| \\
&\overset{(a)}{\geq} U(w) - \|w - z\| + \|w - z\| - \|w - z'\| - \|z' - x\| \\
&\overset{(b)}{\geq} U(w) - \|w - z\| + \|w - z\| - \frac{\gamma}{3} - \|x - z\| - \|z - z'\| + \gamma \\
&\overset{(c)}{\geq} U(w) - \|w - z\| - \|x - z\| + \frac{\gamma}{3} \qquad \Gamma - \ a.e \ w \in B(z', \delta),
\end{aligned}$$

where in (a) we use the triangular inequality, in (b) we use that $\|w - z'\| \leq \gamma/3$ and that $\|x - z'\| + \gamma \leq \|x - z\| + \|z - z'\|$, and in (c) we use that $\|w - z'\| \leq \gamma/3$ and the triangular inequality. Therefore, $V_{B(z',\delta)}(x) + \|x - z\| - \frac{\gamma}{3} \geq V_{B(z',\delta)}(z)$. In turn, this yields $V_{B(z',\delta)}(x) + \|x - z\| - \frac{\gamma}{3} \geq V(z) - \epsilon$. Since $V(x) \geq V_{B(z',\delta)}(x)$ and because $x \in A(z)$ this implies that $V(x) = V(z) - \|x - z\|$ (see Lemma 1.3) we deduce that $V(x) + \|x - z\| - \frac{\gamma}{3} \geq V(x) + \|x - z\| - \epsilon$. Taking $\epsilon > 0$ small enough yields a contradiction. We conclude that z is a sink location. Moreover, because $x \in A(z)$ ($z \in \mathcal{IR}(x|\, p, \tau)$) and $x < y \leq z$ (recall these three points are collinear) Lemma A.3 guarantees that $y \in A(z)$.

□

Proof of Proposition 1.3. With some abuse of notation let

$$A^{\circ}(z|\, p, \tau) = \bigcup_{a \in R_z} (z, X_a(z|\, p, \tau).$$

This result is based on the following properties:

a) For all $(x, y) \in A(z|\, p, \tau)^c \times A(z|\, p, \tau)$, $y \notin \mathcal{IR}(x|\, p, \tau)$.

b) For all $(x,y) \in (A^{\circ}(z|p,\tau)\cup\{z\})\times\Big(A(z|p,\tau)^c\cup L(z|p,\tau)\setminus\{z\}\Big)$, $y \notin \mathcal{IR}(x|p,\tau)$.

Before we provide a formal proof of these properties, we use them to show the statement of the proposition. We will also make use of Lemma A.5 which we prove and state after the present proof.

We begin with the first part of *(i)*, that is, we show that $\tau(A(z|p,\tau)^c\times A(z|p,\tau)) = 0$. If this is not true then by Lemma A.5 we can find $(x,y) \in A(z|p,\tau)^c\times A(z|p,\tau)$ such that $y \in \mathcal{IR}(x|\,p,\tau)$. We obtain a contradiction with property a) above. Therefore it must be the case that $\tau(A(z|\,p,\tau)^c \times A(z|\,p,\tau)) = 0$.

Next, we show the second part of *(i)*, namely, $\tau((A^{\circ}(z|\,p,\tau)\cup\{z\})\times\Big(A(z|\,p,\tau)^c\cup L(z|\,p,\tau)\setminus\{z\}\Big)) = 0$. If this is not true then by Lemma A.5 we can find $(x,y) \in (A^{\circ}(z|\,p,\tau)\cup\{z\})\times\Big(A(z|\,p,\tau)^c\cup L(z|\,p,\tau)\setminus\{z\}\Big)$ such that $y \in \mathcal{IR}(x|\,p,\tau)$ but this contradicts property b) above. Therefore it must be the case that $\tau((A^{\circ}(z|\,p,\tau)\cup\{z\})\times\Big(A(z|\,p,\tau)^c\cup L(z|\,p,\tau)\setminus\{z\}\Big)) = 0$.

Now we provide a proof for *(ii)*. Let $R_1, R_2 \subset R_z$ with $R_1 \cap R_2 = \emptyset$ we show that

$$\tau\Big(\bigcup_{a\in R_1}(z, X_a(z|\,p,\tau)]\times\bigcup_{a\in R_2}(z, X_a(z|\,p,\tau)]\Big) = 0.$$

Suppose by contradiction that this is not true then by Lemma A.5 we can find $(x,y) \in \bigcup_{a\in R_1}(z, X_a(z|\,p,\tau)]\times\bigcup_{a\in R_2}(z, X_a(z|\,p,\tau)]$ such that $y \in \mathcal{IR}(x|\,p,\tau)$. This implies that $x \in A(y|\,p,\tau)$. Moreover, since z is a sink location we have $x \in A(z|\,p,\tau)$ and $y \in A(z|\,p,\tau)$. We use Lemma 1.3 to infer that

$$V(x) = V(y) - \|y-x\|, \quad V(x) = V(z) - \|z-x\|, \quad \text{and} \quad V(y) = V(z) - \|z-y\|.$$

In turn, we can use the first to equalities to obtain $V(y) = V(z) + \|y-x\| - \|z-x\|$. Plugging this into the last equality yields $\|z-x\| = \|y-x\|+\|z-y\|$; however, because $R_1 \cap R_2 = \emptyset$ we have that $x \in (z, X_{a_1}(z|\,p,\tau)]$ and $y \in (z, X_{a_2}(z|\,p,\tau)]$ with $a_1 \neq a_2$. In other words, x and y belong to different rays around z. In turn, the latter equality cannot hold and we must have that $\tau\Big(\bigcup_{a\in R_1}(z, X_a(z|p,\tau)]\times\bigcup_{a\in R_2}(z, X_a(z|p,\tau)]\Big) = 0$.

Next we verify properties a) and b). We start with a). We argue by contradiction. Suppose there exists $x \in A(z|\, p,\tau)^c$ and $y \in A(z|\, p,\tau)$ such that $y \in \mathcal{IR}(x|\, p,\tau)$. Let a index the ray that contains the vector $(x-z)$. Recall that by Lemma 1.2 we have that $A_a(z|\, p,\tau) = [z, X_a(z|\, p,\tau)]$. Since $x \in A(z|\, p,\tau)^c$ we must have that $x \notin [z, X_a(z|\, p,\tau)]$. In particular $\|x-z\| > |X_a(z|\, p,\tau) - z|$. Hence if we show that $z \in \mathcal{IR}(x|\, p,\tau)$ we would contradict the maximality of $X_a(z|\, p,\tau)$. Fix $\epsilon > 0$, then from $z \in \mathcal{IR}(y|\, p,\tau)$ we can always find $\delta_0 > 0$ such that for all $\delta \leq \delta_0$

$$\epsilon + V_{B(z,\delta)}(y) \geq V(y). \tag{A.9}$$

By the Lipschitz property $V_{B(z,\delta)}(x) + \|y-x\| \geq V_{B(z,\delta)}(y)$. Hence, from Eq. (A.9) we can deduce that $V_{B(z,\delta)}(x) + \|y-x\| \geq V(y) - \epsilon$. Also, because $y \in \mathcal{IR}(x|\, p,\tau)$ or, equivalently, $x \in A(y|\, p,\tau)$ Lemma 1.3 yields $V(x) = V(y) - \|y-x\|$. Hence, $V_{B(z,\delta)}(x) \geq V(x) - \epsilon$, that is, $z \in \mathcal{IR}(x|\, p,\tau)$.

Now we show b). Let $x \in (A^{\circ}(z|p,\tau) \cup \{z\})$ and $y \in \Big(A(z|p,\tau)^c \cup L(z|p,\tau) \setminus \{z\}\Big)$. We look into two cases: $x \neq z$ and $x = z$. In both cases we proceed by contradiction assuming that $y \in \mathcal{IR}(x|\, p,\tau)$. Let us start with $x \neq z$. Let a index the ray that contains the vector $(x-z)$. Recall that by Lemma 1.2 we have that $A_a(z|\, p,\tau) = [z, X_a(z|p,\tau)]$. Since, $x \in A^{\circ}(z|p,\tau)$ and $x \neq z$ we must have that $x \in (z, X_a(z|p,\tau))$. Lemma 1.3 delivers

$$V(x) = V(y) - \|y-x\| \text{ and } V(X_a(z|\, p,\tau)) = V(x) - \|x - X_a(z|\, p,\tau)\|,$$

that is,

$$V(y) - V(X_a(z|\, p,\tau)) = \|y-x\| + \|x - X_a(z|\, p,\tau)\|. \tag{A.10}$$

If $y = X_a(z|\, p,\tau)$ the previous equality implies $x = X_a(z|\, p,\tau)$, but since $x \in (z, X_a(z|\, p,\tau))$ this is not possible. If $y \neq X_a(z|\, p,\tau)$ then since $y \in \Big(A(z|\, p,\tau)^c \cup L(z|\, p,\tau) \setminus \{z\}\Big)$ we must have that $y \notin (z, X_a(z|\, p,\tau))$. Also, y cannot be equal to some point $x + t(z-x)$ for some $t > 1$ because that would contradict the fact that

z is a sink location. Therefore, Eq. (A.10) together with the triangular inequality deliver $V(y) - V(X_a(z|\, p, \tau)) > |y - X_a(z|\, p, \tau)|$, but this contradicts the Lipschitz property of $V(\cdot)$.

To conclude, consider the case $x = z$. In this case we would have $z \in A(y|\, p, \tau)$ but this contradicts the fact that z is a sink location.

□

Lemma A.5 *Let $\mathcal{L}_1, \mathcal{L}_2 \subset \mathcal{C}$. If $\tau(\mathcal{L}_1 \times \mathcal{L}_2) > 0$ then there exists $(x, y) \in \mathcal{L}_1 \times \mathcal{L}_2$ such that $y \in \mathcal{IR}(x|p, \tau)$.*

Proof of Lemma A.5. Suppose $\tau(\mathcal{L}_1 \times \mathcal{L}_2) > 0$. We first argue that there exists a pair $(x, y) \in \mathcal{L}_1 \times \mathcal{L}_2$ such that for all $\delta > 0$

$$\tau(B(x, \delta) \times B(y, \delta)) > 0. \tag{A.11}$$

If this is not true then for any $(x, y) \in \mathcal{L}_1 \times \mathcal{L}_2$ we can find $\delta_{x,y} > 0$ such that Eq. (A.11) does not hold when we replace δ with $\delta_{x,y}$, that is, $\tau(B(x, \delta_{x,y}) \times B(y, \delta_{x,y})) = 0$ for all $(x, y) \in \mathcal{L}_1 \times \mathcal{L}_2$. The collection $\mathcal{I}$ defined by

$$\mathcal{I} = \{B(x, \delta_{x,y}) \times B(y, \delta_{x,y})\}_{(x,y) \in \mathcal{L}_1 \times \mathcal{L}_2}$$

is an open cover of $\mathcal{L}_1 \times \mathcal{L}_2$. Moreover the set $\mathcal{L}_1 \times \mathcal{L}_2$ is separable because $\mathcal{C} \times \mathcal{C}$ is separable. This implies that we can find a countable sub-cover of $\mathcal{L}_1 \times \mathcal{L}_2$ in $\mathcal{I}$, that is, there exists $\{B(x_n, \delta_{x_n,y_n}) \times B(y_n, \delta_{x_n,y_n})\}_{n \in \mathbb{N}}$ such that

$$\mathcal{L}_1 \times \mathcal{L}_2 \subset \bigcup_{n \in \mathbb{N}} B(x_n, \delta_{x_n,y_n}) \times B(y_n, \delta_{x_n,y_n}).$$

The existence of the sub-cover is guaranteed by the Lindelöf property of separable metric spaces, see e.g., [63] Theorem 69, p. 116. Since τ is a measure we have

$$\begin{aligned}
\tau(\mathcal{L}_1 \times \mathcal{L}_2) &\leq \tau\Big(\bigcup_{n \in \mathbb{N}} B(x_n, \delta_{x_n,y_n}) \times B(y_n, \delta_{x_n,y_n})\Big) \\
&\leq \sum_{n \in \mathbb{N}} \tau(B(x_n, \delta_{x_n,y_n}) \times B(y_n, \delta_{x_n,y_n})) \\
&= 0,
\end{aligned}$$

a contradiction. Therefore, for some $(x, y) \in \mathcal{L}_1 \times \mathcal{L}_2$, Eq. (A.11) holds for any $\delta > 0$.

We next show that $y \in \mathcal{IR}(x)$, that is,

$$\forall \epsilon > 0, \exists \delta_0 > 0 \;\text{ such that }\; \forall \delta < \delta_0 \quad \epsilon + V_{B(y,\delta)}(x) \geq V(x).$$

Let $\epsilon > 0$ and let $\delta_0 = \frac{\epsilon}{2}$. Consider $\delta < \delta_0$, from Eq. (A.11) and the equilibrium definition we have

$$\begin{aligned}
&0 < \tau(B(x,\delta) \times B(y,\delta)) \\
&\quad = \tau\Big(\Big\{(x', y') \in B(x,\delta) \times B(y,\delta) : \Pi(x', y') = V(x')\Big\}\Big) \\
&\quad \leq \tau_2\Big(\underbrace{\Big\{y' \in B(y,\delta) : \exists x' \in B(x,\delta) \;\text{ such that }\; \Pi(x', y') = V(x')\Big\}}_{\triangleq R^{x,y,\delta}}\Big),
\end{aligned}$$

since $\tau_2 \ll \Gamma$ this implies that $\Gamma(R^{x,y,\delta}) > 0$. Now we argue that $R^{x,y,\delta} \subset \{y' \in B(y,\delta) : \Pi(x, y') \geq V(x) - \epsilon\}$. Indeed, let $y' \in R^{x,y,\delta}$ then there exists $x' \in B(x,\delta)$ for which

$$\begin{aligned}
U(y') &= V(x') + \|y' - x'\| \\
&\geq V(x) - \|x' - x\| + \|y' - x'\| \\
&= V(x) - \|x' - x\| + \|y' - x'\| - \|y' - x\| + \|y' - x\| \\
&\geq V(x) - \|x' - x\| - \|x' - x\| + \|y' - x\|,
\end{aligned}$$

where in the first inequality we used the Lipchitz property of V (see Lemma 1.1), and in the second we use triangular inequality. Since $\|x' - x\| \leq \delta_0 = \frac{\epsilon}{2}$ we have that $U(y') \geq V(x) - \epsilon + \|y' - x\|$, that is, $R^{x,y,\delta} \subset \{y' \in B(y,\delta) : \Pi(x, y') \geq V(x) - \epsilon\}$. Therefore, $\Gamma(\{y' \in B(y,\delta) : \Pi(x, y') \geq V(x) - \epsilon\}) > 0$, which implies that $V_{B(y,\delta)}(x) \geq V(x) - \epsilon$.

□

<u>Proof of Proposition 1.4.</u> For ease of notation we use X_a to denote $X_a(z|p,\tau)$. We show that $\hat{\tau}$ belongs to $\mathcal{F}_{\mathcal{C}}(\mu)$ and that it is an equilibrium in $\mathcal{C}$. First we argue that $\hat{\tau} \in \mathcal{F}_{\mathcal{C}}(\mu)$. Since $\hat{\tau}$ is the sum of two non-negative measures we have that

$\hat{\tau} \in \mathcal{M}(\mathcal{C} \times \mathcal{C})$. In order see why $\hat{\tau}_1$ coincides with μ, let B be a measurable subset of $\mathcal{C}$ then

$$
\begin{aligned}
\hat{\tau}_1(B) &= \hat{\tau}(B \times \mathcal{C}) \\
&= \tau((B \cap (\mathcal{A}^c \cup \mathcal{L})) \times \mathcal{A}^c) + \tilde{\tau}((B \cap \mathcal{A}) \times \mathcal{A}) \\
&\overset{(a)}{=} \tau((B \cap (\mathcal{A}^c \cup \mathcal{L})) \times \mathcal{A}^c) + \tilde{\mu}(B \cap \mathcal{A}) \\
&= \tau((B \cap (\mathcal{A}^c \cup \mathcal{L})) \times \mathcal{A}^c) + \tau((B \cap \mathcal{A}) \times \mathcal{A}) \\
&\overset{(b)}{=} \tau((B \cap \mathcal{A}^c) \times \mathcal{A}^c) + \tau((B \cap \mathcal{L}) \times \mathcal{A}^c) + \tau((B \cap \mathcal{A}) \times \mathcal{A}) \\
&\overset{(c)}{=} \tau((B \cap \mathcal{A}^c) \times \mathcal{C}) + \tau((B \cap \mathcal{A}) \times \mathcal{C}) \\
&= \mu(B),
\end{aligned}
$$

where (a) comes from the fact that $\tilde{\tau}$ belongs to $\mathcal{F}_{\mathcal{A}}(\tilde{\mu})$. In (b) we use the fact that $\mathcal{A}$ is a closed set. Equality (c) comes from Proposition 1.3 part *(i)*. That is, $\hat{\tau}_1$ coincides with μ. Now, we show that $\hat{\tau}_2 \ll \Gamma$. Let B be as before and suppose $\Gamma(B) = 0$ then

$$
\begin{aligned}
\hat{\tau}_2(B) = \hat{\tau}(\mathcal{C} \times B) &= \tau((\mathcal{A}^c \cup \mathcal{L}) \times (B \cap \mathcal{A}^c)) + \tilde{\tau}(\mathcal{A} \times (B \cap \mathcal{A})) \\
&\leq \tau_2(B \cap \mathcal{A}^c) + \tilde{\tau}_2(B \cap \mathcal{A}) \\
&= 0,
\end{aligned}
$$

where the last equality holds because $\tau_2 \ll \Gamma$ and $\tilde{\tau}_2 \ll \Gamma|_{\mathcal{A}}$. Now we show that $\hat{\tau}$ is an equilibrium. We need to verify that $\hat{\tau}(\hat{\mathcal{E}})$ equals $\mu(\mathcal{C})$, where

$$
\hat{\mathcal{E}} \triangleq \left\{(x, y) \in \mathcal{C} \times \mathcal{C} : \Pi(x, y, \hat{p}(y), s^{\hat{\tau}}(y)) = \operatorname*{ess\,sup}_{\mathcal{C}} \Pi\Big(x, \cdot, \hat{p}(\cdot), s^{\hat{\tau}}(\cdot)\Big)\right\}.
$$

In order to verify this we compute first $s^{\hat{\tau}}$ and $V(x|\,\hat{p}, \hat{\tau})$. First we show that $\Gamma-$a.e we have

$$
s^{\hat{\tau}}(x) = \begin{cases} s^{\tau}(x) & \text{if } x \in \mathcal{A}^c \\ s^{\tilde{\tau}}(x) & \text{if } x \in \mathcal{A}. \end{cases}
$$

Let B be a measurable subset of $\mathcal{A}^c$ then

$$
\hat{\tau}_2(B) = \tau((\mathcal{C} \times B) \cap ((\mathcal{A}^c \cup \mathcal{L}) \times \mathcal{A}^c)) = \tau((\mathcal{A}^c \cup \mathcal{L}) \times B) \overset{(a)}{=} \tau(\mathcal{C} \times B) = \tau_2(B),
$$

where (a) comes from Proposition 1.3 part *(i)*. Therefore, $s^{\hat{\tau}}(x)$ equals $s^{\tau}(x)$ $\Gamma - a.e.$ x in $\mathcal{A}^c$. Similarly, for B a measurable subset of $\mathcal{A}$ we have

$$\hat{\tau}_2(B) = \tilde{\tau}(\mathcal{A} \times B) = \tilde{\tau}_2(B),$$

where the second equality holds because from Proposition 1.3 we have $\tau(\mathcal{A}^c \times \mathcal{A}) = 0$, and also because $\tilde{\tau}$ is an equilibrium in $\mathcal{A}$.

Second, we show that $V(x|\,\hat{p},\hat{\tau})$ equals $V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau})$ for all $x \in \mathcal{A}$. Let $x \in \mathcal{A}$, by definition

$$V(x|\,\hat{p},\hat{\tau}) \geq \Pi(x,y,\hat{p}(y),s^{\hat{\tau}}(y)), \quad \Gamma - a.e.\ y \text{ in } \mathcal{C}.$$

In particular, from our choice of $\hat{p}$ and $s^{\hat{\tau}}$ in $\mathcal{A}$ we have

$$V(x|\,\hat{p},\hat{\tau}) \geq \Pi(x,y,\tilde{p}(y),s^{\tilde{\tau}}(y)), \quad \Gamma - a.e.\ y \text{ in } \mathcal{A},$$

implying that $V(x|\,\hat{p},\hat{\tau}) \geq V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau})$. Therefore, we only need to show $V(x|\,\hat{p},\hat{\tau}) \leq V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau})$. To prove this we have to verify that $V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau})$ is a $\Gamma - a.e$ upper bound of $\Pi(x,y,\hat{p}(y),s^{\hat{\tau}}(y))$ for $y \in \mathcal{C}$. From the definition of $V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau})$ this upper bound is true in $\mathcal{A}$, so we just need to check

$$V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau}) \geq \Pi(x,y,p(y),s^{\tau}(y)), \quad \Gamma - a.e.\ y \text{ in } \mathcal{A}^c. \tag{A.12}$$

For the sake of contradiction suppose this is not true. Then,

$$\Gamma(y \in \mathcal{A}^c : \Pi(x,y,p(y),s^{\tau}(y)) > V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau})) > 0$$

For any $y \in \mathcal{A}^c$ consider the segment $[x,y]$. Since $x \in \mathcal{A}$ there must exists $x_y \in [x,y] \cap \partial\mathcal{A}$. From the Lipschitz property (see Lemma 1.1) we have that $V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau}) \geq V_{\mathcal{A}}(x_y|\,\tilde{p},\tilde{\tau}) - \|x_y - x\|$, and since $V_{\mathcal{A}}(\cdot|\,\tilde{p},\tilde{\tau})$ coincides with $V(\cdot|\,p,\tau)$ in $\partial\mathcal{A}$ we can infer that $V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau}) \geq V(x_y|\,p,\tau) - \|x_y - x\|$. Then,

$$\begin{aligned}
0 &< \Gamma(y \in \mathcal{A}^c : \Pi(x,y,p(y),s^{\tau}(y)) > V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau})) \\
&\leq \Gamma(y \in \mathcal{A}^c : \Pi(x,y,p(y),s^{\tau}(y)) > V(x_y|\,p,\tau) - \|x_y - x\|) \\
&= \Gamma(y \in \mathcal{A}^c : U(y,p(y),s^{\tau}(y)) - \|x-y\| > V(x_y|\,p,\tau) - \|x_y - x\|) \\
&\overset{(a)}{=} \Gamma(y \in \mathcal{A}^c : U(y,p(y),s^{\tau}(y)) > V(x_y|\,p,\tau) + \|x_y - y\|) \\
&\overset{(b)}{\leq} \Gamma(y \in \mathcal{A}^c : U(y,p(y),s^{\tau}(y)) > V(y|\,p,\tau)) \\
&\overset{(c)}{=} 0,
\end{aligned}$$

a contradiction. In (a) we use that x, y and x_y are collinear points. In (b) we use the the Lipschitz property, and (c) follows from Lemma A.2. Thus Eq. (A.12) holds. In conclusion, $V(x|\,\hat{p},\hat{\tau})$ equals $V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau})$ for all $x \in \mathcal{A}$.

We next show that $V(x|\,\hat{p},\hat{\tau})$ equals $V(x|\,p,\tau)$ for all $x \in \mathcal{A}^c$. We proceed by contradiction. Let $x \in \mathcal{A}^c$ and suppose that $V(x|\,\hat{p},\hat{\tau}) \neq V(x|\,p,\tau)$. If $V(x|\,\hat{p},\hat{\tau}) > V(x|\,p,\tau)$ then we must have that

$$\begin{aligned}
0 &< \Gamma(y \in \mathcal{C} : \Pi(x,\hat{p}(y),s^{\hat{\tau}}(y),y) > V(x|\,p,\tau)) \\
&= \Gamma(y \in \mathcal{A} : \Pi(x,\tilde{p}(y),s^{\tilde{\tau}}(y),y) > V(x|\,p,\tau)) \\
&+ \Gamma(y \in \mathcal{A}^c : \Pi(x,p(y),s^{\tau}(y),y) > V(x|\,p,\tau)) \\
&\overset{(a)}{=} \Gamma(y \in \mathcal{A} : U(\tilde{p}(y),s^{\tilde{\tau}}(y),y) - \|x-y\| > V(x|\,p,\tau)) \\
&\overset{(b)}{\leq} \Gamma(y \in \mathcal{A} : U(\tilde{p}(y),s^{\tilde{\tau}}(y),y) - \|x-y\| > V_{\mathcal{A}}(x_y|\,\tilde{p},\tilde{\tau}) - \|x_y - x\|) \\
&\overset{(c)}{=} \Gamma(y \in \mathcal{A} : U(\tilde{p}(y),s^{\tilde{\tau}}(y),y) > V_{\mathcal{A}}(x_y|\,\tilde{p},\tilde{\tau}) + \|x_y - y\|) \\
&\overset{(d)}{\leq} \Gamma(y \in \mathcal{A} : U(\tilde{p}(y),s^{\tilde{\tau}}(y),y) > V_{\mathcal{A}}(y|\,\tilde{p},\tilde{\tau})) \\
&\overset{(e)}{=} 0,
\end{aligned}$$

where (a) follows from that the definition of $V(x|\,p,\tau)$ implies that the second term in the previous line is zero. Similarly to what we did before, in (b) we take $x_y \in [x,y] \cap \partial\mathcal{A}$ and then apply the Lipschitz property together with the assumption that

$V_{\mathcal{A}}(x_y|\,\tilde{p},\tilde{\tau}) = V(x_y|\,p,\tau)$. In (c) we made use of the collinearity of x, y and x_y, and in (d) we applied once again the Lipschitz property. The last line (e) follows from Lemma A.2.

Now suppose that $V(x|\,\hat{p},\hat{\tau}) < V(x|\,p,\tau)$ then

$$
\begin{aligned}
0 &< \Gamma(y \in \mathcal{C} : \Pi(x, p(y), s^{\tau}(y), y) > V(x|\,\hat{p},\hat{\tau})) \\
&= \Gamma(y \in \mathcal{A} : \Pi(x, p(y), s^{\tau}(y), y) > V(x|\,\hat{p},\hat{\tau})) \\
&+ \Gamma(y \in \mathcal{A}^c : \Pi(x, p(y), s^{\tau}(y), y) > V(x|\,\hat{p},\hat{\tau})) \\
&\overset{(a)}{=} \Gamma(y \in \mathcal{A} : U(p(y), s^{\tau}(y), y) - \|x-y\| > V(x|\,\hat{p},\hat{\tau})) \\
&\overset{(b)}{\leq} \Gamma(y \in \mathcal{A} : U(p(y), s^{\tau}(y), y) - \|x-y\| > V_{\mathcal{A}}(x_y|\,\tilde{p},\tilde{\tau}) - \|x_y - x\|) \\
&\overset{(c)}{=} \Gamma(y \in \mathcal{A} : U(p(y), s^{\tau}(y), y) > V(x_y|\,p,\tau) + \|x_y - y\|) \\
&\overset{(d)}{\leq} \Gamma(y \in \mathcal{A} : U(p(y), s^{\tau}(y), y) > V(y|\,p,\tau)) \\
&\overset{(e)}{=} 0,
\end{aligned}
$$

where (a) follows from that the definition of $V(x|\,\hat{p},\hat{\tau})$ implies that the second term in the previous line is zero. Similarly to what we did before, in (b) we take $x_y \in [x,y] \cap \partial\mathcal{A}$ and then apply the Lipschitz property together with what we proved before, $V_{\mathcal{A}}(x_y|\,\tilde{p},\tilde{\tau}) = V(x_y|\,\hat{p},\hat{\tau})$ ($x_y \in \mathcal{A}$ because this set is closed). In (c) we made use of the collinearity of x, y and x_y, and that $V_{\mathcal{A}}(x_y|\,\tilde{p},\tilde{\tau}) = V(x_y|\,p,\tau)$. In (d) we applied once again the Lipschitz property. The last line (e) follows from Lemma A.2. Therefore, $V(x|\,\hat{p},\hat{\tau})$ equals $V(x|\,p,\tau)$ in $\mathcal{A}^c$.

Lastly, we verify that $\hat{\tau}(\hat{\mathcal{E}})$ equals $\mu(\mathcal{C})$. Define the sets

$$
\begin{aligned}
\mathcal{E}_1 &\triangleq \Big\{(x,y) \in (\mathcal{A}^c \cup \mathcal{L}) \times \mathcal{A}^c : \Pi(x, y, \hat{p}(y), s^{\hat{\tau}}(y)) = V(x|\,\hat{p},\hat{\tau})\Big\} \\
\mathcal{E}_2 &\triangleq \Big\{(x,y) \in \mathcal{A} \times \mathcal{A} : \Pi(x, y, \hat{p}(y), s^{\hat{\tau}}(y)) = V(x|\,\hat{p},\hat{\tau})\Big\}
\end{aligned}
$$

then $\hat{\tau}(\hat{\mathcal{E}}) = \tau(\mathcal{E}_1) + \tilde{\tau}(\mathcal{E}_2)$. We can replace the definition of $\hat{p}$ and what we have proved about $s^{\hat{\tau}}$ and $V(x|\,\hat{p},\hat{\tau})$ in the expressions above to obtain

$$\tau(\mathcal{E}_1) = \tau\Big(\Big\{(x,y) \in (\mathcal{A}^c \cup \mathcal{L}) \times \mathcal{A}^c : \Pi(x,y,p(y),s^{\tau}(y)) = V(x|\,p,\tau)\Big\}\Big),$$
$$\tilde{\tau}(\mathcal{E}_2) = \tilde{\tau}\Big(\Big\{(x,y) \in \mathcal{A} \times \mathcal{A} : \Pi(x,y,\tilde{p}(y),s^{\tilde{\tau}}(y)) = V_{\mathcal{A}}(x|\,\tilde{p},\tilde{\tau})\Big\}\Big) = \tilde{\mu}(\mathcal{A}),$$

where the second line comes from the fact that $\tilde{\tau}$ is an equilibrium in $\mathcal{A}$. Let $\mathcal{E}$ be defined analogously to $\hat{\mathcal{E}}$ but with $(\hat{p},\hat{\tau})$ replaced by (p,τ), then

$$\begin{aligned}
\hat{\tau}(\hat{\mathcal{E}}) &= \tau(\mathcal{E}_1) + \tilde{\mu}(\mathcal{A}) \\
&\overset{(a)}{=} \tau(\mathcal{E}_1) + \tau(\mathcal{A} \times \mathcal{A}) \\
&\overset{(b)}{=} \tau(\mathcal{E} \cap ((\mathcal{A}^c \cup \mathcal{L}) \times \mathcal{A}^c)) + \tau(\mathcal{E} \cap (\mathcal{A} \times \mathcal{A})) \\
&\overset{(c)}{=} \tau(\mathcal{E}) \\
&\overset{(d)}{=} \mu(\mathcal{C}),
\end{aligned}$$

where in (a) we use the definition of $\tilde{\mu}$. In (b) and (d) we use the fact that τ only puts mass in $\mathcal{E}$, and in (c) we use Proposition 1.3 part *(i)*.

□

A.2 Proofs for Section 1.5

Proof of Lemma 1.4. Suppose $\lambda(x) > 0$ and recall that the price achieving the maximum in the definition of $R_x^{loc}(s)$ is $\rho_x^{loc}(s) = \max\{\rho_x^{bal}(s), \rho_x^u\}$. Let s^u be equal to $\lambda(x) \cdot \overline{F}_x(\rho_x^u)$, that is, $\rho_x^{bal}(s^u) = \rho_x^u$ (here we are using that $q \mapsto q \cdot \overline{F}_y(q)$ is continuous and unimodular in q). Then, since $\rho_x^{bal}(\cdot)$ is decreasing we have that $\rho_x^{loc}(s) = \rho_x^{bal}(s)$ for all $0 < s \leq s^u$ and, therefore,

$$\frac{R_x^{loc}(s)}{s} = \rho_x^{bal}(s) = F^{-1}(1 - \frac{s}{\lambda(x)}), \quad \text{for all } 0 < s \leq s^u.$$

Since F is strictly increasing, the quotient above is strictly decreasing for $s \in (0, s^u]$. Moreover, since $F^{-1}(1) = \overline{V}$, the point just made also includes $s = 0$. Now, for $s > s^u$

we have $\rho_x^{loc}(s) = \rho_x^u$, thus

$$\frac{R_x^{loc}(s)}{s} = \rho_x^u \cdot \frac{\lambda(x) \cdot \overline{F}_x(\rho_x^u)}{s},$$

which is strictly decreasing. In any case, we conclude that $\psi_x(\cdot)$ is strictly decreasing when $\lambda(x) > 0$.

$\square$

Proof of Proposition 1.5. Define the set $B \triangleq \{x \in C_\lambda : V(x) > \psi_x(s^\tau(x))\}$. We want to show that $\Gamma(B) = 0$. First we argue that $B \subseteq \{x \in \mathcal{C} : U(x) \neq V(x)\}$, indeed, let $x \in B$ then

$$V(x) > \psi_x(s^\tau(x)) \geq U\Big(x, p(x), s^\tau(x)\Big),$$

that is, $V(x) > U(x)$ as desired. By Lemma A.2 we know that $\tau_2(\{x \in \mathcal{C} : U(x) \neq V(x)\}) = 0$ and, therefore, $\tau_2(B) = 0$. This yields,

$$0 = \tau_2(B) = \int_B s^\tau(x)\, d\Gamma(x). \tag{A.13}$$

If $\Gamma(B) = 0$ then we are done. Suppose $\Gamma(B) > 0$, from equation (A.13) we can conclude that $s^\tau(x) = 0$, $\Gamma - a.e.$ $x \in B$. Since in B we have $\lambda(x) > 0$ this implies that $\Gamma - a.e$ in B we have that $\psi_x(s^\tau(x))$ equals $\alpha \cdot \overline{V}$. Because $\alpha \cdot \overline{V}$ is the maximum value that $V(\cdot)$ can attain (see Lemma A.1), we conclude that

$$\alpha \cdot \overline{V} \geq V(x) > \psi_x(s^\tau(x)) = \alpha \cdot \overline{V} \qquad \Gamma - a.e.\ x \in B.$$

But since we are assuming that $\Gamma(B) > 0$, this yields a contradiction. $\square$

Proof of Theorem 1.1. The proof of this theorem consists of several parts. In the first part we specialize the upper bound derived in Proposition 1.5 to account for the case when $\lambda(x) = 0$. Next, we pose an optimization problem which is a relaxation of platform's optimization problem restricted to the attraction region $A(z)$. Then we introduce some notation. Given this, the relaxation has a similar structure to a continuous bounded knapsack problem, and we characterize the structure of the optimal solution as stated in the statement of the theorem. Next we construct a local price-equilibrium pair $(\hat{p}, \hat{\tau})$ in $A(z)$ that implements the relaxation's solution. We conclude by applying the pasting result of Proposition 1.4 to globally extend our price-equilibrium pair $(\hat{p}, \hat{\tau})$ in $\mathcal{C}$ as in the statement of the theorem. In summary the parts of the proof are: Upper bound specialization, Relaxation, Notation, Knapsack, Implementation and Conclusion. We enumerate all these parts from 1 to 6, and present them in boldface to make the presentation clearer.

Part 1: Upper bound specialization. For ease of notation we use X_a and $A(z)$ to denote $X_a(z|p,\tau)$, and $A(z|p,\tau)$, respectively. Recall $z \in \mathcal{C}$ is a sink location, so the following is well defined

$$X_a^{\text{supp}} \triangleq \inf\{x \in [z, X_a] \,\cap\, \text{supp}(\Gamma)\}.$$

We use $A^{\text{supp}}(z)$ to denote $\cup_{a \in R_z}[z, X_a^{\text{supp}}]$, and (with abuse of notation)

$$L^{\text{supp}}(z) = \bigcup_{a \in R_z} \{X_a^{\text{supp}}\}, \quad \text{and} \quad A^{\text{supp}}(z)^{\circ} = \bigcup_{a \in R_z} [z, X_a^{\text{supp}}).$$

We define the function

$$H_x(V) \triangleq \begin{cases} \psi_x^{-1}(V) & \text{if } \lambda(x) > 0; \\ 0 & \text{if } \lambda(x) = 0,\ x \in A^{\text{supp}}(z)^{\circ}; \\ \frac{d\mu}{d\Gamma}(x) & \text{if } \lambda(x) = 0,\ x \in L^{\text{supp}}(z); \\ 0 & \text{if } x \in A(z) \setminus A^{\text{supp}}(z)^{\circ}. \end{cases} \tag{A.14}$$

In this part of the proof we will show that

$$s^{\tau}(x) \;\leq\; H_x(V(x|\, p, \tau)), \quad \Gamma - a.e.\ x \ \text{ in } \ A(z). \tag{A.15}$$

In order to prove Eq. (A.15) first note that from Proposition 1.5 we have

$$s^{\tau}(x) \;\leq\; H_x(V(x|\, p, \tau)), \quad \Gamma - a.e.\ x \ \text{ in } \ A(z) \cap \mathcal{C}_{\lambda},$$

so we only need to show that the set

$$B \triangleq \{x \in A(z) : \lambda(x) = 0,\ s^{\tau}(x) > H_x(V(x|\, p, \tau))\},$$

satisfies $\Gamma(B) = 0$. From the definition of X_a^{supp} we have that $\Gamma(A(z) \backslash A^{\text{supp}}(z))$ equals zero (beyond these X_a^{supp} the city measure does not put mass). Hence, showing that $\Gamma(B)$ equals zero is equivalent to showing that $\Gamma(B_1 \cup B_2)$ equals zero, where

$$B_1 \triangleq \{x \in A^{\text{supp}}(z)^\circ : \lambda(x) = 0,\ s^\tau(x) > 0\},$$
$$B_2 \triangleq \{x \in L^{\text{supp}}(z) : \lambda(x) = 0,\ s^\tau(x) > \frac{d\mu}{d\Gamma}(x)\}.$$

For the sake of contradiction assume that $\Gamma(B_1) > 0$ then $\tau_2(B_1) = \int_{B_1} s^\tau \, d\Gamma > 0$. This, together with Lemma A.2, yields that $\tau_2(B_1 \cap \{x : U(x) = V(x)\}) > 0$, which in turn implies the existence of $x \in B_1 \cap \{x : U(x) = V(x)\}$. Such an x satisfies that $x \in A^{\text{supp}}(z)^\circ$ and $V(x) = 0$ and, therefore, $V(x') < 0$ for some $x' \in L^{\text{supp}}(z)$ (recall that by Lemma 1.3, $V(\cdot)$ is linear on any ray $[z, X_a]$ around z). However, any x' in $L^{\text{supp}}(z)$ belongs to $\text{supp}(\Gamma)$ and, hence, Lemma A.1 guarantees that $V(x') \geq 0$, yielding a contradiction. Thus, $\Gamma(B_1) = 0$.

If $\Gamma(B_2) > 0$ then by the definition of B_2 we must have that $\tau_2(B_2) > \mu(B_2)$. We will also argue that $\mu(B_2) \geq \tau_2(B_2)$ to obtain a contradiction. Indeed,

$$\begin{aligned}
\mu(B_2) &\geq \tau(B_2 \times B_2) \\
&= \tau_2(B_2) - \tau(\mathcal{C} \setminus B_2 \times B_2) \\
&= \tau_2(B_2) - \tau(A(z) \setminus B_2 \times B_2) \\
&\overset{(a)}{=} \tau_2(B_2) - \tau(A^{\text{supp}}(z) \setminus B_2 \times B_2) \\
&\overset{(b)}{=} \tau_2(B_2) - \tau(L^{\text{supp}}(z) \setminus B_2 \times B_2) \\
&= \tau_2(B_2),
\end{aligned}$$

where (a) comes from $\tau(A^{\text{supp}}(z)^\circ \times L^{\text{supp}}(z)) = 0$ (recall that by Lemma 1.3, $V(\cdot)$ is linear and decrease on any ray $[z, X_a]$ around z). And (b) holds because τ does not send mass across rays, so the mass can only me sent to from in the pairs $(X_a^{\text{supp}}, X_a^{\text{supp}})$; but this pairs do not belong to $L^{\text{supp}}(z) \setminus B_2 \times B_2$. In conclusion, $\Gamma(B) = 0$ and Eq. (A.15) is proven.

Part 2: Relaxation. We consider the attraction region $A(z)$. In it, the upper bound we just proved in Eq. (A.15) must be satisfied. Moreover, due to our flow separation result in Proposition 1.3 part *(i)* we have $\tau_2(A(z)) = \tau(A(z) \times A(z))$. Also, since flow is not transported across rays (see Proposition 1.3 part *(ii)*), the total supply in the ray $(z, X_a]$ cannot be larger than its initial supply. Therefore, in $A(z)$ the platform's problem is bounded above by

$$\max_{s(\cdot)} \quad \int_{A(z)} V(x) \cdot s(x)\, d\Gamma(x) \tag{$\mathcal{P}_{KP}(z)$}$$

$$\text{s.t} \quad s(x) \leq H_x(V(x|\, p, \tau)), \quad \Gamma - a.e.\ x \ \text{in}\ A(z) \tag{CB}$$

$$\int_{A(z)} s(x)\, d\Gamma(x) = \tau(A(z) \times A(z)) \tag{FC}$$

$$\int_{(z,X_a]} s(x)\, d\Gamma_a(x) \leq \int_{(z,X_a]} s^{\tau}(x)\, d\Gamma_a(x), \quad \Gamma^{\mathrm{p}} - a.e.\ a \in R_z. \tag{FR$_a$}$$

Observe that s^{τ} (which defines τ_2) is a feasible solution for $(\mathcal{P}_{KP}(z))$. The supply density $s^{\hat{\tau}}$(as in the statement of the present theorem) will be shown to be an optimal solution for this relaxation.

Part 3: Notation.

1. Next we rename the quantities on the RHS of equations (FC) and (FR$_a$).

$$\tau_{\text{total}} = \tau(A(z) \times A(z)),$$
$$\tau_a = \int_{(z,X_a]} s^{\tau}(x)\, d\Gamma_a(x),$$
$$\tau_c = \tau(A(z) \times \{z\}).$$

2. For any measurable set $B \subseteq Az$ we define the measure

$$S^H(B) \triangleq \int_B H_x(V(x))\, d\Gamma(x),$$

   $S^H(\cdot)$ is the measure with density $H_x(V(x))$ (see Eq. (A.14)) with respect to the Γ measure. Moving forward we will use $s^H(x)$ to denote its density.

Part 4: Knapsack. We show that any optimal solution to $(\mathcal{P}_{KP}(z))$ is as $s^{\hat{\tau}}$ in the statement of the theorem. There are two cases.

Case 1. First suppose that $0 < \tau_{\text{total}} \leq S^H(\{z\})$ (so that there is an atom at z). Then, we define $r_a = z$ for all $a \in R_z$, and let the solution to $(\mathcal{P}_{KP}(z))$ be

$$s^\star(x) = \frac{\tau_{\text{total}}}{\Gamma(\{z\})} \cdot \mathbf{1}_{\{x=z\}},$$

which is feasible, and optimal because for any feasible s we have

$$\int_{A(z)} V(x) \cdot s(x) d\Gamma(x) \leq V(z) \cdot \int_{A(z)} s(x)\, d\Gamma(x) = V(z) \cdot \tau_{\text{total}},$$

which is equal to the objective function at $s^\star$. So in this case the optimal solution coincides with the description of $s^{\hat{\tau}}$ as in the statement of the theorem.

Case 2. Now let us assume that $\tau_{\text{total}} > S^H(\{z\})$. We start by showing that in this case we have $s^\star(z) = s^H(z)$. If z is not a point with positive Γ- mass then setting $s^\star(z)$ in this way is without loss of generality. If the point z has positive mass then we argue by contradiction that $s^\star(z)$ must be choose in this way. Let $s^\star$ be an optimal solution to $(\mathcal{P}_{KP}(z))$ such that $s^\star(z) < s^H(z)$. Then,

$$\tau_{\text{total}} = \int_{A(z)\setminus\{z\}} s^\star d\Gamma + s^\star(z) \cdot \Gamma(\{z\}) < \underbrace{\int_{A(z)\setminus\{z\}} s^\star d\Gamma}_{K} + s^H(z) \cdot \Gamma(\{z\}). \tag{A.16}$$

Let $\epsilon \in (0, 1)$ be such that $(\tau_{\text{total}} - \epsilon \cdot K)/\Gamma(\{z\}) = s^H(z)$, this is well defined because we are assuming $\tau_{\text{total}} > S^H(\{z\})$. Next define a new solution $\bar{s}$ by

$$\bar{s}(x) = \begin{cases} s^H(z) & \text{if } x = z, \\ \epsilon \cdot s^\star(x) & \text{if } x \neq z. \end{cases}$$

Note that $\bar{s}$ is feasible: it satisfies (FR_a) for all $a \in R_z$ and (CB), and for (FC) we have

$$\int_{A(z)} \bar{s} d\Gamma = \epsilon \cdot K + s^H(z) \cdot \Gamma(\{z\}) = \tau_{\text{total}}.$$

Furthermore, $\bar{s}$ yields an strictly larger objective than $s^\star$,

$$\begin{aligned}
\int_{A(z)} V(x)\cdot s^\star(x)\,d\Gamma(x) &= \int_{A(z)\setminus\{z\}} V(x)\cdot s^\star(x)\,d\Gamma(x) + V(z)\cdot s^\star(z)\cdot\Gamma(\{z\}) \\
&= \epsilon\cdot\int_{A(z)\setminus\{z\}} V(x)\cdot s^\star(x)\,d\Gamma(x) \\
&+ (1-\epsilon)\cdot\int_{A(z)\setminus\{z\}} V(x)\cdot s^\star(x)\,d\Gamma(x) \\
&+ V(z)\cdot s^\star(z)\cdot\Gamma(\{z\}) \\
&\overset{(a)}{<} \int_{A_{\mathrm{p}}(z)\setminus\{z\}} V(x)\cdot \bar{s}(x)\,d\Gamma(x) + (1-\epsilon)\cdot V(z)\cdot K \\
&+ V(z)\cdot s^\star(z)\cdot\Gamma(\{z\}) \\
&= \int_{A(z)\setminus\{z\}} V(x)\cdot \bar{s}(x)\,d\Gamma(x) + V(z)\cdot(\tau_{\text{total}}-\epsilon\cdot K) \\
&\overset{(b)}{=} \int_{A(z)} V(x)\cdot \bar{s}(x)\,d\Gamma(x),
\end{aligned}$$

where (a) comes from Eq. (A.16), and (b) holds because $(\tau_{\text{total}}-\epsilon\cdot K)/\Gamma(\{z\}) = s^H(z)$. Hence, whenever $\tau_{\text{total}} > S^H(\{z\})$, we can assume that $s^\star(z) = s^H(z)$. We assume this for the reminder of the proof.

Let $s^\star(z)$ be an optimal solution. We show how to build $\hat{s}$ with the properties described in the theorem's statement. Next we construct r_a. First, note that

$$\int_{A(z)\setminus\{z\}} s^\star(x)d\Gamma = \int_{A(z)\setminus\{z\}} s^\star(x)\mathbf{1}_{\{s^\star\le s^H\}}d\Gamma = \int_{R_z} \underbrace{\int_{(z,X_a]} s^\star(x)\mathbf{1}_{\{s^\star\le s^H\}}d\Gamma_a(x)}_{q_a}\,d\Gamma^{\mathrm{p}}(a)$$

define r_a by

$$r_a \triangleq \inf\left\{r\in(0,X_a] : \int_{(0,r]} s^H(x)\mathbf{1}_{\{s^\star\le s^H\}}d\Gamma_a(x) \ge q_a\right\}.$$

Observe that for $r = X_a$ the integral in the definition of r_a is larger or equal than q_a. Therefore, r_a is well defined. Let us define (with some abuse of notation)

$$A_r(z) \triangleq \bigcup_{a\in R_z}[z, r_a], \quad \text{and} \quad L_r(z) \triangleq \bigcup_{a\in R_z}\{r_a\}.$$

Let's define a new solution $\hat{s}$ by

$$\hat{s}(x) \triangleq \begin{cases} s^\star(z) = s^H(z) & \text{if } x = z \\ s^H(x) & \text{if } x \in A_r(z) \setminus (L_r(z) \cup \{z\}), \\ \mathbf{1}_{\{\Gamma_a(\{x\})>0\}} \frac{\left(q_a - \int_{(0,x)} s^H(y) \mathbf{1}_{\{s^\star \le s^H\}} d\Gamma_a(y)\right)}{\Gamma_a(\{x\})} & \text{if } x \in L_r(z), \end{cases}$$

and $\hat{s}(x) = 0$ otherwise. We show that $\hat{s}$ weakly revenue dominates $s^\star$ and that is feasible. Let us do first the <u>revenue dominance</u>. Note that the objective in $\{z\}$ of both solutions coincide; thus, we only need to compare the objective in the set $Q \triangleq A(z) \setminus \{z\}$. Note that $A_r(z) \setminus \{z\} \subset Q$, then

$$\begin{aligned} \int_Q V(x) \cdot s^\star(x) \, d\Gamma(x) &= \int_{A_r(z)\setminus\{z\}} V(x) \cdot s^\star(x) \, d\Gamma(x) + \int_{Q\setminus(A_r(z)\setminus\{z\})} V(x) \cdot s^\star(x) \, d\Gamma(x) \\ &= \int_{A_r(z)\setminus\{z\}} V(x) \cdot \hat{s}(x) \, d\Gamma(x) + \int_{A_r(z)\setminus\{z\}} V(x) \cdot (s^\star - \hat{s})(x) \, d\Gamma(x) \\ &+ \underbrace{\int_{Q\setminus(A_r(z)\setminus\{z\}))} V(x) \cdot s^\star(x) \mathbf{1}_{\{s^\star \le s^H\}} \, d\Gamma(x)}_{I}, \end{aligned}$$

for the last term above we have

$$\begin{aligned} I &\le \int_{R_z} V(r_a) \Big[\int_{(r_a, X_a]} s^\star(x) \mathbf{1}_{\{s^\star \le s^H\}} \, d\Gamma_a(x) \Big] d\Gamma^{\mathtt{P}}(a) \\ &= \int_{R_z} V(r_a) \Big[q_a - \int_{(z, r_a]} s^\star(x) \mathbf{1}_{\{s^\star \le s^H\}} \, d\Gamma_a(x) \Big] d\Gamma^{\mathtt{P}}(a) \\ &= \int_{R_z} V(r_a) \Big[\int_{(z, r_a)} (s^H - s^\star)(x) \mathbf{1}_{\{s^\star \le s^H\}} \, d\Gamma_a(x) + (\hat{s} - s^\star \mathbf{1}_{\{s^\star \le s^H\}})(r_a) \Gamma_a(r_a) \Big] d\Gamma^{\mathtt{P}}(a) \\ &\le \int_{R_z} \Big[\int_{(z, r_a)} V(x)(s^H - s^\star)(x) \mathbf{1}_{\{s^\star \le s^H\}} \, d\Gamma_a(x) \\ &+ V(r_a)(\hat{s} - s^\star \mathbf{1}_{\{s^\star \le s^H\}})(r_a) \Gamma_a(r_a) \Big] d\Gamma^{\mathtt{P}}(a) \\ &= \int_{A_r(z)\setminus\{z\}} V(x)(\hat{s} - s^\star)(x) d\Gamma(x), \end{aligned}$$

hence

$$\int_Q V(x) \cdot s^\star(x) \, d\Gamma(x) \le \int_{A_r(z)\setminus\{z\}} V(x) \cdot \hat{s}(x) \, d\Gamma(x).$$

Since the right hand side above equals the objective under $\hat{s}$ in $A_r(z)$ we conclude that $\hat{s}$ is an optimal solution.

For the feasibility of $\hat{s}$, by construction and the definition of r_a we have that $\hat{s}$ satisfies (CB). Furthermore, because $s^\star$ satisfies (FR$_a$) and since $\hat{s}$ only redistributes the mass of $s^\star$ across rays but no between rays that originate in z, $\hat{s}$ also satisfies (FR$_a$). In order to verify (FC) note that

$$\begin{aligned}
\int_{A(z)} \hat{s}(x)\, d\Gamma(x) &= \hat{s}(z) \cdot \Gamma(\{z\}) + \int_{A_r(z)\setminus\{z\}} \hat{s}(x)\, d\Gamma(x) \\
&= s^\star(z) \cdot \Gamma(\{z\}) + \int_{A_r(z)\setminus\{z\}} \hat{s}(x)\, d\Gamma(x) \\
&= s^\star(z) \cdot \Gamma(\{z\}) \\
&+ \int_{R_z} \Big[\int_{(z,r_a)} s^H(z) \mathbf{1}_{\{s^\star \le s^H\}}\, d\Gamma_a(x) + \hat{s}(r_a)\Gamma_a(\{r_a\}) \Big] d\Gamma^{\mathrm{p}}(a) \\
&= s^\star(z) \cdot \Gamma^{\mathrm{p}}(\{z\}) + \int_{R_z} \Big[q_a \Big] d\Gamma^{\mathrm{p}}(a) \\
&= s^\star(z) \cdot \Gamma^{\mathrm{p}}(\{z\}) + \int_{R_z} \Big[\int_{(z,X_a]} s^\star(x) \mathbf{1}_{\{s^\star \le s^H\}} d\Gamma_a(x) \Big] d\Gamma^{\mathrm{p}}(a) \\
&= s^\star(z) \cdot \Gamma^{\mathrm{p}}(\{z\}) + \int_{A(z)\setminus\{z\}} s^\star(x)\, d\Gamma(x) \\
&= \tau_{\text{total}}.
\end{aligned}$$

In conclusion, the solution $\hat{s}$ constructed is as defined in the statement of the theorem. Next, we use this solution to define prices and flows. We use $\hat{S}$ to denote the measure induced by $\hat{s}$. Observe that $\hat{S}$ has support in $A_r(z)$.

Part 5: Implementation. We construct a price-equilibrium pair $(\hat{p}, \hat{\tau})$ in $A(z)$ with $\hat{\tau} \in \mathcal{F}_{A(z)}(\tilde{\mu})$ and

$$\tilde{\mu}(\mathcal{B}) \triangleq \tau((\mathcal{B} \cap A(z)) \times A(z)), \quad \mathcal{B} \subseteq \mathcal{C} \text{ measurable.}$$

- **Prices.** Define $\hat{p} : A(z) \to [0, \overline{V}]$ by

$$\hat{p}(x) = \begin{cases} \rho_x^{loc}(\hat{s}(x)) & \text{if } x \in A_r(z) \setminus L_r(z); \\ p_a & \text{if } x = r_a,\ a \in R_z; \\ \overline{V} & \text{otherwise,} \end{cases}$$

where p_a is such that $U(r_a, p_a, \hat{s}(r_a)) = V(r_a|\, p, \tau) \cdot \mathbf{1}_{\{\lambda(r_a)>0\}}$ for $a \in R_z$. By the way we constructed $\hat{s}(r_a)$, it is bounded by $H_{r_a}(V(r_a))$ and, therefore, the value p_a is always well defined (Γ-a.e).

- **Flows:** We define $\hat{\tau}$ as a transport plan between $\tilde{\mu}$ and $\hat{S}$. We start by defining the flow that $\hat{\tau}$ sends to z and then the flow along rays.

  Flow to the center. Next we define the flow that $\hat{\tau}$ sends to $\{z\}$. We define

$$\tilde{\mu}_a(\mathcal{B}) \triangleq \int_{\mathcal{B}\cap(z,X_a]} \frac{d\tilde{\mu}}{d\Gamma}(y)d\Gamma_a(y) \text{ and } \hat{S}_a(\mathcal{B}) \triangleq \int_{\mathcal{B}\cap(z,X_a]} \frac{d\hat{S}}{d\Gamma}(y)d\Gamma_a(y).$$

Then,

$$\tilde{\mu}(\mathcal{B}) = \tau(\{z\}\times\{z\})\mathbf{1}_{\{z\in\mathcal{B}\}} + \int_{R_z} \tilde{\mu}_a(\mathcal{B})d\Gamma^{\mathrm{p}}(a),$$

and

$$\hat{S}(\mathcal{B}) = \hat{S}(\{z\})\mathbf{1}_{\{z\in\mathcal{B}\}} + \int_{R_z} \hat{S}_a(\mathcal{B})d\Gamma^{\mathrm{p}}(a).$$

We define the quantities

$$\Delta_a \triangleq \tilde{\mu}_a((z,X_a]) - \hat{S}_a((z,X_a]),$$

note that because of (FR$_a$), $\Delta_a \geq 0$, $\Gamma^{\mathrm{p}} - a.e\ a$ in R_z. Further define

$$h_a \triangleq z + \inf\{\delta \geq 0 : \tilde{\mu}_a((z, z+\delta]) \geq \Delta_a\}.$$

For any set $\mathcal{B} \subseteq A(z)$ we define the mass going to the center from ray $a \in R_z$ by the measures

$$\mu_a^c(\mathcal{B}) \triangleq \tilde{\mu}_a(\mathcal{B}\cap(z,h_a)) + \mathbf{1}_{\{h_a\in\mathcal{B}\cap(z,X_a]\}} \cdot (\Delta_a - \tilde{\mu}_a(z,h_a)),$$

observe that by the definition h_a, the atoms above have non-negative mass, $\Gamma^{\mathrm{p}} - a.e$ a in R_z . Let $\mathcal{Q}_z \triangleq \{z\}\times\{z\}$. For any measurable set $\mathcal{R} \subseteq A(z)\times A(z)$, the measure that sends flow to the origin is defined by

$$\tau^c(\mathcal{R}) \triangleq \tau(\mathcal{R}\cap\mathcal{Q}_z) + \int_{R_z} \mu_a^c(\pi_1(\mathcal{R}\cap A(z)\times\{z\}))d\Gamma^{\mathrm{p}}(a),$$

where π_1 is the mapping that to each pair (x, y) assigns the first component x. Using Lemma A.6 (which we state and prove after the present proof) we can verify that $\tau^c \in \mathcal{M}(A(z) \times A(z))$.

Flow along rays. For any ray $a \in R_z$ define the flow $\tilde{\gamma}_a$ along that ray to be the solution to the following optimal transport problem:

$$\min \int_{(z,X_a]\times(z,X_a]} \|x-y\| \, d\gamma_a(x,y)$$
$$\text{s.t } \gamma_a \in \Pi(\tilde{\mu}_a^{\mathfrak{r}}, \hat{S}_a),$$

where

$$\tilde{\mu}_a^{\mathfrak{r}}(\mathcal{B}) \triangleq \tilde{\mu}_a(\mathcal{B} \cap (h_a, X_a]) + \mathbf{1}_{\{h_a \in \mathcal{B}\cap(z,X_a]\}} \cdot (\tilde{\mu}_a(z, h_a] - \Delta_a),$$

where $\Pi(\tilde{\mu}_a^{\mathfrak{r}}, \hat{S}_a)$ is the set of transport plans between $\tilde{\mu}_a^{\mathfrak{r}}$ and $\hat{S}_a$. Any solution to this problem satisfies:

$$\tilde{\gamma}_a\Big(\{(x,y) \in (z,X_a] \times (z,X_a] : y > x\}\Big) = 0, \quad \Gamma^{\mathfrak{p}} - a.e.\ a \in R_z. \tag{A.17}$$

We provide a proof Eq. (A.17) after **Part 6**.

We will argue that $\hat{\tau}$ defined by

$$\hat{\tau}(\mathcal{R}) = \tau^c(\mathcal{R}) + \int_{R_z} \tilde{\gamma}_a(\mathcal{R}) d\Gamma^{\mathfrak{p}}(a)$$

yields an equilibrium, that is, for the set

$$\tilde{\mathcal{E}} \triangleq \Big\{(x,y) \in A(z) \times A(z) : U(y, \hat{p}(y), s^{\hat{\tau}}(y)) - |y-x| = V_{A(z)}(x|\, \hat{p}, \hat{\tau})\Big\},$$

we have that $\hat{\tau}(\tilde{\mathcal{E}})$ equals $\tilde{\mu}(A(z))$. Note that with this definition of $\hat{\tau}$ there is not flow being transported across rays but only within rays. Before verifying the equilibrium condition we check that $\hat{\tau} \in \mathcal{F}_{A(z)}(\tilde{\mu})$. Clearly $\hat{\tau}$ is a non-negative measure in $A(z) \times A(z)$ because is the sum of non-negative measures. Now we check that $\hat{\tau}_1 = \tilde{\mu}$. Consider a measurable set $\mathcal{B} \subseteq A(z)$ then

$$\begin{aligned}
\tilde{\tau}_1(\mathcal{B}) &= \tilde{\tau}(\mathcal{B} \times A(z)) \\
&= \tau^c(\mathcal{B} \times A(z)) + \int_{R_z} \tilde{\gamma}_a(\mathcal{B} \times A(z)) d\Gamma^{\mathtt{p}}(a) \\
&= \tau(\mathcal{Q}_z)\mathbf{1}_{\{z\in\mathcal{B}\}} + \int_{R_z} \mu_a^c(\mathcal{B}) d\Gamma^{\mathtt{p}}(a) + \int_{R_z} \tilde{\mu}_a^{\mathtt{r}}(\mathcal{B} \cap (z, X_a]) d\Gamma^{\mathtt{p}}(a) \\
&= \tau(\mathcal{Q}_z)\mathbf{1}_{\{z\in\mathcal{B}\}} + \int_{R_z} \Big[\tilde{\mu}_a(\mathcal{B} \cap (z, h_a)) + \mathbf{1}_{\{h_a \in \mathcal{B}\cap(z,X_a]\}} \cdot (\Delta_a - \tilde{\mu}_a(z, h_a)) \\
&+ \tilde{\mu}_a(\mathcal{B} \cap (h_a, X_a]) + \mathbf{1}_{\{h_a \in \mathcal{B}\cap(z,X_a]\}} \cdot (\tilde{\mu}_a(z, h_a] - \Delta_a)\Big] d\Gamma^{\mathtt{p}}(a) \\
&= \tau(\mathcal{Q}_z)\mathbf{1}_{\{z\in\mathcal{B}\}} + \int_{R_z} \tilde{\mu}_a(\mathcal{B}) d\Gamma^{\mathtt{p}}(a) \\
&= \tilde{\mu}(\mathcal{B})
\end{aligned}$$

and from the definition of $\tilde{\mu}$ we also have $\hat{\tau}_1 \ll \Gamma$. For the second marginal of $\hat{\tau}$ we have

$$
\begin{aligned}
\hat{\tau}_2(\mathcal{B}) &= \tilde{\tau}(A(z)\times\mathcal{B}) \\
&= \tau(\mathcal{Q}_z)\mathbf{1}_{\{z\in\mathcal{B}\}} + \int_{R_z}\mu_a^c(A(z))\mathbf{1}_{\{z\in\mathcal{B}\}}d\Gamma^{\mathrm{p}}(a) + \int_{R_z}\hat{S}_a(\mathcal{B})d\Gamma^{\mathrm{p}}(a) \\
&= \tau(\mathcal{Q}_z)\mathbf{1}_{\{z\in\mathcal{B}\}} \\
&+ \mathbf{1}_{\{z\in\mathcal{B}\}}\int_{R_z}\Big[\tilde{\mu}_a((z,h_a)) + \mathbf{1}_{\{h_a\in(z,X_a]\}}\cdot(\Delta_a - \tilde{\mu}_a(z,h_a))\Big]d\Gamma^{\mathrm{p}}(a) \\
&+ \int_{R_z}\hat{S}_a(\mathcal{B})d\Gamma^{\mathrm{p}}(a) \\
&= \tau(\mathcal{Q}_z)\mathbf{1}_{\{z\in\mathcal{B}\}} + \mathbf{1}_{\{z\in\mathcal{B}\}}\int_{R_z}\Delta_a d\Gamma^{\mathrm{p}}(a) + \int_{R_z}\hat{S}_a(\mathcal{B})d\Gamma^{\mathrm{p}}(a) \\
&= \tau(\mathcal{Q}_z)\mathbf{1}_{\{z\in\mathcal{B}\}} + \mathbf{1}_{\{z\in\mathcal{B}\}}\int_{R_z}[\tilde{\mu}_a((z,X_a]) - \hat{S}_a((z,X_a])]d\Gamma^{\mathrm{p}}(a) \\
&+ \int_{R_z}\hat{S}_a(\mathcal{B})d\Gamma^{\mathrm{p}}(a) \\
&= \tau(\mathcal{Q}_z)\mathbf{1}_{\{z\in\mathcal{B}\}} + \mathbf{1}_{\{z\in\mathcal{B}\}}\int_{R_z}[\tilde{\mu}_a(A(z)) - \hat{S}_a(A(z))]d\Gamma^{\mathrm{p}}(a) + \int_{R_z}\hat{S}_a(\mathcal{B})d\Gamma^{\mathrm{p}}(a) \\
&= \tau(\mathcal{Q}_z)\mathbf{1}_{\{z\in\mathcal{B}\}} + \mathbf{1}_{\{z\in\mathcal{B}\}}\Big[\tilde{\mu}(A(z)) - \tau(\mathcal{Q}_z) - \hat{S}(A(z)) + \hat{S}(\{z\})\Big] \\
&+ \int_{R_z}\hat{S}_a(\mathcal{B})d\Gamma^{\mathrm{p}}(a) \\
&= \hat{S}(\{z\})\mathbf{1}_{\{z\in\mathcal{B}\}} + \mathbf{1}_{\{z\in\mathcal{B}\}}\underbrace{\Big[\tilde{\mu}(A(z)) - \hat{S}(A(z))\Big]}_{=0} + \int_{R_z}\hat{S}_a(\mathcal{B})d\Gamma^{\mathrm{p}}(a) \\
&= \hat{S}(\mathcal{B}).
\end{aligned}
$$

Since $\hat{S}$ is such that $\hat{S} \ll \Gamma$, we conclude that $\hat{\tau} \in \mathcal{F}_{A(z)}(\tilde{\mu})$. Also, $s^{\hat{\tau}}$ coincides with $\hat{s}$ Γ almost everywhere. Before we move to verify that $\hat{\tau}$ is an equilibrium, we next compute $V_{A(z)}(x|\,\hat{p},\hat{\tau})$ and $U(y,\hat{p}(y),s^{\hat{\tau}}(y))$.

- **Equilibrium utilities:** We show that $V_{A(z)}(x|\,\hat{p},\hat{\tau})$ equals $V(z|\,p,\tau) - |z-x|$ for all $x \in A(z)$. First, from the definition of $\hat{p}$ and the value of $s^{\hat{\tau}}$ we have that $\Gamma - a.e.$ y in $A(z)$

$$
U(y,\hat{p}(y),s^{\hat{\tau}}(y)) = \begin{cases} (V(z|\,p,\tau) - |z-y|)\cdot\mathbf{1}_{\{\lambda(y)>0\}} & \text{if } y \in A_r(z), \\ 0 & \text{if } y \in A(z)\setminus A_r(z). \end{cases}
$$

Second, for any $x \in A(z)$ we argue that $V(x|\, p,\tau) = V(z|\, p,\tau) - |z-x| \geq V_{A(z)}(x|\, \hat{p},\hat{\tau})$. It is enough to show that

$$\Gamma(y \in A(z) : U(y, \hat{p}(y), s^{\hat{\tau}}(y)) - |y-x| > V(x|\, p,\tau)) = 0.$$

Suppose this is not true. Lemma A.1 implies that $V(y|p,\tau)$ is non-negative $\Gamma - a.e.$ Also $V(y|\, p,\tau)$ equals $V(z|\, p,\tau) - |z-y|$ for any $y \in A(z)$. Hence it must be true that $V(y|\, p,\tau)$ is larger or equal than $U(y, \hat{p}(y), s^{\hat{\tau}}(y))$ $\Gamma - a.e$ (see the value of this expression above). Thus our current assumption implies

$$\Gamma(y \in A(z) : V(y|\, p,\tau) - |y-x| > V(x|\, p,\tau)) > 0,$$

but this contradicts the Lipschitz property of $V(\cdot|\, p,\tau)$.

Third, we show that the upper bound we just proved is tight, that is, for all $\epsilon > 0$

$$\Gamma(y \in A(z) : U(y, \hat{p}(y), s^{\hat{\tau}}(y)) - |y-x| > V(x|\, p,\tau) - \epsilon) > 0. \tag{A.18}$$

Recall that z is an in-demand sink if $\forall Q \subset R_z$ such that $\Gamma^{\mathfrak{p}}(Q) > 0$ then

$$\Gamma(\{z\})\mathbf{1}_{\{\lambda(z)>0\}} + \int_Q \int_{(z,z+\delta]} \mathbf{1}_{\{\lambda(x)>0\}} d\Gamma_a(x) d\Gamma^{\mathfrak{p}}(a) > 0, \quad \forall \delta > 0. \tag{A.19}$$

Next, define $\bar{r}$ as the essential supremum of $\{\|r_a - z\|\}$ with respect to the measure $\Gamma^{\mathfrak{p}}$

$$\bar{r} \triangleq \inf\{c \in \mathbb{R} : \Gamma^{\mathfrak{p}}\Big(a \in R_z : \|r_a - z\| > c\Big) = 0\}.$$

Let us analyze two cases. First, assume that $\bar{r} > 0$. In this case there exists $\delta_0 > 0$ such that $\bar{r} > \delta_0$ and the set $Q_{\delta_0} = \{a \in R_z : \|r_a - z\| > \delta_0\}$ has $\Gamma^{\mathfrak{p}}$ positive measure. Choose $0 < \delta \leq \min\{\epsilon/2, \delta_0\}$ then from Eq. (A.19) we have

$$\Gamma(\{z\})\mathbf{1}_{\{\lambda(z)>0\}} + \int_{Q_{\delta_0}} \int_{(z,z+\delta]} \mathbf{1}_{\{\lambda(x)>0\}} d\Gamma_a(x) d\Gamma^{\mathfrak{p}}(a) > 0, \tag{A.20}$$

but this implies that

$$
\begin{aligned}
0 &< \Gamma(y \in B(z,\delta) \cap A_r(z) : \lambda(y) > 0) \\
&= \Gamma(y \in B(z,\delta) \cap A_r(z) : \lambda(y) > 0,\ 2\delta + \|z - x\| \geq \|y - x\| + \|y - z\|) \\
&\leq \Gamma(y \in B(z,\delta) \cap A_r(z) : \lambda(y) > 0,\ \epsilon + \|z - x\| > \|y - x\| + \|y - z\|) \\
&\leq \Gamma(y \in A_r(z) : \lambda(y) > 0,\ \epsilon + \|z - x\| > \|y - x\| + \|y - z\|) \\
&= \Gamma(y \in A_r(z) : \lambda(y) > 0,\ U(y, \hat{p}(y), s^{\hat{\tau}}(y)) - \|y - x\| > V(x) - \epsilon) \\
&\leq \Gamma(y \in A(z) : U(y, \hat{p}(y), s^{\hat{\tau}}(y)) - \|y - x\| > V(x) - \epsilon),
\end{aligned}
$$

this shows that Eq. (A.18) holds. For the other case suppose that assume that $\bar{r} = 0$. This implies that $r_a = z$ for Γ^{p} almost all $a \in R_z$. Then, we must have

$$
\begin{aligned}
0 < \tau(A(z) \times A(z)) &= \int_{A_r(z)} s^{\hat{\tau}}(x)\, d\Gamma(x) \\
&= s^{\hat{\tau}}(z) \cdot \Gamma(\{z\}) + \int_{R_z} \int_{(z, r_a]} s^{\hat{\tau}}(x) \cdot d\Gamma_a(x) d\Gamma^{\mathrm{p}}(a) \\
&= s^{\hat{\tau}}(z) \cdot \Gamma(\{z\})
\end{aligned}
$$

Thus both $s^{\hat{\tau}}(z)$ and $\Gamma(\{z\})$ are strictly positive. If $\lambda(z) > 0$ then the same series of inequalities that we used for the previous case applies to this case, and so the desired Eq. (A.18) holds. If $\lambda(z) = 0$ then since by feasibility we have $0 < s^{\hat{\tau}}(z) \leq H_z(V(z|\, p, \tau))$, it must be the case that z belongs to $L^{\mathrm{supp}}(z)$. WLOG suppose that $z = X_a^{\mathrm{supp}}$ for some $a \in R_z$ then by the previous inequality we have that $0 < s^{\hat{\tau}}(X_a^{\mathrm{supp}}) \leq \frac{d\mu}{d\Gamma}(X_a^{\mathrm{supp}})$. In turn, this implies that $\mu(\{X_a^{\mathrm{supp}}\}) > 0$. This means that z has and initial mass of supply. Since z is a sink location, it does not belong to the indifference region of any other location and, therefore, by Lemma A.5 it does not send flow to any other location. Hence, $\tau_2(\{z\}) > 0$ and by Lemma A.2 we deduce that $U(z, p(z), s^{\tau}(z)) = V(z|\, p, \tau)$. Since, $\lambda(z) = 0$ this implies that

$V(z|\, p, \tau) = 0$. To conclude, note

$$
\begin{aligned}
\Gamma(y \in A(z) : \Pi(x, y, \hat{p}(y), s^{\hat{\tau}}(y)) > V(x|\, p, \tau) - \epsilon) &\geq \Gamma(y \in \{z\} : -|y - x| \\
&> -|z - x| - \epsilon) \\
&= \Gamma(\{z\}) \\
&> 0,
\end{aligned}
$$

hence Eq. (A.18) holds.

- **Equilibrium condition:** Consider the equilibrium set

$$
\tilde{\mathcal{E}} \triangleq \Big\{(x, y) \in A(z) \times A(z) : U(y, \hat{p}(y), s^{\hat{\tau}}(y)) - |y - x| = V_{A(z)}(x|\, \hat{p}, \hat{\tau})\Big\},
$$

we need to verify that $\hat{\tau}(\tilde{\mathcal{E}})$ equals $\tilde{\mu}(A(z))$. First, for $\hat{\tau}(\tilde{\mathcal{E}})$ we have

$$
\begin{aligned}
\hat{\tau}(\tilde{\mathcal{E}}) &\overset{(a)}{=} \hat{\tau}\Big(\Big\{(x, y) \in A(z) \times A_r(z) : \lambda(y) > 0,\ |z - y| + |y - x| = |z - x|\Big\}\Big) \\
&\quad + \hat{\tau}\Big(\Big\{(x, y) \in A(z) \times A_r(z) : \lambda(y) = 0,\ -|y - x| = V(x|\, p, \tau)\Big\}\Big)
\end{aligned}
$$

In (a) we use what we have just proved about $V_{A(z)}(x|\hat{p}, \hat{\tau})$, that $U(y, \hat{p}(y), s^{\hat{\tau}}(y)) = 0$ when $\lambda_(y) = 0$ and that $\hat{\tau}_2$ only puts mass in $A_r(z)$. Denote by Z the second term above, then

$$
\begin{aligned}
Z &= \hat{\tau}\Big(\Big\{(x,y) \in A(z) \times A_r(z) \setminus \{z\} : \lambda(y) = 0,\ -|y-x| = V(x|\,p,\tau)\Big\}\Big) \\
&\quad + \hat{\tau}\Big(\Big\{(x,y) \in A(z) \times \{z\} : \lambda(z) = 0,\ -|z-x| = V(z|\,p,\tau) - |z-x|\Big\}\Big) \\
&\overset{(a)}{=} \int_{R_z} \tilde{\gamma}_a\Big(\Big\{(x,y) \in (z, X_a] \times (z, r_a] : \lambda(y) = 0,\ -|y-x| = V(x|\,p,\tau)\Big\}\Big) d\Gamma^{\mathrm{p}}(a) \\
&\quad + \hat{\tau}_2(\{z\}) \cdot \mathbf{1}_{\{V(z)=0, \lambda(z)=0\}} \\
&\overset{(b)}{=} \int_{R_z} \tilde{\gamma}_a\Big(\Big\{(x,y) \in [r_a, X_a] \times \{r_a\} : \lambda(y) = 0,\ -|y-x| = V(x|\,p,\tau)\Big\}\Big) \mathbf{1}_{\{r_a \neq z\}} d\Gamma^{\mathrm{p}}(a) \\
&\quad + \hat{S}(\{z\}) \cdot \mathbf{1}_{\{V(z)=0, \lambda(z)=0\}} \\
&= \int_{R_z} \tilde{\gamma}_a\Big(\Big\{(x,y) \in [r_a, X_a] \times \{r_a\} : \lambda(y) = 0,\ V(r_a|\,p,\tau) = 0\Big\}\Big) \mathbf{1}_{\{r_a \neq z\}} d\Gamma^{\mathrm{p}}(a) \\
&\quad + \hat{S}(\{z\}) \cdot \mathbf{1}_{\{V(z)=0, \lambda(z)=0\}} \\
&\overset{(c)}{=} \int_{R_z} \hat{S}_a(\{r_a\}) \mathbf{1}_{\{\lambda(r_a)=0,\, V(r_a)=0\}} \mathbf{1}_{\{r_a \neq z\}} d\Gamma^{\mathrm{p}}(a) + \hat{S}(\{z\}) \cdot \mathbf{1}_{\{V(z)=0, \lambda(z)=0\}} \\
&= \hat{S}((\partial A_r \cup \{z\}) \cap \{y : \lambda(y) = 0, V(y|\,p,\tau) = 0\})
\end{aligned}
$$

where (a) comes from the fact that $\tilde{\gamma}_a$ only puts mass in $(z, X_a] \times (z, X_a]$. The equality in (b) follows from the congestion bound in Eq. (A.15) which makes $\hat{s}(y)$ equal to zero when $\lambda(y) = 0$ and $y \in (z, r_a)$, and also the fact that $\tilde{\gamma}_a$ only sends flows towards z and not in the opposite direction, that is, $\tilde{\gamma}_a((z, r_a) \times \{r_a\}) = 0$ (see Eq. (A.17)). The last equality, (c), uses the latter fact once again.

Consider the sets

$$\tilde{\mathcal{E}}_c \triangleq A(z) \times \{y \in \{z\} : \lambda(y) > 0\},$$

and

$$\tilde{\mathcal{E}}_a \triangleq \Big\{(x,y) \in (z, X_a] \times (z, r_a] : \lambda(y) > 0,\ y \leq x\Big\}.$$

Then,

$$\hat{\tau}(\tilde{\mathcal{E}}) = \tilde{\tau}(\tilde{\mathcal{E}}_c) + \int_{R_z} \tilde{\gamma}_a(\tilde{\mathcal{E}}_a) d\Gamma^{\mathrm{p}}(a) + Z.$$

For the first term we have

$$\hat{\tau}(\tilde{\mathcal{E}}_c) = \hat{\tau}_2(\{y \in \{z\} : \lambda(y) > 0\}) = \hat{\tau}_2(\{z\}) \cdot \mathbf{1}_{\{\lambda(z)>0\}}.$$

For the second term we have that for any ray a, $\tilde{\gamma}_a(\tilde{\mathcal{E}}_a)$ equals $\hat{S}_a((z, r_a] \cap \{y : \lambda(y) > 0\})$. This is true because the plan $\tilde{\gamma}_a$ only sends mass to $(z, r_a]$(this is the support of $\hat{S}_a$) and it does not send mass in the opposite direction of z, see Eq. (A.17). Therefore,

$$\begin{aligned}
\hat{\tau}(\tilde{\mathcal{E}}) &= \hat{\tau}_2(\{z\}) \cdot \mathbf{1}_{\{\lambda(z)>0\}} + \int_{R_z} \hat{S}_a((z, r_a] \cap \{y : \lambda(y) > 0\}) d\Gamma^{\mathrm{p}}(a) + Z \\
&= \hat{S}(\{y \in A_r(z) : \lambda(y) > 0\}) + \hat{S}((L_r(z) \cup \{z\}) \cap \{y : \lambda(y) = 0, V(y|\, p, \tau) = 0\})
\end{aligned}$$

Now, recall that $\tilde{\mu}(A(z)) = \hat{S}(A_r(z))$ and

$$\begin{aligned}
\hat{S}(A_r(z)) &= \hat{S}(A_r(z) \cap \{y : \lambda(y) > 0\}) + \hat{S}(A_r(z) \cap \{y : \lambda(y) = 0\}) \\
&\overset{(a)}{=} \hat{S}(\{y \in A_r(z) : \lambda(y) > 0\}) + \hat{S}(\{z\}) \mathbf{1}_{\{\lambda(z)=0\}} \\
&+ \int_{R_z} \hat{S}_a(A_r(z) \cap \{y : \lambda(y) = 0\}) d\Gamma^{\mathrm{p}}(a) \\
&\overset{(b)}{=} \hat{S}(\{y \in A_r(z) : \lambda(y) > 0\}) + \hat{S}(\{z\}) \mathbf{1}_{\{\lambda(z)=0\}} \\
&+ \int_{R_z} \hat{S}_a(\{r_a\}) \mathbf{1}_{\{\lambda(r_a)=0\}} \mathbf{1}_{\{r_a \neq z\}} d\Gamma^{\mathrm{p}}(a) \\
&= \hat{S}(\{y \in A_r(z) : \lambda(y) > 0\}) + \hat{S}((\partial A_r \cup \{z\}) \cap \{y : \lambda(y) = 0\})
\end{aligned}$$

For the second term in (a) we use the disintegration of $\hat{S}$, and in (b) we use the congestion bound in Eq. (A.15). Hence, if we show that

$$\hat{S}((L_r(z) \cup \{z\}) \cap \{y : \lambda(y) = 0, V(y|p, \tau) = 0\}) = \hat{S}((L_r(z) \cup \{z\}) \cap \{y : \lambda(y) = 0\}),$$

the proof will be complete. Let $Q = (\partial A_r \cup \{z\}) \cap \{y : \lambda(y) = 0, V(y|\, p, \tau) > 0\}$ then

$$\begin{aligned}
\hat{S}(Q) &= \hat{S}\Big((\partial A_r \cup \{z\}) \cap \{y : \lambda(y) = 0, V(y|\, p, \tau) > 0\}\Big) \\
&= \hat{S}\Big(\underbrace{(\partial A_r \cup \{z\}) \cap \{y \in L^{\mathrm{supp}}(z) : \lambda(y) = 0, V(y|\, p, \tau) > 0\}}_{Q_2}\Big).
\end{aligned}$$

Assume that $\hat{S}(Q_2) > 0$ then

$$\begin{aligned}\tau_2(\cup_{X_a^{\text{supp}} \in Q_2}(z, X_a^{\text{supp}})) &\leq \hat{S}(\cup_{X_a^{\text{supp}} \in Q_2}(z, X_a^{\text{supp}})) \\ &< \hat{S}(\cup_{X_a^{\text{supp}} \in Q_2}(z, X_a^{\text{supp}})) + \hat{S}(Q_2) \\ &= \hat{S}(\cup_{X_a^{\text{supp}} \in Q_2}(z, X_a^{\text{supp}}]) \\ &\leq \tau_2(\cup_{X_a^{\text{supp}} \in Q_2}(z, X_a^{\text{supp}}])\end{aligned}$$

where the first inequality comes from the congestion bound in Eq. (A.15) and the definition of Q_2, the second from $\hat{S}(Q_2) > 0$ and the last from the feasibility of $\hat{s}$. Hence $\tau_2(Q_2) > 0$ and, therefore, Lemma A.2 implies that in Q_2 we have $U(y, p(y), s^\tau(y)) = V(y|\, p, \tau)$. Since, $\lambda(y) = 0$ for $y \in Q_2$ we conclude that in this case we cannot have $\hat{S}(Q_2) > 0$. This completes the proof.

Part 6: Conclusion. We conclude by applying Proposition 1.4. The price-equilibrium pair $(\hat{p}, \hat{\tau})$ satisfies the hypothesis in Proposition 1.4, so we can create a global price-equilibrium pair which we still denote by $(\hat{p}, \hat{\tau})$ in $\mathcal{C}$. This new solution has the same objective that (p, τ) in $A(z)^c$, but it dominates the platform revenue in $A(z)$. Therefore, $(\hat{p}, \hat{\tau})$ revenue dominates (p, τ).

Proof of Eq. (A.17): We show that

$$\tilde{\gamma}_a\Big(\{(x, y) \in (z, X_a] \times (z, X_a] : y > x\}\Big) = 0, \quad \Gamma^{\mathfrak{p}} - a.e.\ a \in R_z.$$

First we show that both measures $\tilde{\mu}_a^{\mathfrak{r}}$ and $\hat{S}_a$ satisfy:

$$\tilde{\mu}_a^{\mathfrak{r}}((z, b_a]) \leq \hat{S}_a((z, c_a]) \quad \forall b_a, c_a \in (z, X_a],\, b_a \leq c_a, \quad \Gamma^{\mathfrak{p}} - a.e.\ a \in R_z, \tag{A.21}$$

where b_a and c_a lie in the ray indexed by $a \in R_z$. To see why this is true let us proceed by contradiction. Let us denote by Q de set where Eq. (A.21) is not satisfied, we have that $\Gamma^{\mathfrak{p}}(Q) > 0$. Note that for any $a \in Q$ we can find b_a and c_a for which the inequality in Eq. (A.21) is not satisfied, so let us thus fix such collection of b_a and c_a. Moreover, from the definition of $\tilde{\mu}_a^{\mathfrak{r}}$ we deduce that for any $a \in Q$ we have $h_a \leq b_a$

(otherwise $\tilde{\mu}^{\mathtt{r}}_a((z,b_a]) = 0$ and, as a consequence, a could not belong to Q). Then,

$$\begin{aligned}
\int_Q \tilde{\mu}^{\mathtt{r}}_a((z,b_a])d\Gamma^{\mathtt{p}}(a) &= \int_Q \tilde{\mu}_a((z,b_a] \cap (h_a, X_a]) + \mathbf{1}_{\{h_a \in (z,b_a]\}} \cdot (\tilde{\mu}_a(z,h_a] - \Delta_a) d\Gamma^{\mathtt{p}}(a) \\
&\overset{(a)}{\leq} \int_Q \tilde{\mu}_a((z,b_a] \cap (h_a, X_a]) + \mathbf{1}_{\{h_a \leq b_a\}} \cdot (\tilde{\mu}_a(z,h_a] - \Delta_a) d\Gamma^{\mathtt{p}}(a) \\
&= \int_Q (\tilde{\mu}_a((h_a,b_a]) + \tilde{\mu}_a((z,h_a]) - \Delta_a) d\Gamma^{\mathtt{p}}(a) \\
&= \int_Q (\tilde{\mu}_a((z,b_a]) - \Delta_a) d\Gamma^{\mathtt{p}}(a) \\
&= \underbrace{\int_{Q \cap \{a : r_a \leq b_a\}} (\tilde{\mu}_a((z,b_a]) - \Delta_a) d\Gamma^{\mathtt{p}}(a)}_{(*)} \\
&\quad + \underbrace{\int_{Q \cap \{a : r_a > b_a\}} (\tilde{\mu}_a((z,b_a]) - \Delta_a) d\Gamma^{\mathtt{p}}(a)}_{(**)},
\end{aligned}$$

where (a) follows from $\tilde{\mu}_a(z,h_a] \geq \Delta_a$. For $(*)$ we have

$$\begin{aligned}
(*) &= \int_{Q \cap \{a : r_a \leq b_a\}} (\tilde{\mu}_a((z,b_a]) - \tilde{\mu}((z,X_a]) + \hat{S}_a((z,X_a])) d\Gamma^{\mathtt{p}}(a) \\
&= \int_{Q \cap \{a : r_a \leq b_a\}} (-\tilde{\mu}((b_a,X_a]) + \hat{S}_a((z,X_a])) d\Gamma^{\mathtt{p}}(a) \\
&\leq \int_{Q \cap \{a : r_a \leq b_a\}} \hat{S}_a((z,X_a]) d\Gamma^{\mathtt{p}}(a) \\
&= \int_{Q \cap \{a : r_a \leq b_a\}} \hat{S}_a((z,c_a]) d\Gamma^{\mathtt{p}}(a),
\end{aligned}$$

the last inequality holds because

Now we analyze $(**)$. Denote by Q^r the set of rays $a \in R_z$ such that $r_a > b_a$ and

$a \in Q$. Then

$$\begin{aligned}
\int_{Q^r} \tilde{\mu}_a((z,b_a])d\Gamma^{\mathfrak{p}}(a) &= \tilde{\mu}\Big(\bigcup_{a\in Q^r}(z,b_a]\Big) \\
&= \tau\Big(\bigcup_{a\in Q^r}(z,b_a] \times \bigcup_{a\in Q^r}(z,X_a]\Big) + \underbrace{\tau\Big(\bigcup_{a\in Q^r}(z,b_a]\times\{z\}\Big)}_{\triangleq \ell_r} \\
&= \tau\Big(\bigcup_{a\in Q^r}(z,b_a] \times \bigcup_{a\in Q^r}(z,b_a]\Big) + \tau\Big(\bigcup_{a\in Q^r}(z,b_a] \times \bigcup_{a\in Q^r}(b_a,X_a]\Big) \\
&+ \ell_r \\
&= \tau\Big(\bigcup_{a\in Q^r}(z,b_a] \times \bigcup_{a\in Q^r}(z,b_a]\Big) + \ell_r \\
&\leq \tau_2\Big(\bigcup_{a\in Q^r}(z,b_a]\Big) + \ell_r \\
&\leq \hat{S}\Big(\bigcup_{a\in Q^r}(z,b_a]\Big) + \ell_r \\
&= \int_{Q^r} \hat{S}_a((z,c_a])d\Gamma^{\mathfrak{p}}(a) + \ell_r,
\end{aligned}$$

the first equality comes from the definition of $\tilde{\mu}_a$ and then integrating this disintegration of measures. The second and fourth equality come from Proposition 1.3 part *(ii)*. The last inequality comes from the congestion bound. For Δ_a we have

$$\begin{aligned}
\int_{Q^r} \Delta_a d\Gamma^{\mathfrak{p}}(a) &= \tilde{\mu}\Big(\bigcup_{a\in Q^r}(z,X_a]\Big) - \hat{S}\Big(\bigcup_{a\in Q^r}(z,X_a]\Big) \\
&= \tau\Big(\bigcup_{a\in Q^r}(z,X_a] \times \bigcup_{a\in Q^r}(z,X_a]\Big) + \tau\Big(\bigcup_{a\in Q^r}(z,X_a]\times\{z\}\Big) \\
&- \hat{S}\Big(\bigcup_{a\in Q^r}(z,X_a]\Big) \\
&\geq \tau\Big(\bigcup_{a\in Q^r}(z,X_a] \times \bigcup_{a\in Q^r}(z,X_a]\Big) + \ell_r - \hat{S}\Big(\bigcup_{a\in Q^r}(z,X_a]\Big) \\
&= \tau_2\Big(\bigcup_{a\in Q^r}(z,X_a]\Big) + \ell_r - \hat{S}\Big(\bigcup_{a\in Q^r}(z,X_a]\Big) \\
&\geq \ell_r,
\end{aligned}$$

where the last inequality comes from Eq. (FR_a). As a consequence we deduce that

$$(**) = \int_{Q\cap\{a:r_a>b_a\}} (\tilde{\mu}_a((z,b_a]) - \Delta_a)d\Gamma^{\mathfrak{p}}(a) \leq \int_{Q\cap\{a:r_a>b_a\}} \hat{S}_a((z,b_a])d\Gamma^{\mathfrak{p}}(a)$$

Putting together the bounds for $(*)$ and $(**)$ we deduce that

$$\int_Q \tilde{\mu}_a^{\mathfrak{r}}((z,b_a])d\Gamma^{\mathfrak{p}}(a) \leq \int_Q \hat{S}_a((z,c_a])d\Gamma^{\mathfrak{p}}(a),$$

since $\Gamma^{\mathfrak{p}}(Q) > 0$ the previous inequality yields a contradiction. We conclude that Eq. (A.21) holds.

To finalize the proof of Eq. (A.17). Consider the set where Eq. (A.21) holds (the complement of this set has $\Gamma^{\mathfrak{p}}$ measure equal to zero). for any ray a in this suppose that

$$\tilde{\gamma}_a\Big(\{(x,y) \in (z,X_a] \times (z,X_a] : y > x\}\Big) > 0.$$

From the proof of Lemma A.5 we deduce that there exists $(x,y) \in (z,X_a] \times (z,X_a]$ such that $y > x$ and $\tilde{\gamma}_a((z,x+\delta] \times (y-\delta,X_a)) > 0$, where $\delta > 0$ can be taken small enough such that $x+\delta < y-\delta$. Then,

$$\begin{aligned}
\hat{S}_a((z,x+\delta]) &\geq \tilde{\mu}_a^{\mathfrak{r}}((z,x+\delta]) \\
&= \tilde{\gamma}_a((z,x+\delta] \times (z,X_a]) \\
&= \tilde{\gamma}_a((z,x+\delta] \times (z,x+\delta]) + \tilde{\gamma}_a((z,x+\delta] \times (x+\delta,X_a]) \\
&> \tilde{\gamma}_a((z,x+\delta] \times (z,x+\delta]) \\
&= \hat{S}_a((z,x+\delta]) - \tilde{\gamma}_a((x+\delta,X_a] \times (z,x+\delta]),
\end{aligned}$$

Thus,

$$\tilde{\gamma}_a((x+\delta,X_a] \times (z,x+\delta]) > 0, \text{ and we also have } \tilde{\gamma}_a((z,x+\delta] \times (y-\delta,X_a)) > 0,$$

but this is not possible because $\tilde{\gamma}_a$ is an optimal transport and, therefore, it is concentrated on a c-cyclically monotone set where $c(x,y) = \|x-y\|$, see [66]. This concludes the proof of Eq. (A.17).

□

Lemma A.6 *Let ν be a non-negative measure in $\mathcal{C}$, and π_1 a mapping be such that $\pi_1(x,y) = x$. Consider any measurable subset K of $\mathcal{C}$ and some $z \in \mathcal{C}$ then the mappings $\nu(\pi_1(\,\cdot\, \cap \mathcal{D}) \cap K)$ and $\nu(\pi_1(\,\cdot\, \cap (K \times \{z\})))$, defined on the Borel sets of $\mathcal{C} \times \mathcal{C}$, belong to $\mathcal{M}(\mathcal{C} \times \mathcal{C})$.*

Proof of Lemma A.6. For any Borel set $\mathcal{L} \subset \mathcal{C} \times \mathcal{C}$ define

$$\tau_a(\mathcal{L}) \triangleq \nu(\pi_1(\mathcal{L} \cap \mathcal{D}) \cap K) \text{ and } \tau_b(\mathcal{L}) \triangleq \nu(\pi_1(\mathcal{L} \cap (K \times \{z\}))).$$

We show that $\tau_a, \tau_b \in \mathcal{M}(\mathcal{C} \times \mathcal{C})$. Note that because $\nu \in \mathcal{M}(\mathcal{C})$ for $i \in \{a, b\}$ we have that $\tau_i(\emptyset) = 0$, and for any Borel set $\mathcal{L} \subseteq \mathcal{C} \times \mathcal{C}$ that $\tau_i(\mathcal{L}) \in [0, \infty)$. To verify $\sigma-$additivity consider a countable partition $\{\mathcal{L}_n\}_{n \in \mathbb{N}} \subseteq \mathcal{C} \times \mathcal{C}$, we need to show that

$$\tau_i\Big(\bigcup_{n \in \mathbb{N}} \mathcal{L}_n\Big) = \sum_{n \in \mathbb{N}} \tau_i(\mathcal{L}_n).$$

Note that from the definition of $\mathcal{D}$ and the fact the set $K \times \{z\}$ has second component equal to 0, both collections $\{\pi_1(\mathcal{L}_n \cap \mathcal{D})\}_{n \in \mathbb{N}}$ and $\{\pi_1(\mathcal{L}_n \cap (K \times \{z\}))\}_{n \in \mathbb{N}}$ form a partition. Given this we can verify $\sigma-$ additivity, we do it for both τ_a and τ_b at the same time

$$\begin{aligned}
\tau_a\Big(\bigcup_{n \in \mathbb{N}} \mathcal{L}_n\Big) + \tau_b\Big(\bigcup_{n \in \mathbb{N}} \mathcal{L}_n\Big) &= \nu(\pi_1(\bigcup_{n \in \mathbb{N}} \mathcal{L}_n \cap \mathcal{D}) \cap K) + \nu(\pi_1(\bigcup_{n \in \mathbb{N}} \mathcal{L}_n \cap K \times \{z\})) \\
&= \nu(\bigcup_{n \in \mathbb{N}} \pi_1(\mathcal{L}_n \cap \mathcal{D}) \cap K) + \nu(\bigcup_{n \in \mathbb{N}} \pi_1(\mathcal{L}_n \cap K \times \{z\})) \\
&= \sum_{n \in \mathbb{N}} \nu(\pi_1(\mathcal{L}_n \cap \mathcal{D}) \cap K) + \sum_{n \in \mathbb{N}} \nu(\pi_1(\mathcal{L}_n \cap K \times \{z\})) \\
&= \sum_{n \in \mathbb{N}} \tau_a(\mathcal{L}_n) + \sum_{n \in \mathbb{N}} \tau_b(\mathcal{L}_n),
\end{aligned}$$

where the third line comes from the $\sigma-$additivity of the ν measure. Thus $\tau \in \mathcal{M}(\mathcal{C} \times \mathcal{C})$.

□

A.3 Proofs for Section 1.6

A.3.1 Preliminars

We use $m \in \mathcal{M}(\mathcal{C})$ denotes the Lebesgue measure in $\mathcal{C}$. We use $\mathcal{D}$ to denote the subset of $\mathcal{C}\times\mathcal{C}$ with equal first and second components, that is, $\mathcal{D} = \{(x, y) \in \mathcal{C}\times\mathcal{C} : x = y\}$. For any measurable set $\mathcal{B} \subseteq \mathcal{C}$ and a price-equilibrium pair (p, τ) we denote the platform's revenue in $\mathcal{B}$ under (p, τ) by $\mathbf{Rev}_{\mathcal{B}}(p, \tau)$. In case that $\mathcal{B}$ is $\mathcal{C}$ we simply use $\mathbf{Rev}(p, \tau)$.

Proof of Proposition 1.6. Let (p, τ) be any feasible price-equilibrium pair by Lemma A.7 (which state and prove after this proof) we have $V(x|p, \tau) \leq \psi_1$, Γ almost everywhere in $\mathcal{C}_\lambda = \mathcal{C} \setminus \{0\}$. This yields the following upper bound for the platform's objective

$$\int_{\mathcal{C}_\lambda} V(x|\, p, \tau) \cdot s^\tau(x)\, dx \leq \psi_1 \cdot \int_{\mathcal{C}_\lambda} s^\tau(x)\, dx \leq \psi_1 \cdot \mu_1 \cdot m(\mathcal{C}).$$

The maximum revenue the platform can achieve in this case is bounded above by $\gamma \cdot \psi_1 \cdot \mu_1 \cdot m(\mathcal{C})$. Next, we show that the solution given in the statement of the lemma is feasible and achieves the upper bound.

Flow feasibility. We show that $\tau \in \mathcal{F}(\mu)$. A complete definition of the measure τ is $\tau(\mathcal{L}) = \mu(\pi_1(\mathcal{L} \cap \mathcal{D})$. From the definition of τ it is clear that $\tau \in \mathcal{M}(\mathcal{C})$. Furthermore, τ_1 coincides with μ and so does τ_2. Since μ is the Lebesgue measure times a constant and Γ is the Lebesgue measure plus an atom, we have $\tau_1, \tau_2 \ll \Gamma$. From this we can deduce that $m-$ a.e in $\mathcal{C}_\lambda$, $s^\tau(x)$ equals μ_1.

Equilibrium utilities. We show that $V(x|\, p, \tau)$ equals ψ_1. Note that

$$U(y, p(y), s^\tau(y)) = \psi_1, \quad \Gamma - a.e.\ y \text{ in } \mathcal{C}_\lambda.$$

Fix $x \in \mathcal{C}$, we have that

$$\Gamma(\{y \in \mathcal{C} : U(y, p(y), s^\tau(y)) - |y - x| > \psi_1\}) = \mathbf{1}_{\{0 - |0-x| > \psi_1\}} + \Gamma(\{y \in \mathcal{C} \setminus \{0\} : -|y - x| > 0\}) = 0.$$

Moreover, for any $\epsilon > 0$

$$\Gamma(\{y \in \mathcal{C} : U(y, p(y), s^{\tau}(y)) - |y - x| > \psi_1 - \epsilon\}) \geq \Gamma(\{y \in \mathcal{C}_{\lambda} : -|y - x| > \epsilon\}) > 0,$$

where the last inequality comes from the fact that Γ corresponds to the Lebesgue measure (plus an atom). That is, $V(x|\, p, \tau)$ equals ψ_1.

Equilibrium condition. Consider the equilibrium set

$$\mathcal{E} \triangleq \Big\{(x, y) \in \mathcal{C} \times \mathcal{C} : U(y, p(y), s^{\tau}(y)) - |y - x| = V(x|\, p, \tau)\Big\}.$$

Then,

$$\tau(\mathcal{E}) = \tau\Big(\Big\{(x, y) \in \mathcal{C} \times \{0\} : -|y - x| = \psi_1\Big\}\Big) + \tau\Big(\Big\{(x, y) \in \mathcal{C} \times \mathcal{C}_{\lambda} : -|y - x| = 0\Big\}\Big) = \mu(\mathcal{C}).$$

We have proven that the solution is the statement is feasible, and because of the values of $V(\cdot|\, p, \tau)$ and $s^{\tau}(\cdot)$ we conclude that this solution achieves the upper bound.

□

Lemma A.7 *Let p be any price mapping and τ a corresponding equilibrium flow. Then for any measurable set $\mathcal{B} \subseteq \mathcal{C}_{\lambda}$ such that $0 \notin \mathcal{B}$ and $\tau(\mathcal{B} \times \mathcal{B}^c) = 0$ we have*

$$V(x|p, \tau) \;\leq\; \psi_1, \quad \Gamma - a.e.\ x \ \text{ in } \mathcal{B}.$$

Furthermore, in the pre-shock environment we can replace $\mathcal{B}$ with $\mathcal{C}_{\lambda}$ in the inequality above.

Proof of Lemma A.7. Define the set

$$\mathcal{L} \triangleq \{x \in \mathcal{B} : V(x|p, \tau) \;\leq\; \psi_1\}.$$

We would like to show that $\Gamma(\mathcal{L}^c) = 0$ where the complement is taken with respect to $\mathcal{B}$. Suppose this is not the case, and note that

$$\mu_1 \cdot m(\mathcal{L}^c) = \mu(\mathcal{L}^c) = \tau(\mathcal{L}^c \times \mathcal{C}) = \tau(\mathcal{L}^c \times \mathcal{B}) + \tau(\mathcal{L}^c \times \mathcal{B}^c),$$

since $\mathcal{L}^c \subseteq \mathcal{B}$ and $\tau(\mathcal{B} \times \mathcal{B}^c) = 0$, the second term in the expression above is zero. This yields,

$$\begin{aligned}
\mu_1 \cdot m(\mathcal{L}^c) &= \tau(\mathcal{L}^c \times \mathcal{B}) \\
&= \tau(\mathcal{L}^c \times \mathcal{B} \cap \mathcal{L}^c) + \tau(\mathcal{L}^c \times \mathcal{B} \cap \mathcal{L}) \\
&= \tau(\mathcal{L}^c \times \mathcal{L}^c) + \tau(\mathcal{L}^c \times \mathcal{L})
\end{aligned}$$

There are two cases. First, if $\tau(\mathcal{L}^c \times \mathcal{L}) > 0$ then by Lemma A.5 there exists a pair $(x, y) \in \mathcal{L}^c \times \mathcal{L}$ such that $y \in \mathcal{IR}(x|p, \tau)$. Therefore, by Lemma 1.3 we have

$$V(y|p, \tau) = V(x|p, \tau) + |x - y|.$$

However, since $(x, y) \in \mathcal{L}^c \times \mathcal{L}$

$$V(y|p, \tau) \leq \psi_1 \text{ and } V(x|p, \tau) > \psi_1.$$

Using the previous equation we can deduce that $\psi_1 > \psi_1$, which is not possible. The second case is $\tau(\mathcal{L}^c \times \mathcal{L}) = 0$. Note that

$$\tau_2(\mathcal{L}^c) = \tau(\mathcal{C} \times \mathcal{L}^c) \geq \tau(\mathcal{L}^c \times \mathcal{L}^c) = \mu_1 \cdot m(\mathcal{L}^c).$$

We also have that

$$\tau_2(\mathcal{L}^c) = \int_{\mathcal{L}^c} s^\tau(x) d\Gamma(x) \leq \int_{\mathcal{L}^c} \psi_x^{-1}(V(x|\, p, \tau)) d\Gamma(x) < \mu_1 \cdot \Gamma(\mathcal{L}^c),$$

where the first inequality comes from Proposition 1.5, and the second from the fact that $\psi_x(\cdot)$ is a strictly decreasing function, the definition of $\mathcal{L}^c$ and $\Gamma(\mathcal{L}^c) > 0$. Note that this inequality holds in both of the cases in the statement of the lemma. In both cases we have $0 \notin \mathcal{B}$ so $\Gamma(\mathcal{L}^c)$ equals $m(\mathcal{L}^c)$, yielding

$$\mu_1 \cdot m(\mathcal{L}^c) \leq \tau_2(\mathcal{L}^c) < \mu_1 \cdot \Gamma(\mathcal{L}^c) = \mu_1 \cdot m(\mathcal{L}^c).$$

□

A.3.2 Proofs for Section 1.6.2

Proof of Proposition 1.7. The proof of this proposition consists of several steps. In the first step we establish that the origin is an attraction region, characterize some properties of it and compute the value of the equilibrium utility function outside the attraction region. After this step, the drivers utility function will be pinned down in the entire city as a function of its value in the origin, $V(0|\, p, \tau)$. The second step supplies us with a full characterization, up to $V(0|\, p, \tau)$, of the post-relocation supply τ_2 in the entire city. Finally, in step three we show how to solve for the optimal value of $V(0|\, p, \tau)$ and, therefore, we pin down both $V(\cdot|\, p, \tau)$ and τ_2. We further show how to find the optimal $p(0)$ and the corresponding optimal flow τ.

Step 1: We show that we can restrict attention to solutions (p, τ) such that $X_l < 0 < X_r$, $X_r = V(0) - \psi_1$ and $X_l = -X_r$. Furthermore, such solutions have $V(x|\, p, \tau) = \psi_1$ for all $x \in \mathcal{C} \setminus [X_l, X_r]$.

Proof of Step 1: Let (p, τ) be a feasible solution. First, we show that at any optimal solution we must have $X_l < 0 < X_r$. By Lemma A.8 (which we state and prove after the proof of the present proposition) we have that if either of the sets $\{x \in (0, H] : 0 \in \mathcal{IR}(x|p, \tau)\}$ or $\{x \in [-H, 0) : 0 \in \mathcal{IR}(x|p, \tau)\}$ is empty then the revenue the platform makes satisfies

$$\frac{1}{\gamma} \cdot \mathbf{Rev}(p, \tau) \leq \psi_1 \cdot \mu_1 \cdot 2 \cdot H.$$

Now we construct a new feasible solution $(\tilde{p}, \tilde{\tau})$ for which both sets are non-empty and such that

$$\frac{1}{\gamma} \cdot \mathbf{Rev}(\tilde{p}, \tilde{\tau}) > \psi_1 \cdot \mu_1 \cdot 2 \cdot H, \tag{A.22}$$

where $\tilde{p}$ equals ρ_1 in $\mathcal{C} \setminus \{0\}$ and $p(0)$ is appropriately chosen. This will imply that any optimal solution must satisfy $\{x \in (0, H] : 0 \in \mathcal{IR}(x|p, \tau)\} \neq \emptyset$ and $\{x \in [-H, 0) : 0 \in \mathcal{IR}(x|p, \tau)\} \neq \emptyset$ and, therefore, $X_l < 0 < X_r$. This also implies that the optimal revenue in this case is strictly larger than the one in the pre-shock environment.

Our solution will send flow in $[-h, h]$ to the origin, where $h > 0$ is to be determined. Inside this interval, all the flow in the subinterval $[-\bar{h}(h), \bar{h}(h)]$ goes to the origin where $0 \leq \bar{h}(h) \leq h$. The rest of the flow in $[-h, h]$ partially stays at its original position and partially goes to the origin. We now show how to determine $\bar{h}(h)$ and h. For any given $h > 0$ we define

$$\bar{h}(h) \triangleq (\psi_1 + h - \alpha \cdot \rho_1)^+,$$

note that when ψ_1 equals $\alpha \cdot \rho_1$ we have that $\bar{h}(h)$ equals h, and we will send all the flow in $[-h, h]$ to the origin. However, when $\psi_1 < \alpha \cdot \rho_1$ not all the flow will be sent to the origin. Define

$$\mu_1(x) \triangleq \alpha \cdot \rho_1 \cdot \frac{\lambda_1 \overline{F}(\rho_1)}{\psi_1 + h - |x|},$$

then

$$\frac{\lambda_1 \overline{F}(\rho_1)}{\mu_1(x)} \leq 1, \quad x \in [-h, h] \setminus [-\bar{h}(h), \bar{h}(h)].$$

The idea is that for every location $x \in K(h) \triangleq [-h, h] \setminus [-\bar{h}(h), \bar{h}(h)]$ we will leave a density $\mu_1(x)$ of flow there and send $\mu_1 - \mu_1(x)$ (note that this difference is non-negative) to the origin. In order to make this possible, we need to chose h appropriately. Observe that the total supply we will send to the origin is

$$S_T(h) = 2\bar{h}(h)\mu_1 + 2\int_{\bar{h}(h)}^{h} (\mu_1 - \mu_1(x))\, dm(x),$$

where $\lim_{h \to 0} S_T(h) = 0$. Hence, since $\psi_1 < \overline{\alpha} \cdot \overline{V}$, we can always find $h > 0$ such that

$$\alpha \cdot \overline{V} - h \geq \alpha \cdot F^{-1}\Big(1 - \frac{S_T(h)}{\lambda_0}\Big) - h \geq \psi_1. \tag{A.23}$$

This yields

$$\overline{F}\Big(\frac{\psi_1 + h}{\alpha}\Big) \geq \frac{S_T(h)}{\lambda_0}.$$

Now we construct the solution $(\tilde{p}, \tilde{\tau})$. Fix any h satisfying Eq. (A.23) and consider prices defined by

$$\tilde{p}(x) = \begin{cases} \frac{\psi_1 + h}{\alpha} & \text{if } x = 0 \\ \rho_1 & \text{if } x \in \mathcal{C} \setminus \{0\}, \end{cases}$$

and flows for any measurable set $\mathcal{L} \subseteq \mathcal{C} \times \mathcal{C}$ defined by

$$\tilde{\tau}(\mathcal{L}) = \mu(\pi_1(\mathcal{L} \cap \mathcal{D}) \cap [-h, h]^c) + \mu(\pi_1(\mathcal{L} \cap [-\bar{h}(h), \bar{h}(h)] \times \{0\}))$$
$$+ G_0(\pi_1(\mathcal{L} \cap K(h) \times \{0\})) + G_1(\pi_1(\mathcal{L} \cap \mathcal{D}) \cap K(h)),$$

where G_0, G_1 are measures defined for any measurable set $\mathcal{B} \subseteq K(h)$ by

$$G_0(\mathcal{B}) \triangleq \int_{\mathcal{B}} (\mu_1 - \mu_1(x))\, dm(x), \quad G_1(\mathcal{B}) \triangleq \int_{\mathcal{B}} \mu_1(x)\, dm(x).$$

We argue that $(\tilde{p}, \tilde{\tau})$ is a feasible solution that complies with Eq. (A.22). From Lemma A.6 we have that $\tilde{\tau} \in \mathcal{M}(\mathcal{C} \times \mathcal{C})$, also note that for any measurable set $\mathcal{B} \subseteq \mathcal{C}$ the first marginal of $\tilde{\tau}$ satisfies

$$\tilde{\tau}_1(\mathcal{B}) = \mu(\mathcal{B} \cap [-h, h]^c) + \mu(\mathcal{B} \cap [-\bar{h}(h), \bar{h}(h)]) + G_0(\mathcal{B} \cap K(h)) + G_1(\mathcal{B} \cap K(h)) = \mu(\mathcal{B}).$$

The post-relocation supply measure is

$$\tilde{\tau}_2(\mathcal{B}) = \mu(\mathcal{B} \cap [-h, h]^c) + S_T(h) \cdot \mathbf{1}_{\{0 \in \mathcal{B}\}} + G_1(\mathcal{B} \cap K(h)),$$

clearly $\tilde{\tau}_2 \ll \Gamma$. Therefore, $\tilde{\tau} \in \mathcal{F}(\mu)$. Next, we need to show that $\tilde{\tau}$ is a supply equilibrium. The Radon-Nikodym derivative of $\tilde{\tau}_2$ with respect the city measure is (Γ-a.e)

$$s(x) = \begin{cases} S_T(h) & \text{if } x = 0 \\ 0 & \text{if } x \in [-\bar{h}(h), \bar{h}(h)] \setminus \{0\} \\ \mu_1(x) & \text{if } x \in K(h) \\ \mu_1 & \text{if } x \in [-h, h]^c. \end{cases}$$

Indeed,

$$\int_{\mathcal{L}} s(x)\, d\Gamma(x) = S_T(h) \mathbf{1}_{\{0 \in \mathcal{L}\}} + \int_{\mathcal{L} \cap [-h,h]^c} \mu_1\, dm(x) + \int_{\mathcal{L} \cap K(h)} \mu_1(x)\, dm(x) = \tilde{\tau}_2(\mathcal{L}),$$

that is, $\frac{d\tilde{\tau}_2}{d\Gamma}(\cdot)$ equals $s(\cdot)$ Γ-a.e. From this we can compute $V(\cdot|\tilde{p},\tilde{\tau})$. Note that ($\Gamma$-a.e)

$$\tilde{U}(y) = U\Big(y, \tilde{p}(y), \frac{d\tilde{\tau}_2}{d\Gamma}(y)\Big) = \begin{cases} \psi_1 + h & \text{if } y = 0; \\ \alpha \cdot \rho_1 & \text{if } y \in [-\bar{h}(h), \bar{h}(h)] \setminus \{0\}; \\ \alpha \cdot \rho_1 \cdot \frac{\lambda_1 \overline{F}(\rho_1)}{\mu_1(x)} & \text{if } y \in K(h); \\ \psi_1 & \text{if } y \in [-h,h]^c. \end{cases}$$

Let $a(x)$ be defined by

$$a(x) \triangleq \begin{cases} \psi_1 + h - |x| & \text{if } x \in [-h,h], \\ \psi_1 & \text{if } x \in [-h,h]^c. \end{cases}$$

We argue that $V(\cdot|\tilde{p},\tilde{\tau}) \equiv a(\cdot)$. Fix $x \in \mathcal{C}$, it is not hard to verify that

$$\Gamma(y \in \mathcal{C} : \tilde{U}(y) - |y - x| > a(x)) = 0,$$

and, thus, $a(x) \geq V(x|\,\tilde{p},\tilde{\tau})$. Suppose that $x \in [-h,h]$ and $a(x) > V(x|\,\tilde{p},\tilde{\tau})$ then, because $\Gamma(\{0\}) > 0$, we have that

$$\psi_1 + h - |x| = a(x) > V(x|\,\tilde{p},\tilde{\tau}) \geq \Pi(x,0) = \psi_1 + h - |0 - x|,$$

a contradiction. Thus, for $x \in [-h,h]$ we have $a(x) = V(x|\tilde{p},\tilde{\tau})$. For any other x we can use a similar argument to conclude that $a(x) = V(x|\tilde{p},\tilde{\tau})$.

Now we are ready to verify the equilibrium condition. Observe that

$$\begin{aligned} \mathcal{E} &= \Big\{(x,y) \in \mathcal{C} \times \mathcal{C} : \Pi(x,y) = V(x|\tilde{p},\tilde{\tau})\Big)\Big\} \\ &= ([-h,h] \times \{0\}) \cup ([-h,h]^c \times [-h,h]^c \cap \mathcal{D}) \cup (K(h) \times K(h) \cap \mathcal{D}), \end{aligned}$$

then

$$\begin{aligned} \tilde{\tau}(\mathcal{E}) &= \mu(\pi_1(\mathcal{E} \cap \mathcal{D}) \cap [-h,h]^c) + \mu(\pi_1(\mathcal{E} \cap [-\bar{h}(h), \bar{h}(h)] \times \{0\})) \\ &\quad + G_1(\pi_1(\mathcal{E} \cap \mathcal{D}) \cap K(h)) + G_0(\pi_1(\mathcal{E} \cap K(h) \times \{0\})) \\ &= \mu([-h,h]^c) + \mu([-\bar{h}(h), \bar{h}(h)]) + G_1(K(h)) + G_0(K(h)) \\ &= \mu(\mathcal{C}). \end{aligned}$$

This proves that $\tilde{\tau}$ is an equilibrium. Next we need to show $(\tilde{p}, \tilde{\tau})$ satisfies Eq. (A.22). From Proposition 1.1 we have

$$\begin{aligned}
\gamma \mathbf{Rev}(\tilde{p}, \tilde{\tau}) &= \int_{\mathcal{C}} V(x) \cdot \frac{d\tilde{\tau}_2}{d\Gamma}(x) d\Gamma(x) \\
&= (\psi_1 + h) \cdot S_T(h) + 2\int_{\bar{h}(h)}^{h} (\psi_1 + h - x)\mu_1(x)\, dm(x) + \psi_1 \cdot \mu_1 \cdot 2(H-h) \\
&\geq h \cdot S_T(h) + \psi_1 \Big(S_T(h) + 2\int_{\bar{h}(h)}^{h} \mu_1(x) dx \Big) + \psi_1 \cdot \mu_1 \cdot 2(H-h) \\
&= h \cdot S_T(h) + \psi_1 \Big(2\bar{h}(h)\mu_1 + 2\int_{\bar{h}(h)}^{h} (\mu_1 - \mu_1(x)) dx + 2\int_{\bar{h}(h)}^{h} \mu_1(x) dx \Big) \\
&+ \psi_1 \cdot \mu_1 \cdot 2(H-h) \\
&= h \cdot S_T(h) + \psi_1 \cdot \mu_1 \cdot 2 \cdot H.
\end{aligned}$$

Since $h \cdot S_T(h) > 0$, Eq. (A.22) obtains. This proves that $X_l < 0 < X_r$ in any optimal solution.

The next step of the proof of Step 1 consists on arguing that given $V(0)$, $X_r = V(0) - \psi_1$ and $X_l = -(V(0) - \psi_1)$. Consider a feasible solution (p, τ) where $p(\cdot)$ equals ρ_1 everywhere but at the origin, and $X_l < 0 < X_r$. From Proposition 1.3 and the fact that $\mu(\{X_r\}) = 0$ we have that

$$\tau([X_r, H] \times [X_r, H]^c) \leq \mu(\{X_r\}) + \tau((X_r, H] \times [X_r, H]^c) = 0.$$

Then by Lemma A.7 we have that $V(x) \leq \psi_1$, $\Gamma - a.e.\ x$ in $[X_r, H]$. This, together with the continuity of $V(\cdot)$ imply that $V(x) \leq \psi_1$ for all $x \in [X_r, H]$.

Suppose first that $X_r < V(0) - \psi_1$ then

$$V(X_r |\, p, \tau) = V(0) - X_r > \psi_1,$$

but this violates the continuity of V to the right of X_r. Thus $X_r \geq V(0) - \psi_1$. On the other hand, suppose $X_r > V(0) - \psi_1$ then we must have that $\psi_1 > V(x|p, \tau) = V(0) - x$ for all $x \in (V(0) - \psi_1, X_r]$. Observe that

$$\mu([V(0) - \psi_1, X_r]) \geq \tau_2([V(0) - \psi_1, X_r]) = \int_{[V(0)-\psi_1, X_r]} s^{\tau}(x)\, d\Gamma(x). \tag{A.24}$$

Define the set

$$K \triangleq \{y \in [V(0) - \psi_1, X_r] : s^{\tau}(y) \leq \mu_1\},$$

it must be that $\Gamma(K) = 0$; otherwise, from the definition of $V(X_r |\, p, \tau)$ we have

$$\begin{aligned} V(0) - X_r = V(X_r) &\geq U(y, \rho_1, s^{\tau}(y)) - |y - X_r|, \quad \Gamma - a.e.\ y \text{ in } K \\ &\geq U(y, \rho_1, \mu_1) - |y - X_r|, \quad \Gamma - a.e.\ y \text{ in } K \\ &= \psi_1 - (X_r - y), \quad \Gamma - a.e.\ y \text{ in } K, \end{aligned}$$

and $\Gamma(K) > 0$ implies that $V(0) - y \geq \psi_1$ for some $y \in (V(0) - \psi_1, X_r]$. However, we know that $\psi_1 > V(0) - y$ for $y \in (V(0) - \psi_1, X_r]$ and, therefore, we must have $\Gamma(K) = 0$. Using this in Eq. (A.24) yields

$$\mu([V(0) - \psi_1, X_r]) > \mu_1 \cdot \Gamma([V(0) - \psi_1, X_r]) = \mu([V(0) - \psi_1, X_r]),$$

which is not possible. Hence, $X_r = V(0) - \psi_1$ and the same arguments applies to X_l, yielding $X_l = -(V(0) - \psi_1)$.

In order to conclude the proof for Step 1 we show that we can restrict attention to solutions (p, τ) such that $V(x|\, p, \tau)$ equals ψ_1 for all $x \in [X_l, X_r]^c$. In turn, this will show that $s^{\tau}(x)$ equals μ_1, $\Gamma - a.e.\ x$ in $[X_l, X_r]^c$. We base the proof of the latter statements in Lemma A.9 (which we state and prove after the proof of the present result), this lemma enables us to separate the city into two regions $[X_l, X_r]$ and $[X_l, X_r]^c$. For each region we can modify the prices and equilibria, and then paste them together to obtain a new solution that is an equilibrium for the entire city.

Consider a feasible solution (p, τ) such that $X_l < 0 < X_r$, $X_r = V(0) - \psi_1$ and $X_l = -X_r$. Since $\tau([X_l, X_r] \times [X_l, X_r]^c) = 0$ and $0 \notin [X_l, X_r]^c$, Lemma A.7 delivers

$$\frac{1}{\gamma} \cdot \mathbf{Rev}(p, \tau) \leq \frac{1}{\gamma} \cdot \mathbf{Rev}_{[X_l, X_r]}(p, \tau) + 2 \cdot \mu_1 \cdot \psi_1 \cdot (H - X_r). \tag{A.25}$$

We show that we can always modify (p, τ) so that the previous upper bound is achieve. Let $\mathcal{B} = [X_l, X_r]$, since $\tau(\mathcal{B} \times \mathcal{B}^c) = 0$ and $\tau(\mathcal{B}^c \times \mathcal{B}) = 0$, Lemma A.9 ensures that

$(p, \tau)|_{\mathcal{B}}$ is a price equilibrium pair in $\mathcal{B}$. Such equilibrium satisfies $V_{\mathcal{B}}(x) = \psi_1$ for $x \in \partial\mathcal{B}$.

Now, we choose prices $p^{\mathcal{B}^c}(x)$ equal to ρ_1 for all $x \in \mathcal{B}^c$ and a flow $\tau^{\mathcal{B}^c}$ defines by for any measurable set $\mathcal{L}_1 \times \mathcal{L}_2 \subseteq \mathcal{B}^c \times \mathcal{B}^c$

$$\tau^{\mathcal{B}^c}(\mathcal{L}_1 \times \mathcal{L}_2) = \mu(\mathcal{L}_1 \cap \mathcal{L}_2).$$

Then, it is easy to verify (as we did in the pre-shock environment, see Proposition 1.6) that $(p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c})$ forms a price-equilibrium pair in $\mathcal{B}^c$. This solution satisfy that $V_{\mathcal{B}^c}(x) = \psi_1$ for $x \in \mathcal{B}^c$, and that $s^{\tau^{\mathcal{B}^c}}(x)$ equals μ_1, $\Gamma - a.e.\ x$ in $\mathcal{B}^c$.

Lemma A.9 enables us to paste the solutions $(p, \tau)|_{\mathcal{B}}$ and $(p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c})$, and generate a new solution in the entire city. Such solution preserve the prices and flows in both $\mathcal{B}$ and $\mathcal{B}^c$ and, therefore, the upper bound in Eq. (A.25) is achieved. In conclusion, we can restrict attention to solutions (p, τ) such that $V(x|\, p, \tau)$ equals ψ_1 for all $x \in [X_l, X_r]^c$, and that $s^{\tau}(x)$ equals μ_1, $\Gamma - a.e.\ x$ in $[X_l, X_r]^c$.

Step 2: We characterize $s^{\tau}(\cdot)$ (this completely characterizes τ_2). Let

$$X_r^0 = (V(0) - \alpha \cdot \rho_1))^+ \quad \text{and} \quad X_l^0 = -X_r^0,$$

and

$$\mu_1(y) \triangleq \alpha \cdot \rho_1 \cdot \frac{\lambda_1 \cdot \overline{F}(\rho_1)}{V(0) - |y|}, \quad S_T = 2 \cdot \mu_1 \cdot X_r^0 + 2 \int_{X_r^0}^{X_r} (\mu_1 - \mu_1(x))dx.$$

In this step we show that $(\Gamma - a.e)$

$$s^{\tau}(y) = \begin{cases} S_T & \text{if } y = 0 \\ 0 & \text{if } y \in [X_l^0, X_r^0] \setminus \{0\} \\ \mu_1(y) & \text{if } y \in [X_l, X_r] \setminus [X_l^0, X_r^0] \\ \mu_1 & \text{if } y \in [X_l, X_r]^c. \end{cases}$$

Proof of Step 2: Note that at the end of the previous step we showed the result for $y \in [X_l, X_r]^c$. So first we show

$$s^{\tau}(y) = 0, \quad \Gamma - a.e.\ x \text{ in } [X_l^0, X_r^0] \setminus \{0\}.$$

Define the set $K_1 \triangleq \{y \in [X_l^0, X_r^0] \setminus \{0\} : s^\tau(y) > 0\}$. We argue that $\Gamma(K_1) = 0$. If this is not the case then $\Gamma(K_1) > 0$ and, therefore,

$$\tau_2(K_1) = \int_{K_1} s^\tau(x)\, d\Gamma(x) > 0.$$

Then Lemma A.2 ensures that

$$U\Big(x, \rho_1, s^\tau(x)\Big) \;=\; V(x|\, p, \tau) \qquad \tau_2 - a.e.\ x \in K_1, \tag{A.26}$$

but for $x \in K_1 \subseteq [X_l^0, X_r^0] \setminus \{0\}$ we have $V(x|\, p, \tau) = V(0) - |x|$ and $V(0) - |x| \geq \alpha \cdot \rho_1$. Then Eq. (A.26) implies the existence of $x \in (X_l^0, X_r^0) \setminus \{0\}$ such that $\alpha \cdot \rho_1 < U\Big(x, \rho_1, s^\tau(x)\Big) \leq \alpha \cdot \rho_1$, yielding a contradiction. Next we show that

$$s^\tau(y) = \mu_1(y), \quad \Gamma - a.e.\ y \text{ in } [X_l, X_r] \setminus [X_l^0, X_r^0].$$

By Lemma A.2 we have that

$$U\Big(x, \rho_1, s^\tau(x)\Big) = V(x) = V(0) - |x|, \quad \Gamma - a.e.\ x \text{ in } [X_l, X_r] \setminus [X_l^0, X_r^0], \tag{A.27}$$

but for any $x \in [X_l, X_r] \setminus [X_l^0, X_r^0]$ the definition of X_l^0 and X_r^0 imply that $V(0) - |x| < \alpha \cdot \rho_1$. Thus Eq. (A.27) and the definition of $U\Big(x, \rho_1, s^\tau(x)\Big)$ deliver

$$\lambda_1 \cdot \overline{F}(\rho_1)/s^\tau(x) < 1, \quad \Gamma - a.e.\ x \text{ in } [X_l, X_r] \setminus [X_l^0, X_r^0].$$

Using the again Eq. (A.27) and the definition of $U\Big(x, \rho_1, s^\tau(x)\Big)$ we conclude that

$$s^\tau(x) = \alpha \cdot \rho_1 \cdot \frac{\overline{F}(\rho_1)}{V(0) - |x|}, \quad \Gamma - a.e.\ x \text{ in } [X_l, X_r] \setminus [X_l^0, X_r^0],$$

as needed. Next we compute $s^\tau(0)$,

$$\begin{aligned} s^\tau(0) \cdot \Gamma(\{0\}) = \int_{\{0\}} s^\tau(x)\, d\Gamma &= \tau_2(\{0\}) \\ &= \tau(\mathcal{C} \times \{0\}) \\ &= \tau([X_l, X_r] \times \{0\}) \\ &= \underbrace{\tau([X_l^0, X_r^0] \times \{0\})}_{(1)} \\ &+ \underbrace{\tau([X_l, X_r] \setminus [X_l^0, X_r^0] \times \{0\})}_{(2)}, \end{aligned}$$

for (1) we have

$$\begin{aligned}\tau([X_l^0, X_r^0] \times \{0\}) &= \mu([X_l^0, X_r^0]) - \tau([X_l^0, X_r^0] \times \mathcal{C} \setminus \{0\}) \\ &= 2\mu_1 \cdot X_r^0 - \tau([X_l^0, X_r^0] \times [X_l^0, X_r^0] \setminus \{0\}) \\ &\overset{(a)}{=} 2\mu_1 \cdot X_r^0,\end{aligned}$$

in (a) we use $s^\tau(x) = 0,\ \Gamma - a.e.\ x\ \text{ in }\ [X_l^0, X_r^0] \setminus \{0\}$. For (2) we have

$$\begin{aligned}\tau([X_l, X_r] \setminus [X_l^0, X_r^0] \times \{0\}) &= \mu([X_l, X_r] \setminus [X_l^0, X_r^0]) \\ &\quad - \tau([X_l, X_r] \setminus [X_l^0, X_r^0] \times [X_l, X_r] \setminus \{0\}) \\ &= 2\mu_1 \cdot (X_r - X_r^0) \\ &\quad - \tau([X_l, X_r] \setminus [X_l^0, X_r^0] \times [X_l^0, X_r^0] \setminus \{0\}) \\ &\quad - \tau([X_l, X_r] \setminus [X_l^0, X_r^0] \times [X_l, X_r] \setminus [X_l^0, X_r^0]) \\ &= 2\mu_1 \cdot (X_r - X_r^0) - 0 - \tau_2([X_l, X_r] \setminus [X_l^0, X_r^0]) \\ &= 2\mu_1 \cdot (X_r - X_r^0) - \int_{[X_l, X_r] \setminus [X_l^0, X_r^0]} \mu_1(x)\, d\Gamma,\end{aligned}$$

from this we conclude that

$$s^\tau(0) = 2 \cdot \mu_1 \cdot X_r^0 + 2 \int_{X_r^0}^{X_r} (\mu_1 - \mu_1(x))\, dx.$$

Step 3: Now we can provide a full solution for the optimization problem. Recall that we are only optimizing over $p(0)$ or, equivalently, over $V(0)$. By our congestion bound (see Proposition 1.5), any solution has to satisfy $V(0|\, p, \tau) \leq \psi_0(s^\tau(0))$. Moreover, Step 2 characterizes the supply-demand ratio at every location as a function of $V(0)$. Thus, the following formulation is a natural relaxation for the platform's problem

$$\begin{aligned}\max_{V(0)} \quad & V(0) \cdot S_T + 2 \cdot \psi_1 \cdot \mu_1 \cdot (H - X_r^0) && (\mathcal{P}_{loc-reac}) \\ \text{s.t} \quad & X_r^0 = (V(0) - \alpha \cdot \rho_1))^+, \quad X_r = V(0) - \psi_1 \\ & S_T = 2X_r^0\mu_1 + 2\int_{X_r^0}^{X_r} (\mu_1 - \mu_1(x))dx, \qquad \psi_1 < V(0) \leq \psi_0(S_T).\end{aligned}$$

We show that the optimal $V^\star(0)$ in $(\mathcal{P}_{loc-reac})$ is the unique solution to

$$V^\star(0) = \psi_0(S_T(V^\star(0))).$$

The optimal solution to the platform's problem set price at the origin $p^\star(0) = \rho_0^{loc}(S_T(V^\star(0)))$ such that $p^\star(0) \geq \rho_1$, and flows for any measurable set $\mathcal{B} \subset \mathcal{C} \times \mathcal{C}$ given by

$$\begin{aligned}\tau(\mathcal{B}) &= \mu(\pi_1(\mathcal{B} \cap \mathcal{D}) \cap [X_l, X_r]^c) + \mu(\pi_1(\mathcal{B} \cap [X_l^0, X_r^0] \times \{0\})) \\ &\quad + G_1(\pi_1(\mathcal{B} \cap \mathcal{D}) \cap [X_l, X_r] \setminus [X_l^0, X_r^0]) + G_0(\pi_1(\mathcal{B} \cap [X_l, X_r] \setminus [X_l^0, X_r^0] \times \{0\})),\end{aligned}$$

where G_0, G_1 are measures defined for any measurable set $\mathcal{L} \subset [X_l, X_r] \setminus [X_l^0, X_r^0]$ by

$$G_0(\mathcal{L}) \triangleq \int_{\mathcal{L}} (\mu_1 - \mu_1(x))\, dm(x), \quad G_1(\mathcal{L}) \triangleq \int_{\mathcal{L}} \mu_1(x)\, dm(x).$$

Proof of Step 3: The proof consists of two parts. First, we show that $V^\star(0)$ as stated above is an optimal solution for $(\mathcal{P}_{loc-reac})$. To do this we prove that $S_T(V(0))$ is increasing for $V(0) > \psi_1$, with $S_T(\psi_1) = 0$. This implies that $\psi_0(S_T(V(0)))$ is decreasing and, therefore, it crosses with $V(0)$ at only one point. Then, we show the objective function increases with $V(0)$. These two facts imply the optimality of $V^\star(0)$. Second, we show that (p, τ) with $p(0) = p^\star(0)$ (and equal to ρ_1 for $x \neq 0$) and and τ as stated above, are a feasible price-equilibrium pair that achieve the same revenue than the optimal solution of $(\mathcal{P}_{loc-reac})$. Since this problem is a relaxation to our original optimization problem we have optimality.

We begin with the first part. Note that

$$S_T(V(0)) = 2\mu_1 \cdot (V(0) - \psi_1) + 2\psi_1 \cdot \mu_1 \cdot \log\Big(\frac{\psi_1}{V(0) - (V(0) - \alpha\rho_1)^+}\Big).$$

From this it follows that $S_T(\psi_1) = 0$. If $V(0) \geq \alpha\rho_1$ then $S_T(V(0))$ is clearly increasing. If $V(0) \in (\psi_1, \alpha\rho_1)$ then the derivative of $S_T(V(0))$ with respect to $V(0)$ equals

$$2\mu_1 - 2\psi_1 \cdot \mu_1 \cdot \frac{V(0)}{\psi_1} \cdot \frac{\psi_1}{V(0)^2} = 2\mu_1 - 2\psi_1 \cdot \mu_1 \cdot \frac{1}{V(0)},$$

which is nonnegative if and only if $V(0) \geq \psi_1$. Since this is in our domain, we conclude that $S_T(\cdot)$ is increasing in $(\psi_1, \alpha\rho_1)$ and, therefore, is increasing for all $V(0) > \psi_1$.

Next, we show the objective is increasing in $V(0)$, the objective function is

$$V(0) \cdot S_T(V(0)) + 2 \cdot \psi_1 \cdot \mu_1 \cdot (H - (V(0) - \alpha\rho_1)^+),$$

when $V(0) \geq \alpha \cdot \rho_1$, the objective becomes

$$2\mu_1 \cdot V(0) \cdot (V(0) - \psi_1) + 2\psi_1 \cdot \mu_1 \cdot V(0) \cdot \log\left(\frac{\psi_1}{\alpha\rho_1}\right) + 2 \cdot \psi_1 \cdot \mu_1 \cdot (H - V(0) + \alpha\rho_1).$$

Its derivative is non-negative if and only if $2\frac{V(0)}{\psi_1} \geq 2 + \log\left(\frac{\alpha\rho_1}{\psi_1}\right)$, but from $V(0) \geq \alpha \cdot \rho_1$ and that the logarithm is a concave function the latter inequality is always true. Similarly, for $V(0) \in (\psi_1, \alpha \cdot \rho_1)$ the objective's derivative is non-negative if and only if $2\frac{V(0)}{\psi_1} \geq 2 + \log\left(\frac{V(0)}{\psi_1}\right)$, which, since $V(0) > \psi_1$, is always true. Observe that in both cases the inequalities for the sign of the objective's derivative is strict except when $V(0) = \psi_1$. Thus, the objective is strictly increasing in the domain.

For the second part we need to show that (p, τ) with $p(0) = p^\star(0)$ (and equal to ρ_1 for $x \neq 0$) and and τ, implement the solution of $(\mathcal{P}_{loc-reac})$. To do this we first need to argue that this solution is feasible. It can be easily seen that this flow yields the exact same flows as in Step 2, only this time we replace $V^\star(0)$ in all the quantities that depend on $V(0)$. Given the value of s^τ and the fact that under $p^\star(0)$ we have $U(0, p(0), s^\tau(0)) = V(0|\, p, \tau) = V^\star(0)$, we can do the same as we did in Step 1(to show that $\tilde{\tau}$ is an equilibrium) and show that τ is an equilibrium. Since we have pinned the value of $V(0|\, p, \tau)$ (and thus the value of $V(|\, p, \tau)$ in the entire city) and the value of $s^\tau(\cdot)$, it is easy to see (using Proposition 1.1) that $\frac{1}{\gamma} \cdot \mathbf{Rev}(p, \tau)$ coincides with the optimal value of $(\mathcal{P}_{loc-reac})$. Therefore, (p, τ) is the optimal solution.

To conclude we argue that $p^\star(0) \geq \rho_1$. There are two cases. If $\mu_1 \leq \lambda_1 \cdot \overline{F}(\rho_1)$ then ψ_1 equals $\alpha \cdot \rho_1$. Since $V^\star(0) > \psi$ and $V^\star(0) = \psi_0(S_T(V^\star(0)))$ we have have that

$$\alpha \cdot \rho_1 = \psi_1 < V^\star(0) = \psi_0(S_T(V^\star(0))) \leq \alpha \cdot \rho_0^{loc}(S_T(V^\star(0))) = \alpha \cdot p^\star(0),$$

that is, $\rho_1 < p^\star(0)$. The second case is $\mu_1 > \lambda_1 \cdot \overline{F}(\rho_1)$. Here ρ_1 equals ρ^u and, since $\rho_0^{loc}(S_T(V^\star(0)))$ equals $\max\{\rho_0^{bal}, \rho^u\}$, we have that $\rho_1 \leq p^\star(0)$.

□

Lemma A.8 *Let (p, τ) be a feasible price-equilibrium pair for either the local price response environment (Section 1.6.2) or the global price response environment (Section 1.6.3). If either $\{x \in (0, H] : 0 \in \mathcal{IR}(x|p, \tau)\} = \emptyset$ or $\{x \in [-H, 0) : 0 \in \mathcal{IR}(x|p, \tau)\} = \emptyset$, then the platform's objective satisfies*

$$\gamma \cdot \mathbf{Rev}(p, \tau) \leq \psi_1 \cdot \mu_1 \cdot 2 \cdot H.$$

Proof of Lemma A.8. WLOG let us just assume that $\{x \in (0, H] : 0 \in \mathcal{IR}(x|p, \tau)\} = \emptyset$. That is, for all $x \in (0, H]$ we have $0 \notin \mathcal{IR}(x|p, \tau)$. In turn, this implies that $\tau((0, H] \times [-H, 0]) = 0$ and, therefore, by Lemma A.7 we conclude that

$$V(x|p, \tau) \leq \psi_1 \qquad \Gamma - a.e. \text{ in } \quad (0, H],$$

which, from the continuity of $V(\cdot|p, \tau)$, implies that $V(x|p, \tau) \leq \psi_1$ for all $x \in [0, H]$. Now, we show that the same bound holds for $x \in [-H, 0)$. If $\tau([-H, 0) \times \mathcal{B}) = 0$ for any $\mathcal{B} \subset [0, H]$, we can use Lemma A.7 to obtain the upper bound. On the other hand, if there exists $\mathcal{B} \subset [0, H]$ such that $\tau([-H, 0) \times \mathcal{B}) > 0$ then by Lemma A.5 we know there exists a pair $(x, y) \in [-H, 0) \times \mathcal{B}$ for which $y \in \mathcal{IR}(x|\, p, \tau)$. Thus, we can define

$$\underline{x} = \inf\{z \in [-H, 0) : y \in \mathcal{IR}(z|\, p, \tau)\},$$

and by Proposition 1.3, $y \in \mathcal{IR}(\underline{x}|\, p, \tau)$. Also, from Lemma 1.3 we have

$$V(z|p, \tau) = V(\underline{x}|p, \tau) + z - \underline{x}, \quad \forall\, z \in [\underline{x}, y].$$

This implies $V(z|p, \tau) \leq V(y|p, \tau)$ for all $z \in [\underline{x}, y]$, and because $y \in \mathcal{B} \subset [0, H]$ we have $V(y|p, \tau) \leq \psi_1$, yielding $V(z|p, \tau) \leq \psi_1$ for all $z \in [\underline{x}, y]$. Furthermore, from Lemma A.5 and the definition of $\underline{x}$ we can conclude that $\tau([-H, \underline{x}] \times (\underline{x}, H]) = 0$

which together with Lemma A.7 and the continuity of V imply that $V(x|p,\tau) \leq \psi_1$ for all $x \in [-H, \underline{x}]$. Completing the argument for the upper bound.

In order to bound the revenue, simply note that

$$\frac{1}{\gamma} \cdot \mathbf{Rev}(p,\tau) = \int_{\mathcal{C}} V(x) s^\tau(x)\, d\Gamma(x) \leq \psi_1 \cdot \int_{\mathcal{C}} s^\tau(x)\, d\Gamma(x) = \psi_1 \cdot \mu_1 \cdot 2 \cdot H.$$

□

Lemma A.9 *(Equilibria Separation and Pasting) Consider a set $\mathcal{B} \subset \mathcal{C}$ such that both $\mathcal{B}$ and $\mathcal{B}^c$ are intervals or union of intervals with $\Gamma(\partial\mathcal{B}) = 0$.*

1. *(Separation) Let (p,τ) be a price-equilibrium in $\mathcal{C}$, if $\tau(\mathcal{B}\times\mathcal{B}^c) = 0$ and $\tau(\mathcal{B}^c\times\mathcal{B}) = 0$ then $(p|_{\mathcal{B}}, \tau|_{\mathcal{B}\times\mathcal{B}})$ and $(p|_{\mathcal{B}^c}, \tau|_{\mathcal{B}^c\times\mathcal{B}^c})$ are price-equilibrium pairs in $\mathcal{B}$ and $\mathcal{B}^c$, respectively. Moreover, $V(\cdot|p|_{\mathcal{B}}, \tau|_{\mathcal{B}\times\mathcal{B}})$ equals $V(\cdot|p|_{\mathcal{B}^c}, \tau|_{\mathcal{B}^c\times\mathcal{B}^c})$ in $\partial\mathcal{B}$, $V(\cdot|p|_{\mathcal{B}}, \tau|_{\mathcal{B}\times\mathcal{B}})$ coincides with $V(\cdot|\, p,\tau)|_{\mathcal{B}}$ and the same holds for $\mathcal{B}^c$.*

2. *(Pasting) Suppose we have two price-equilibrium pairs $(p^{\mathcal{B}}, \tau^{\mathcal{B}})$ and $(p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c})$ in $\mathcal{B}$ and $\mathcal{B}^c$ such that $\tau^{\mathcal{B}} \in \mathcal{F}_{\mathcal{B}}(\mu|_{\mathcal{B}})$ and $\tau^{\mathcal{B}^c} \in \mathcal{F}_{\mathcal{B}^c}(\mu|_{\mathcal{B}^c})$, respectively. If $V(\cdot|\, p^{\mathcal{B}}, \tau^{\mathcal{B}})$ equals $V(\cdot|\, p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c})$ in $\partial\mathcal{B}$ then the flow τ defined by for any measurable set $\mathcal{L} \subseteq \mathcal{C}\times\mathcal{C}$*

$$\tau(\mathcal{L}) = \tau^{\mathcal{B}}(\mathcal{L}\cap\mathcal{B}\times\mathcal{B}) + \tau^{\mathcal{B}^c}(\mathcal{L}\cap\mathcal{B}^c\times\mathcal{B}^c),$$

belongs to $\mathcal{F}(\mu)$ and is an equilibrium in $\mathcal{C}$ for a price p equal to $p^{\mathcal{B}}$ in $\mathcal{B}$ and equal to $p^{\mathcal{B}^c}$ in $\mathcal{B}^c$. Moreover, $V(x|p,\tau) = V(x|\, p^{\mathcal{B}}, \tau^{\mathcal{B}})$ in $\mathcal{B}$ and $V(x|p,\tau) = V(x|\, p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c})$ in $\mathcal{B}^c$.

Proof of Lemma A.9. **Separation.** Suppose that $\tau(\mathcal{B}\times\mathcal{B}^c) = 0$ and $\tau(\mathcal{B}^c\times\mathcal{B}) = 0$. Let $\tau^{\mathcal{B}} = \tau|_{\mathcal{B}\times\mathcal{B}}$ and $p^{\mathcal{B}} = p|_{\mathcal{B}}$, we show that $(p^{\mathcal{B}}, \tau^{\mathcal{B}})$ is a price-equilibrium pair. The proof for $(p|_{\mathcal{B}^c}, \tau|_{\mathcal{B}^c\times\mathcal{B}^c})$ is analogous and, thus, omitted. We need to prove that $\tau^{\mathcal{B}} \in \mathcal{F}_{\mathcal{B}}(\mu^{\mathcal{B}})$, where $\mu^{\mathcal{B}}$ coincides with $\mu|_{\mathcal{B}}$, and that the set

$$\mathcal{E}|_{\mathcal{B}} \triangleq \left\{(x,y)\in\mathcal{B}\times\mathcal{B} : \Pi(x,y,p^{\mathcal{B}}(y), \frac{d\tau_2^{\mathcal{B}}}{d\Gamma|_{\mathcal{B}}}(y)) = \operatorname*{ess\,sup}^{\mathcal{B}} \Pi(x,\cdot,p^{\mathcal{B}}(\cdot), \frac{d\tau_2^{\mathcal{B}}}{d\Gamma|_{\mathcal{B}}}(\cdot)\Big)\right\},$$

satisfies $\tau_{\mathcal{B}}(\mathcal{E}|_{\mathcal{B}}) = \mu|_{\mathcal{B}}(\mathcal{B})$.

First we verify that $\tau^{\mathcal{B}} \in \mathcal{F}_{\mathcal{B}}(\mu^{\mathcal{B}})$. Since $\tau^{\mathcal{B}}$ is the restriction of τ to $\mathcal{B} \times \mathcal{B}$ it clearly belongs to $\mathcal{M}(\mathcal{B} \times \mathcal{B})$. Also, for any $\mathcal{L}_1$ measurable subset of $\mathcal{B}$ we have that $\tau_1^{\mathcal{B}}(\mathcal{L}_1)$ equals

$$\tau^{\mathcal{B}}(\mathcal{L}_1 \times \mathcal{B}) = \tau((\mathcal{L}_1 \times \mathcal{B}) \cap (\mathcal{B} \times \mathcal{B})) = \tau(\mathcal{L}_1 \times \mathcal{B}) = \tau(\mathcal{L}_1 \times \mathcal{C}) = \tau_1(\mathcal{L}_1) = \mu(\mathcal{L}_1).$$

Thus, $\tau_1^{\mathcal{B}} = \mu|_{\mathcal{B}}$. Now we need to prove that $\tau_2^{\mathcal{B}} \ll \Gamma|_{\mathcal{B}}$. Observe that for any $\mathcal{L}_2$ measurable subset of $\mathcal{B}$ we have that $\tau_2^{\mathcal{B}}(\mathcal{L}_2)$ equals

$$\tau_{\mathcal{B}}(\mathcal{B} \times \mathcal{L}_2) = \tau((\mathcal{B} \times \mathcal{L}_2) \cap (\mathcal{B} \times \mathcal{B})) = \tau(\mathcal{B} \times \mathcal{L}_2) = \tau(\mathcal{C} \times \mathcal{L}_2) = \tau_2(\mathcal{L}_2),$$

that is, $\tau_2^{\mathcal{B}} = \tau_2|_{\mathcal{B}}$. Therefore, since $\tau_2 \ll \Gamma$, we have that $\tau_2^{\mathcal{B}} \ll \Gamma|_{\mathcal{B}}$. In turn, $\tau^{\mathcal{B}} \in \mathcal{F}_{\mathcal{B}}$.

Now we show $\tau^{\mathcal{B}}(\mathcal{E}|_{\mathcal{B}}) = \mu|_{\mathcal{B}}(\mathcal{B})$. It suffices to prove that $\tau^{\mathcal{B}}(\mathcal{E}|_{\mathcal{B}}^c) = 0$ where the complement is taken with respect to $\mathcal{B} \times \mathcal{B}$, we do this by contradiction. Assume that $\tau^{\mathcal{B}}(\mathcal{E}|_{\mathcal{B}}^c) > 0$, this implies that $0 < \tau^{\mathcal{B}}(\mathcal{E}|_{\mathcal{B}}^c) = \tau(\mathcal{E}|_{\mathcal{B}}^c)$, and we must have that $\tau_2(\mathcal{B}) > 0$, indeed

$$0 < \tau(\mathcal{E}|_{\mathcal{B}}^c) \leq \tau(\mathcal{C} \times \mathcal{B}) = \tau_2(\mathcal{B}).$$

Next, observe that for any $\mathcal{L}_2$ measurable subset of $\mathcal{B}$

$$\tau_2^{\mathcal{B}}(\mathcal{L}_2) = \tau_2(\mathcal{L}_2) = \int_{\mathcal{L}_2} s^\tau(x)\, d\Gamma(x) = \int_{\mathcal{L}_2} s^\tau(x)\, d\Gamma|_{\mathcal{B}}(x),$$

therefore,

$$\frac{d\tau_2^{\mathcal{B}}}{d\Gamma|_{\mathcal{B}}}(x) \ = \ s^\tau(x), \qquad \Gamma - a.e.\ x \ \text{ in } \mathcal{B}. \tag{A.28}$$

This implies that

$$V(x|p^{\mathcal{B}}, \tau^{\mathcal{B}}) = \operatorname*{ess\,sup}_{\mathcal{B}} \Pi(x, \cdot, p^{\mathcal{B}}(\cdot), \frac{d\tau_2^{\mathcal{B}}}{d\Gamma|_{\mathcal{B}}}(\cdot)\Big) = \operatorname*{ess\,sup}_{\mathcal{B}} \Pi(x, \cdot, p(\cdot), \frac{d\tau_2}{d\Gamma}(\cdot)\Big) = V_{\mathcal{B}}(x|p, \tau). \tag{A.29}$$

Consider the set $\mathcal{G} \triangleq \{y \in \mathcal{B} : \frac{d\tau_2^{\mathcal{B}}}{d\Gamma|_{\mathcal{B}}}(y) \ = \ s^\tau(y)\}$. Then, by Eq. (A.28) we have

$$\tau(\mathcal{E}|_{\mathcal{B}}^c \cap (\mathcal{B} \times \mathcal{G}^c)) \leq \tau(\mathcal{C} \times \mathcal{G}^c) = \tau_2(\mathcal{G}^c) = 0,$$

where the complement is take with respect to $\mathcal{B}$. Therefore, $0 < \tau(\mathcal{E}|_{\mathcal{B}}^c) = \tau(\mathcal{E}|_{\mathcal{B}}^c \cap (\mathcal{B} \times \mathcal{G}))$ and we can conclude that

$$\tau\Big(\underbrace{\{(x,y) \in \mathcal{B} \times \mathcal{B} : \Pi(x,y,p(y),\frac{d\tau_2}{d\Gamma}(y)) \neq V_{\mathcal{B}}(x|p,\tau)\}}_{\triangleq R}\Big) > 0.$$

Define the sets R^- and R^+ by

$$\begin{aligned} R^- &= \{(x,y) \in \mathcal{B} \times \mathcal{B} : \Pi(x,y,p(y),\frac{d\tau_2}{d\Gamma}(y)) > V_{\mathcal{B}}(x|p,\tau)\} \\ R^+ &= \{(x,y) \in \mathcal{B} \times \mathcal{B} : \Pi(x,y,p(y),\frac{d\tau_2}{d\Gamma}(y)) < V_{\mathcal{B}}(x|p,\tau)\}, \end{aligned}$$

and note that $R = R^- \cup R^+$. To obtain a contradiction we argue that $\tau(R^- \cup R^+) = 0$. Consider first the set R^+, and note that $\tau(R^+) = \tau(R^+ \cap \mathcal{E})$. However, any $(x,y) \in R^+ \cap \mathcal{E}$ satisfies

$$\Pi(x,y,(p(y),\frac{d\tau_2}{d\Gamma}(y)) < V_{\mathcal{B}}(x|p,\tau) \text{ and } \Pi(x,y,(p(y),\frac{d\tau_2}{d\Gamma}(y)) = V(x|p,\tau),$$

but $V(x) \geq V_{\mathcal{B}}(x)$ implies that $R^+ \cap \mathcal{E} = \emptyset$ and, therefore, $\tau(R^+) = 0$.

Consider R^-. Define $A \triangleq \{y \in \mathcal{B} : U(y) = V_{\mathcal{B}}(y|p,\tau)\}$, then by Lemma A.2 we have $\tau(R^-) = \tau(R^- \cap (\mathcal{B} \times A))$. Take any $(x,y) \in R^- \cap (\mathcal{B} \times A)$ then $V_{\mathcal{B}}(y|p,\tau) - |y - x| > V_{\mathcal{B}}(x|p,\tau)$, which, because of the Lipchitz property (see Lemma 1.1), is not possible. Thus, $R^- \cap (\mathcal{B} \times A)) = \emptyset$ and we have that $\tau(R^-) = 0$. This proves that $\tau^{\mathcal{B}}$ is an equilibrium in $\mathcal{B}$.

Now we show that $V(x|p^{\mathcal{B}}, \tau^{\mathcal{B}}) = V(x|p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c})$ for all $x \in \partial\mathcal{B}$. From equation (A.29) we have

$$V(x|p^{\mathcal{B}}, \tau^{\mathcal{B}}) = V_{\mathcal{B}}(x|p,\tau) \quad \text{and} \quad V(x|p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c}) = V_{\mathcal{B}^c}(x|p,\tau),$$

so we just need to show $V_{\mathcal{B}}(x|p,\tau)$ equals $V_{\mathcal{B}^c}(x|p,\tau)$ for all $x \in \partial B$. We first show that $V_{\mathcal{B}}(x|p,\tau) = V(x|p,\tau)$ for all $x \in \mathcal{B}$. Let $x \in \mathcal{B}$, since $\mathcal{B}$ is an interval or a union of intervals we must have $\mu(B(x,\frac{1}{n})\cap\mathcal{B})>0$ for all $n\in\mathbb{N}$. In turn, this implies

$$\begin{aligned}0&<\tau(B(x,\frac{1}{n})\cap\mathcal{B}\times\mathcal{B})\\&=\tau(B(x,\frac{1}{n})\cap\mathcal{B}\times\mathcal{B}^{\circ})+\tau(B(x,\frac{1}{n})\cap\mathcal{B}\times\partial\mathcal{B})\\&=\tau(B(x,\frac{1}{n})\cap\mathcal{B}\times\mathcal{B}^{\circ}),\end{aligned}$$

where the third line comes from $\tau_2\ll\Gamma$ and $\Gamma(\partial\mathcal{B})=0$. Thus, from Lemma A.5 there exists $(z_n,y_n)\in B(x,\frac{1}{n})\cap\mathcal{B}\times\mathcal{B}^{\circ}$ such that $y_n\in\mathcal{IR}(z_n|\,p,\tau)$. Then,

$$\forall n\in\mathbb{N},\ \exists\,\delta(n)>0\ \text{ such that }\forall\delta\leq\delta(n)\quad\frac{1}{n}+V_{B(y_n,\delta)}(z_n)\geq V(z_n).\tag{A.30}$$

Note that since $y_n\in\mathcal{B}^{\circ}$ we can always find δ_0 such that $B(y_n,\delta)\subseteq\mathcal{B}$ for all $\delta\leq\delta_0$. So we can consider $\delta\leq\min\{\delta_0,\delta(n)\}$ in Eq. (A.30). Using that $z_n\in B(x,\frac{1}{n})$ and the Lipschitz property (see Lemma 1.1) we have

$$V_{B(y_n,\delta)}(z_n)-V_{B(y_n,\delta)}(x)\leq\frac{1}{n}\quad\text{and}\quad V(z_n)-V(x)\geq-\frac{1}{n},$$

plugging this into Eq. (A.30) yields

$$\forall n\in\mathbb{N},\ \exists\,\delta(n)>0\ \text{ such that }\forall\delta\leq\min\{\delta_0,\delta(n)\}\quad\frac{3}{n}+V_{B(y_n,\delta)}(x)\geq V(x).$$

Since $B(y_n,\delta)\subseteq\mathcal{B}$ we have $V_{\mathcal{B}}(x)\geq V_{B(y_n,\delta)}(x)$ thus the former expression implies that $V_{\mathcal{B}}(x)\geq V(x)$. But we always have that $V(x)\geq V_{\mathcal{B}}(x)$ and, therefore, $V(x)=V_{\mathcal{B}}(x)$. The same argument shows that $V(x)=V_{\mathcal{B}^c}(x)$ for all $x\in\mathcal{B}^c$.

To conclude we need to prove that $V_{\mathcal{B}}(x|p,\tau)$ equals $V_{\mathcal{B}^c}(x|p,\tau)$ for all $x\in\partial B$. Consider $x\in\partial\mathcal{B}$. Let $\{x_n\}_{n\in\mathbb{N}}\subset\mathcal{B}$ be a sequence converging to x. Then the continuity of $V_{\mathcal{B}}$ implies $V_{\mathcal{B}}(x_n)\to V_{\mathcal{B}}(x)$. At the same time, since $x_n\in\mathcal{B}$ we have $V_{\mathcal{B}}(x_n)=V(x_n)$ and by continuity $V(x_n)\to V(x)$. Then $V_{\mathcal{B}}(x)=V(x)$ and the same is true for $\mathcal{B}^c$, which implies $V_{\mathcal{B}}(x|p,\tau)=V_{\mathcal{B}^c}(x|p,\tau)$ for all $x\in\partial B$.

Pasting. First we check that $\tau \in \mathcal{F}(\mu)$. Let $\mathcal{L}_1$ be any measurable subset of $\mathcal{C}$ we have that

$$\begin{aligned}
\tau_1(\mathcal{L}_1) &= \tau(\mathcal{L}_1 \times \mathcal{C}) \\
&= \tau^{\mathcal{B}}((\mathcal{L}_1 \times \mathcal{C}) \cap (\mathcal{B} \times \mathcal{B})) + \tau^{\mathcal{B}^c}((\mathcal{L}_1 \times \mathcal{C}) \cap (\mathcal{B}^c \times \mathcal{B}^c)) \\
&= \tau^{\mathcal{B}}((\mathcal{L}_1 \cap \mathcal{B}) \times \mathcal{B}) + \tau^{\mathcal{B}^c}((\mathcal{L}_1 \cap \mathcal{B}^c) \times \mathcal{B}^c) \\
&= \mu|_{\mathcal{B}}(\mathcal{L}_1 \cap \mathcal{B}) + \mu|_{\mathcal{B}^c}(\mathcal{L}_1 \cap \mathcal{B}^c) \\
&= \mu(\mathcal{L}_1).
\end{aligned}$$

Also, if $\Gamma(\mathcal{L}_1) = 0$ then $\Gamma|_{\mathcal{B}}(\mathcal{L}_1) = \Gamma|_{\mathcal{B}^c}(\mathcal{L}_1) = 0$. Therefore, $\tau_2^{\mathcal{B}}(\mathcal{L}_1) = \tau_2^{\mathcal{B}^c}(\mathcal{L}_1) = 0$, which in turn implies $\tau_2 \ll \Gamma$. Hence $\tau \in \mathcal{F}(\mu)$.

Now we show the set

$$\mathcal{E} \triangleq \Big\{(x, y) \in \mathcal{C} \times \mathcal{C} : \Pi(x, y, p(y), s^{\tau}(y)) = \operatorname*{ess\,sup}_{\mathcal{C}} \Pi(x, \cdot, p(\cdot), s^{\tau}(\cdot)\Big)\Big\},$$

satisfies $\tau(\mathcal{E}) = \mu(\mathcal{C})$. Note that

$$\mathcal{E} \cap \mathcal{B} \times \mathcal{B} = \Big\{(x, y) \in \mathcal{B} \times \mathcal{B} : \Pi(x, y, p(y), s^{\tau}(y)) = V(x|p, \tau)\Big\}.$$

It is enough to prove that $\tau^{\mathcal{B}}(\mathcal{E} \cap \mathcal{B} \times \mathcal{B}) = \mu(\mathcal{B})$. As we did in the first part of the proof (see Eq. (A.28)) we can show that

$$\frac{d\tau_2^{\mathcal{B}}}{d\Gamma|_{\mathcal{B}}}(x) \;=\; s^{\tau}(x), \qquad \Gamma - a.e.\ x \ \text{in} \ \mathcal{B},$$

so if we prove that $V(\cdot|p, \tau)|_{\mathcal{B}} \equiv V(\cdot|p^{\mathcal{B}}, \tau^{\mathcal{B}})$ we will be done (the proof for $\mathcal{B}^c$ is analogous). Fix $x \in \mathcal{B}$, as in Eq. (A.29) we have

$$V(x|p^{\mathcal{B}}, \tau^{\mathcal{B}}) = \operatorname*{ess\,sup}_{\mathcal{B}} \Pi(x, \cdot, p^{\mathcal{B}}(\cdot), \frac{d\tau_2^{\mathcal{B}}}{d\Gamma|_{\mathcal{B}}}(\cdot)\Big) = \operatorname*{ess\,sup}_{\mathcal{B}} \Pi(x, \cdot, p(\cdot), \frac{d\tau_2}{d\Gamma}(\cdot)\Big) = V_{\mathcal{B}}(x|p, \tau).$$

So we just need to verify that $V(x|p, \tau) = V_{\mathcal{B}}(x|p, \tau)$. We show that $V(x|p, \tau) \leq V_{\mathcal{B}}(x|p, \tau)$, the other inequality always holds. Let $I(x)$ be the interval in $\mathcal{B}$ to which x belongs to. Let $y_L = \inf I(x)$ and $y_U = \sup I(x)$, note that y_L and y_U do not necessarily belong to $\mathcal{B}$ but they do belong to $\partial\mathcal{B}$. By assumption $V(y|\, p^{\mathcal{B}}, \tau_{\mathcal{B}}) = V(y|\, p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c})$ for $y \in \{y_L, y_U\}$, in turn this implies that $V_{\mathcal{B}}(y|p,\tau)$ equals $V_{\mathcal{B}^c}(y|p,\tau)$ for $y \in \{y_L, y_U\}$. Consider the sets $\mathcal{B}^c_L = [H, y_L] \cap \mathcal{B}^c$ and $\mathcal{B}^c_U = [y_U, H] \cap \mathcal{B}^c$ then

$$\begin{aligned}
V_{\mathcal{B}}(x|p,\tau) &\overset{(a)}{\geq} V_{\mathcal{B}}(y_U|p,\tau) - |x - y_U| \\
&= V_{\mathcal{B}}(y_U|p,\tau) - (y_U - x) \\
&\overset{(b)}{\geq} U(w, s^\tau(w)) - |y_U - w| - (y_U - x), \quad \Gamma - a.e.\ w \text{ in } \mathcal{B}^c_U \\
&\overset{(c)}{\geq} U(w, s^\tau(w)) - (w - y_U) - (y_U - x), \quad \Gamma - a.e.\ w \text{ in } \mathcal{B}^c_U \\
&\overset{(d)}{\geq} U(w, s^\tau(w)) - |w - x|, \quad \Gamma - a.e.\ w \text{ in } \mathcal{B}^c_U,
\end{aligned}$$

where (a) follows from the Lipschitz property (see Lemma 1.1), and (b) from the definition of $V_{\mathcal{B}}(y_U|p,\tau)$ and $\Gamma(\mathcal{B}^c_U) > 0$; $(c), (d)$ hold because for $w \in \mathcal{B}^c_U$ we have $x \leq y_U \leq w$. Similarly,

$$\begin{aligned}
V_{\mathcal{B}}(x|p,\tau) &\geq V_{\mathcal{B}}(y_L|p,\tau) - |x - y_L| \\
&= V_{\mathcal{B}}(y_L|p,\tau) - (x - y_L) \\
&\geq U(w, s^\tau(w)) - |y_L - w| - (x - y_L), \quad \Gamma - a.e.\ w \text{ in } \mathcal{B}^c_L \\
&= U(w, s^\tau(w)) - (y_L - w) - (x - y_L), \quad \Gamma - a.e.\ w \text{ in } \mathcal{B}^c_L \\
&= U(w, s^\tau(w)) - |w - x|, \quad \Gamma - a.e.\ w \text{ in } \mathcal{B}^c_L.
\end{aligned}$$

Since $\mathcal{B}^c_L \cup \mathcal{B}^c_U = \mathcal{B}^c$ this implies that $V_{\mathcal{B}}(x|p,\tau) \geq V(x|p,\tau)$. This concludes the proof.

□

A.3.3 Proofs for Section 1.6.3

Proof of Lemma 1.5. Let (p,τ) be a feasible solution. We show that at any optimal solution we must have $X_l < 0 < X_r$, in turn this implies that 0 is a sink location. By Lemma A.8 we have that if either of the sets $\{x \in (0, H] : 0 \in \mathcal{IR}(x|p,\tau)\}$

or $\{x \in [-H, 0) : 0 \in \mathcal{IR}(x|p,\tau)\}$ is empty then the revenue the platform makes satisfies $\frac{1}{\gamma} \cdot \mathbf{Rev}(p,\tau) \leq \psi_1 \cdot \mu_1 \cdot 2 \cdot H$. However, the solution (p,τ) given in Proposition 1.7 has both sets non-empty because $0 \in \mathcal{IR}(X_r|\, p,\tau)$ and $0 \in \mathcal{IR}(-X_r|\, p,\tau)$ with $X_r > 0$. Furthermore, $\mathbf{Rev}(p,\tau)$ is strictly large than the revenue of the pre-demand shock environment or, equivalently, strictly larger than $\psi_1 \cdot \mu_1 \cdot 2 \cdot H$. This implies that any optimal solution must satisfy $\{x \in (0, H] : 0 \in \mathcal{IR}(x|p,\tau)\} \neq \emptyset$ and $\{x \in [-H, 0) : 0 \in \mathcal{IR}(x|p,\tau)\} \neq \emptyset$ and, therefore, $X_l < 0 < X_r$.

□

Proof of Lemma 1.6. If $X_r = H$ there is nothing to prove, so let's assume $X_r < H$. Fix $x \in [X_r, H]$. From the Lipschitz property (see Lemma 1.1) we have that $V(x|\, p,\tau) \leq V(X_r|\, p,\tau) + (x - X_r)$. Moreover, Proposition 1.3 ensures that $\tau([X_r, H] \times [X_r, H]^c) = 0$ and, hence, because $0 \notin [X_r, H]$ we can apply Lemma A.7 to deduce that

$$V(x|p,\tau) \;\leq\; \psi_1, \quad \Gamma - a.e.\ x \ \text{ in } [X_r, H]. \tag{A.31}$$

To show that the previous inequality holds everywhere, notice that if $V(x|\, p,\tau) > \psi_1$ the from the Lipschitz continuity property of $V(\cdot|\, p,\tau)$ we could find a subset of of $[X_r, H]$ with positive Γ measure (in this set Γ coincides with the Lebesgue measure) in which $V(\cdot|\, p,\tau)$ is strictly larger than ψ_1. This is not possible because it would contradict Eq. (A.31). Putting together both upper bounds yields the desire result.

□

Proof of Proposition 1.8. Let (p,τ) be optimal for problem $(\mathcal{P}_2)$ as in Lemma 1.5 so we have $0 < X_r$. Note that is $X_r = H$ then the result trivially holds, so let's assume $X_r < H$. Before we begin note that for any $x \geq X_r$, by Lemma 1.6 and the Lipschitz continuity property of $V(\cdot|\, p,\tau)$ (see Lemma 1.1), we must have $V(x) \leq \psi_1$.

We first prove the second statement of the proposition. Suppose $V(X_r) = \psi_1$ and define the set $R \triangleq \{x \in [X_r, H] : V(x) = \psi_1\}$. We show by contradiction that we cannot have $\tau_2(R^c) > 0$ (the complement is taken with respect to $[X_r, H]$). If

$\tau_2(R^c) > 0$, because ψ_1 is an upper bound from Proposition 1.1 we have the following

$$\begin{aligned}
\frac{1}{\gamma} \cdot \mathbf{Rev}_{[X_r,H]}(p,\tau) &= \int_{[X_r,H]} V(x)\, d\tau_2(x) \\
&= \int_R V(x)\, d\tau_2(x) + \int_{R^c} V(x)\, d\tau_2(x) \\
&< \int_R V(x)\, d\tau_2(x) + \int_{R^c} \psi_1\, d\tau_2(x) \\
&\leq \psi_1 \cdot \tau_2([X_r, H]) \\
&= \psi_1 \cdot \mu_1 \cdot (H - X_r),
\end{aligned}$$

where the last line comes Proposition 1.3. Thus, the quantity $\mathbf{Rev}_{[-H,X_r]}(p,\tau) + \gamma \cdot \psi_1 \cdot \mu_1 \cdot (H - X_r)$, strictly upper bounds the platform's objective. So if we are able to construct a solution such that attains the upper bound, we will contradict the optimality of (p,τ). Observe that Lemma A.9 enables us to separate the solution (p,τ) in $[-H, X_r]$ and $(X_r, H]$. The separated solution $(p^{[-H,X_r]}, \tau^{[-H,X_r]})$ (see Lemma A.9 for notation) in $[-H, X_r]$ has revenue equal to $\mathbf{Rev}_{[-H,X_r]}(p,\tau)$, and $V(X_r|\, p^{[-H,X_r]}, \tau^{[-H,X_r]})$ coincides with $V(X_r|\, p,\tau)$ which equals ψ_1. For $(X_r, H]$ consider prices $\tilde{p}(x) = \rho_1$ for all $x \in (X_r, H]^c$, and flows $\tilde{\tau}(\mathcal{L}) = \mu(\pi_1(\mathcal{L} \cap \mathcal{D}))$ for any measurable set $\mathcal{L} \subset (X_r, H] \times (X_r, H]$. The pair $(\tilde{p}, \tilde{\tau})$ is the same solution as in Proposition 1.6 with the sole difference that we have changed the city to be $(X_r, H]$ instead of $\mathcal{C}$. Therefore, $(\tilde{p}, \tilde{\tau})$ is a feasible price-equilibrium in $(X_r, H]$ with revenue equal to $\gamma \cdot \psi_1 \cdot \mu_1 \cdot (H - X_r)$, and such that $V(x|\, \tilde{p}, \tilde{\tau})$ equal to ψ_1 for all $x \in (X_r, H]$. Thus we can use Lemma A.9 to paste both solution and obtain an equilibrium in the entire city. This new equilibrium achieves the upper bound.

Suppose that $\tau_2(R^c) = 0$ and define the sets

$$L_+ \triangleq \{x : \mu_1 > s^\tau(x)\}, \quad L_0 \triangleq \{x : \mu_1 = s^\tau(x)\}, \quad L_- \triangleq \{x : \mu_1 < s^\tau(x)\}.$$

Then by Lemma 1.5 it holds that $\Gamma(R \cap L_-) = 0$. Moreover, if $\Gamma(R \cap L_+) > 0$ we have

$$\mu([X_r,H]) = \tau_2([X_r,H]) \overset{(a)}{=} \tau_2(R) = \int_{R\cap L_+} s^\tau(x)\, d\Gamma(x) + \int_{R\cap L_0} s^\tau(x)\, d\Gamma(x)$$
$$< \mu_1\Gamma(R) \le \mu([X_r,H]),$$

not possible, where (a) comes from Proposition 1.3. Thus $\Gamma(R\cap L_+) = 0$. This implies that $\Gamma(R\cap L_0) = \Gamma(R)$ and

$$\mu_1\Gamma([X_r,H]) = \mu([X_r,H]) = \int_{R\cap L_0} s^\tau(x)\, d\Gamma(x) = \mu_1\Gamma(R),$$

that is, $\Gamma(R) = \Gamma([X_r,H])$ or $\Gamma(R^c) = 0$. In turn, $\Gamma - a.e.$ $x \in [X_r,H]$ we have that $V(x)$ equals ψ_1. Since, $V(\cdot)$ is continuous and $\Gamma|_{[X_r,H]}$ has full support in $[X_r,H]$ which has non-empty interior we conclude that $V(x) = \psi_1$ for all $x \in [X_r,H]$.

For the reminder of the proof we assume $V(X_r) < \psi_1$. We show that if $V(\cdot)$ is not non-decreasing in $[X_r,H]$ then there is an strict objective improvement. In the proof we define several critical points in the interval $[X_r,H]$ which will help us to create a flow separated region (no flow leaves this region). Then we show the objective strict improvement in this region. In Figure A.1 we provide a graphical representation of the points just mentioned.

So assume that $V(x)$ is not non-decreasing in $[X_r,H]$, then there exists $\hat{x} > \hat{y} \ge X_r$ such that $V(\hat{x}) < V(\hat{y})$. Let,

$$\bar{y} \triangleq \sup\{z \in [\hat{y},\hat{x}] : V(z) = V(\hat{y})\},$$

note that since for $z = \hat{y}$, $V(z) = V(\hat{y})$ thus the set over which we take the supremum above is both bounded and non-empty. Hence, $\bar{y}$ is well defined and it corresponds to the last point z in $[\hat{y},\hat{x}]$ such that $V(z)$ equals $V(\hat{y})$. Moreover, because $V(\cdot)$ is continuous $\bar{y} < \hat{x}$, and for all $z \in (\bar{y},\hat{x}]$ we have $V(z) < V(\hat{y}) = V(\bar{y})$. Let

$$y_0 \triangleq \inf\{z \in [X_r,\bar{y}] : \exists x \in (\bar{y},H] \quad \text{such that} \quad z \in \mathcal{IR}(x)\},$$

if for all $z \in [X_r, \bar{y}]$ and for all $x \in (\bar{y}, H]$ we have $z \notin \mathcal{IR}(x)$, we let $y_0 = \bar{y}$. That is, y_0 is the smallest z in $[X_r, \bar{y}]$ to which some location in $(\bar{y}, H]$ is indifferent to travel to. Note that for all $z \in (y_0, \hat{x}]$ we have $V(z) < V(y_0)$. Also, the definition of y_0 and Lemma A.5 imply that $\tau([-H, y_0] \times (y_0, H]) = 0$ and $\tau((y_0, H] \times [-H, y_0]) = 0$. Let

$$y_1 \triangleq \inf\{z \in [\hat{x}, H] : V(z) > V(y_0)\},$$

that is, y_1 is the first value after $\hat{x}$ for which $V(\cdot)$ hits $V(y_0)$. Note that when well defined y_1 satisfies that $\tau([y_1, H] \times [-H, y_1]) = 0$. If this is not the case then since atoms do not have measure we would have $\tau((y_1, H] \times [-H, y_1)) > 0$ and, therefore, by Lemma A.5 we can find $(x, y) \in (y_1, H] \times [-H, y_1)$ such that $y \in \mathcal{IR}(x)$. Then Lemma 1.3 would contradict the minimality of y_1.

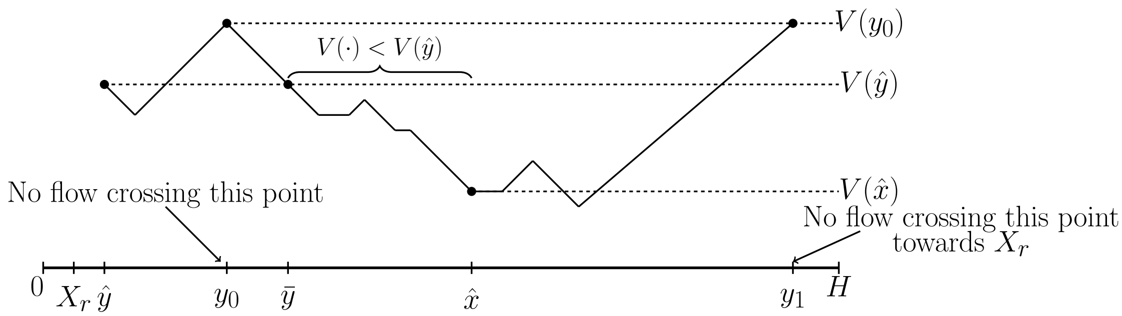


Figure A.1: Graphical representation of $\hat{y}, \hat{x}, \bar{y}, y_0$ and y_1.

There are two cases:

1. y_1 is not well defined: In this case we have that for all $z \in [\hat{x}, H]$, $V(z) \leq V(y_0)$. Recall that from our previous discussion we have that $V(z) < V(y_0)$ for all $z \in (y_0, \hat{x}]$. Also, Property 1 (which we prove at the end of the present proof) establishes that $\tau_2((y_0, \hat{x}]) > 0$. Using this observations we create a new solution $(\tilde{p}, \tilde{\tau})$ with revenue strictly larger than that of (p, τ).

   Let $\mathcal{B} = [-H, y_0]$ and note that we have both $\tau(\mathcal{B} \times \mathcal{B}^c) = 0$ and $\tau(\mathcal{B}^c \times \mathcal{B}) = 0$, so we can use the separation result in Lemma A.9. Hence $(p^{\mathcal{B}}, \tau^{\mathcal{B}})$ (see Lemma A.9 for notation) is a price-equilibrium pair in $\mathcal{B}$. Its revenue equals the revenue of (p, τ) in $\mathcal{B}$, and $V(y_0 | \, p^{\mathcal{B}}, \tau^{\mathcal{B}}) = V(y_0)$.

For $\mathcal{B}^c$ we choose flows $\tau^{\mathcal{B}^c}(\mathcal{L}) = \mu(\pi_1(\mathcal{L} \cap \mathcal{D}))$ for all $\mathcal{L} \subset \mathcal{B}^c \times \mathcal{B}^c$. That is all drivers stay at their initial location. It is not hard to see that $s^{\tau^{\mathcal{B}^c}}(x)$ equals μ_1, $\Gamma - a.e.\ x$ in $\mathcal{B}^c$. We choose prices $p^{\mathcal{B}^c}(x) = p_0$ for all $x \in \mathcal{B}^c$, where p_0 is such that

$$\alpha \cdot p_0 \cdot \min\{1, \frac{\lambda_1 \cdot \overline{F}(p_0)}{\mu_1}\} = V(y_0), \tag{A.32}$$

note that since $V(y_0) \leq \psi_1$, p_0 is well defined. That is, the solution $(p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c})$ is the same solution as in pre-demand shock environment but in smaller city, $\mathcal{B}^c$ and with a larger price across all locations. Using Proposition 1.1 it is not hard to see that the revenue associated with this solution is $\gamma \cdot V(y_0) \cdot \mu_1 \cdot (H - y_0)$.

By Lemma A.9, we can paste the two previous solutions to create a new solution $(\tilde{p}, \tilde{\tau})$ in entire city. This new solution yields a strict objective improvement. Indeed,

$$\begin{aligned}
\mathbf{Rev}_{[y_0,H]}(p, \tau) &= \int_{[y_0,H]} V(x)\, d\tau_2(x) \\
&= \int_{(y_0,\hat{x}]} V(x)\, d\tau_2(x) + \int_{(\hat{x},H]} V(x)\, d\tau_2(x) \\
&\overset{(a)}{<} V(y_0) \cdot \tau_2((y_0, \hat{x}]) + \int_{(\hat{x},H]} V(x)\, d\tau_2(x) \\
&\leq V(y_0) \cdot \tau_2((y_0, \hat{x}]) + V(y_0) \cdot \tau_2((\hat{x}, H]) \\
&\overset{(b)}{=} V(y_0) \cdot \mu([y_0, H]) \\
&= V(y_0) \cdot \mu_1 \cdot (H - y_0) \\
&= \mathbf{Rev}_{[y_0,H]}(\tilde{p}, \tilde{\tau}),
\end{aligned}$$

where (a) comes from $\tau_2((y_0, \hat{x}]) > 0$, (b) comes from the fact that under τ no flow leaves or enters $[y_0, H]$, and the last two lines from the definition of $(\tilde{p}, \tilde{\tau})$ restricted to $[y_0, H]$.

2. y_1 is well defined: In this case there exists $z \in [\hat{x}, H]$ such that $V(z) > V(y_0)$. Also, we must have $y_1 > \hat{x}$, and we already argued that $\tau([y_1, H] \times [-H, y_1]) = 0$. There are two more cases.

   a) $\forall y \in (y_0, y_1]$, $\forall x > y_1$, $x \notin \mathcal{IR}(y)$: This together with Lemma A.5 imply that $\tau([y_0, y_1] \times ([-H, y_0] \cup [y_1, H])) = 0$, and we also have $\tau(([-H, y_0] \cup [y_1, H]) \times [y_0, y_1]) = 0$. From this we can construct a new feasible solution $(\tilde{p}, \tilde{\tau})$ with revenue strictly larger than that of (p, τ).

      Let $\mathcal{B} = [-H, y_0) \cup (y_1, H]$ and note that we have both $\tau(\mathcal{B} \times \mathcal{B}^c) = 0$ and $\tau(\mathcal{B}^c \times \mathcal{B}) = 0$, so we can use the separation result in Lemma A.9. Thus $(p^{\mathcal{B}}, \tau^{\mathcal{B}})$ (see Lemma A.9 for notation) is a price-equilibrium pair in $\mathcal{B}$. Its revenue equals the revenue of (p, τ) in $\mathcal{B}$, and $V(y_0 |\, p^{\mathcal{B}}, \tau^{\mathcal{B}}) = V(y_0)$ and $V(y_1 |\, p^{\mathcal{B}}, \tau^{\mathcal{B}}) = V(y_0)$. For $\mathcal{B}^c$ we choose flows $\tau^{\mathcal{B}^c}(\mathcal{L}) = \mu(\pi_1(\mathcal{L} \cap \mathcal{D}))$ for all $\mathcal{L} \subset \mathcal{B}^c \times \mathcal{B}^c$. We choose prices $p^{\mathcal{B}^c}(x) = p_0$ for all $x \in \mathcal{B}^c$, where p_0 is as in Eq. (A.32). As we argued before this solution forms an price-equilibrium pair with revenue equal to $V(y_0) \cdot \mu_1 \cdot (y_1 - y_0)$.

      We can then paste both solutions (see Lemma A.9) to obtain a solution $(\tilde{p}, \tilde{\tau})$ in the entire city. As before, it yields a strict revenue improvement.

   b) $\exists\, y \in (y_0, y_1]$, $\exists\, x > y_1$ such that $x \in \mathcal{IR}(y)$: Then the following points are well defined

$$\overline{y}_1 \triangleq \sup\{x \in [y_1, H] : \exists\, y \in [y_0, y_1] \text{ such that } x \in \mathcal{IR}(y)\},$$
$$\underline{y}_1 \triangleq \inf\{y \in [y_0, y_1] : \exists\, x \in [y_1, H] \text{ such that } x \in \mathcal{IR}(y)\}.$$

      That is, $\overline{y}_1$ is largest point after y_1 for which some location in $[y_0, y_1]$ has drivers indifferent to travel to it. As for $\underline{y}_1$, it corresponds to the smallest point in $[y_0, y_1]$ that has drivers willing to travel to some location in $[y_1, H]$. Note that from the definition of $\overline{y}_1$ and Lemma A.5 we can deduce that there is no flow crossing $\overline{y}_1$ in any direction, that is, $\tau([-H, \overline{y}_1] \times [\overline{y}_1, H]) = 0$. Also, from

Property 2 (which we prove at the end of the present proof) for any $z \in [\underline{y}_1, \overline{y}_1]$, $\overline{y}_1 \in \mathcal{IR}(z)$. This together with Lemma 1.3 imply that for any $z \in [\underline{y}_1, \overline{y}_1]$, $V(z|\, p, \tau) = V(\overline{y}_1) - |\overline{y}_1 - z|$.

The idea is to again construct an strict objective improvement. First, define y^c to be such that $V(y_0) + (y^c - y_0) = V(\overline{y}_1)$, that is, $y^c = V(\overline{y}_1) - V(y_0) + y_0$. Next we argue that $y_c \in (y_0, \overline{y}_1)$. In fact, by the definition of $\overline{y}_1$ we must have $V(\overline{y}_1) > V(y_0)$ thus $y_c > y_0$. Also, if $y_c \geq \overline{y}_1$ then

$$V(y_0) + (y^c - y_0) \geq V(y_0) + (\overline{y}_1 - y_0) \Leftrightarrow V(\overline{y}_1) \geq V(y_0) + (\overline{y}_1 - y_0),$$

and since $V(\overline{y}_1) = V(y_1) + (\overline{y}_1 - y_1)$ we would have

$$V(y_1) + (\overline{y}_1 - y_1) \geq V(y_0) + (\overline{y}_1 - y_0) \Leftrightarrow V(y_1) - V(y_0) \geq y_1 - y_0,$$

which, since $y_1 > y_0$, implies that $V(y_1) > V(y_0)$, contradicting the definition of y_1. From this we can also infer that $y^c - y_0 = \overline{y}_1 - y_1$.

Second, let $h \triangleq \overline{y}_1 - y^c$ and for any set $\mathcal{L} \subseteq \mathcal{C} \times \mathcal{C}$ define the set

$$\mathcal{L}_h \triangleq \{(x+h, y+h) \in \mathcal{C} \times \mathcal{C} : (x,y) \in \mathcal{L}\}.$$

We now construct a new solution $(\tilde{p}, \tilde{\tau})$. Let $\mathcal{B} = [-H, y_0) \cup (\overline{y}_1, H]$, so that $\mathcal{B}^c = [y_0, \overline{y}_1]$. Following our previous scheme of proof we construct two price-equilibrium pairs one in $\mathcal{B}$ and another in $\mathcal{B}^c$, and then we paste them to create $(\tilde{p}, \tilde{\tau})$. As we did before we can use the separation result (see Lemma A.9) to obtain a solution $(p^{\mathcal{B}}, \tau^{\mathcal{B}})$ in $\mathcal{B}$ such that $V(y_0|p^{\mathcal{B}}, \tau^{\mathcal{B}}) = V(y_0)$ and $V(\overline{y}_1|p^{\mathcal{B}}, \tau^{\mathcal{B}}) = V(\overline{y}_1)$.

For $\mathcal{B}^c$ define the flow $\tau^{\mathcal{B}^c}$ for any $\mathcal{L} \subseteq \mathcal{B}^c \times \mathcal{B}^c$ by

$$\tau^{\mathcal{B}^c}(\mathcal{L}) = \tau\Big((\mathcal{L} \cap ([y_0, y^c] \times [y_0, \overline{y}_1]))_h\Big) + \mu(\pi_1(\mathcal{L} \cap ([y^c, \overline{y}_1] \times [y_0, \overline{y}_1]) \cap \mathcal{D})), \quad \text{(A.33)}$$

We next show that this flow belongs to $\mathcal{F}_{\mathcal{B}^c}(\mu|_{\mathcal{B}^c})$ and that it is an equilibrium for some prices $p^{\mathcal{B}^c}$ yet to be defined. Indeed, for any measurable subset K of

$\mathcal{B}^c$ we have

$$
\begin{aligned}
\tau_1^{\mathcal{B}^c}(K) &= \tau\Big(((K \times \mathcal{B}^c) \cap ([y_0, y^c] \times [y_0, \overline{y}_1]))_h\Big) \\
&\quad + \mu(\pi_1((K \times \mathcal{B}^c) \cap ([y^c, \overline{y}_1] \times [y_0, \overline{y}_1]) \cap \mathcal{D})) \\
&= \tau\Big(((K \cap [y_0, y^c]) \times [y_0, \overline{y}_1])_h\Big) + \mu(K \cap [y^c, \overline{y}_1]) \\
&= \tau\Big(((K+h) \cap [y_0 + h, y^c + h]) \times [y_0 + h, \overline{y}_1 + h]\Big) + \mu(K \cap [y^c, \overline{y}_1]) \\
&= \tau\Big(((K+h) \cap [y_1, \overline{y}_1]) \times [y_1, \overline{y}_1 + h]\Big) + \mu(K \cap [y^c, \overline{y}_1]) \\
&\overset{(a)}{=} \tau\Big(((K+h) \cap [y_1, \overline{y}_1]) \times \mathcal{C}\Big) + \mu(K \cap [y^c, \overline{y}_1]) \\
&= \mu((K+h) \cap [y_1, \overline{y}_1]) + \mu(K \cap [y^c, \overline{y}_1]) \\
&= \mu((K \cap [y_0, y^c]) + h) + \mu(K \cap [y^c, \overline{y}_1]) \\
&\overset{(b)}{=} \mu(K \cap [y_0, y_c]) + \mu(K \cap [y^c, \overline{y}_1]) \\
&= \mu(K),
\end{aligned}
$$

where (a) holds because by construction in $[y_1, \overline{y}_1]$ the flow there can be transported only inside the same set and, therefore, $\tau([y_1, \overline{y}_1] \times [y_1, \overline{y}_1 + h]^c)$ equals zero. Equality (b) comes from the fact that μ is invariant under translation (it is a multiple of the Lebesgue measure). Therefore, $\tau_1^{\mathcal{B}^c}$ coincides with $\mu|_{\mathcal{B}^c}$. Also, it is clear from the definition of $\tau^{\mathcal{B}^c}$ that $\tau_2^{\mathcal{B}^c} \ll \Gamma$. Hence, $\tau^{\mathcal{B}^c}$ belongs to $\mathcal{F}_{\mathcal{B}^c}(\mu|_{\mathcal{B}^c})$. Furthermore, Property 3 (which we prove at the end of the present proof) ensures that

$$
\frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(x) \leq \frac{d\tau_2}{d\Gamma}(x+h)\,\Gamma - a.e.\ x \text{ in } [y_0, y^c], \text{ and } \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(x) = \mu_1\,\Gamma - a.e.\ x \text{ in } [y^c, \overline{y}_1]. \tag{A.34}
$$

We choose the prices $p^{\mathcal{B}^c}$ as follows. In $[y^c, \overline{y}_1]$ we set constant prices equal to p_1 such that

$$
\alpha \cdot p_1 \cdot \min\{1, \frac{\lambda_1 \cdot \overline{F}(p_1)}{\mu_1}\} = V(\overline{y}_1),
$$

this price is well defined because $V(\overline{y}_1) \leq \psi_1$. For locations in $[y_0, y^c]$ consider the set

$$K \triangleq \left\{x \in [y_0, y^c] : \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(x) \leq \frac{d\tau_2}{d\Gamma}(x+h)\right\}, \tag{A.35}$$

note from Eq. A.34 we have $\Gamma(K^c) = 0$. We set prices for $x \in K$ to be such that

$$U\Big(x, p^{\mathcal{B}^c}(x), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(x)\Big) = U\Big(x+h, p(x+h), s^\tau(x+h)\Big), \tag{A.36}$$

such prices are well defined because the new Radon-Nikodym is smaller than the old one (shifted by h) in K. For $x \in K^c$ we set the prices equal to zero. Now we need to verify that this selection of prices and flows yields an equilibrium. That is, we need show that the set

$$\begin{aligned}\mathcal{E}_{\mathcal{B}^c} = \Big\{(x,y) \in \mathcal{B}^c \times \mathcal{B}^c : \Pi(x, y, p^{\mathcal{B}^c}(y), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(y)) \\ = \operatorname*{ess\,sup}_{\mathcal{B}^c} \Pi\Big(x, \cdot, p^{\mathcal{B}^c}(\cdot), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(\cdot)\Big)\Big\},\end{aligned}$$

has $\tau^{\mathcal{B}^c}$ measure equal to $\mu(\mathcal{B}^c)$. First, from Property 3 we have

$$\begin{aligned}V(x|\, p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c}) &= \operatorname*{ess\,sup}_{\mathcal{B}^c} \Pi\Big(x, \cdot, p^{\mathcal{B}^c}(\cdot), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(\cdot)\Big) \\ &= \begin{cases} V(y_1) + (x - y_0) & \text{if } x \in [y_0, y^c] \\ V(\overline{y}_1) & \text{if } [y^c, \overline{y}_1]. \end{cases}\end{aligned} \tag{A.37}$$

For the first term in Eq. (A.33) observe that $\tau((\mathcal{E}_{\mathcal{B}^c} \cap [y_0, y^c] \times [y_0, \overline{y}_1])_h)$ equals

$$\tau\Big(\Big\{(x,y) \in [y_1, \overline{y}_1] \times [y_1, \overline{y}_1] : \Pi(x-h, y-h, p^{\mathcal{B}^c}(y-h), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(y-h)) = V(y_1) + (x - y_1)\Big\}\Big),$$

using that $\Gamma(K^c) = 0$ and Eq. (A.47) one can verify that this expression equals

$$\tau\Big(\Big\{(x,y) \in [y_1, \overline{y}_1] \times [y_1, \overline{y}_1] : \Pi(x, y, p(y), s^\tau(y)) = V(x|\, p, \tau)\Big\}\Big).$$

In turn, from the definition of $\underline{y}_1$ and $\overline{y}_1$, and the fact that τ is an equilibrium flow this last expression equals $\mu([y_1, \overline{y}_1])$. For the second term in Eq. (A.33) we have

$$\mathcal{E}_{\mathcal{B}^c} \cap [y^c, \overline{y}_1] \times [y_0, \overline{y}_1] \cap \mathcal{D} = \Big\{(x,y) \in [y^c, \overline{y}_1] \times [y_0, \overline{y}_1] : \Pi(x, y, p^{\mathcal{B}^c}(y), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(y)) = V(\overline{y}_1)\Big\} \cap \mathcal{D},$$

Thus the second term in Eq. (A.33) equals

$$\mu\Big(\Big\{x \in [y^c, \overline{y}_1] : U(x, p^{\mathcal{B}^c}(x), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(x)) = V(\overline{y}_1)\Big\}\Big) = \mu([y^c, \overline{y}_1]) = \mu([y_0, y_1]),$$

where the first equality comes from Eq. (A.34) and the discussion that it follows it. The second equality comes from μ being invariant under translation and $y^c - y_0 = \overline{y}_1 - y_1$. Putting all these together yields

$$\tau^{\mathcal{B}^c}(\mathcal{E}_{\mathcal{B}^c}) = \mu([\overline{y}_1, y_1]) + \mu([y_0, y_1]) = \mu([y_0, \overline{y}_1]) = \mu(\mathcal{B}^c).$$

In order to create the new solution $(\tilde{p}, \tilde{\tau})$ we just use Lemma A.9 to paste the two solutions we constructed in $\mathcal{B}$ and $\mathcal{B}^c$. Note that the pasting is allowed because $V(y_0 |\, p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c}) = V(y_0)$ and $V(\overline{y}_1 |\, p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c}) = V(\overline{y}_1)$.

We now finally show the objective improvement. It is sufficient to prove that $\mathbf{Rev}_{[y_0,\overline{y}_1]}(\tilde{p}, \tilde{\tau}) > \mathbf{Rev}_{[y_0,\overline{y}_1]}(p, \tau)$,

$$\begin{aligned}
\mathbf{Rev}_{[y_0,\overline{y}_1]}(p, \tau) = \int_{[y_0,\overline{y}_1]} V(x)\, d\tau_2(x) &\overset{(a)}{<} \int_{[y_0,\overline{y}_1]} V(y_0)\, d\tau_2(x) \\
&\overset{(b)}{=} \int_{[y_0,\overline{y}_1]} V(y_0)\, d\tau_2^{\mathcal{B}^c}(x) \\
&\overset{(c)}{\leq} \int_{[y_0,\overline{y}_1]} V(x |\, p^{\mathcal{B}^c}, \tau^{\mathcal{B}^c})\, d\tau_2^{\mathcal{B}^c}(x) \\
&= \mathbf{Rev}_{[y_0,\overline{y}_1]}(\tilde{p}, \tilde{\tau}),
\end{aligned}$$

where in (a) use Property 1. In (b) we use that under τ no flow leaves or enters $\mathcal{B}^c$ and, thus,

$$\tau_2^{\mathcal{B}^c}(\mathcal{B}^c) = \tau^{\mathcal{B}^c}(\mathcal{B}^c \times \mathcal{B}^c) = \mu(\mathcal{B}^c) = \tau(\mathcal{B}^c \times \mathcal{C}) = \tau(\mathcal{B}^c \times \mathcal{B}^c) = \tau(\mathcal{C} \times \mathcal{B}^c) = \tau_2(\mathcal{B}^c).$$

In (c) we simply use Eq. (A.37).

In what follows we provide a complete proof of the three properties that we use to obtain the result.

Property 1. $\tau_2((y_0, \hat{x}]) > 0$.

Proof of Property 1. First we show that $\exists\, h \in (0, \hat{x} - y_0)$ such that $\tau((y_0, y_0 + h) \times [\hat{x}, y_1]) = 0$. Suppose this is not true then for all $n \in \mathbb{N}$ large enough we have that $\tau((y_0, y_0 + \frac{1}{n}) \times [\hat{x}, y_1]) > 0$, which thanks to Lemma A.5 implies that for all $n \in \mathbb{N}$ large enough there exists $(x_n, y_n) \in (y_0, y_0 + \frac{1}{n}) \times [\hat{x}, y_1]$ such that $y_n \in \mathcal{IR}(x_n)$. Our envelope result (see Lemma 1.3) ensures that $V(x_n) = V(y_n) - |y_n - x_n|$. Since $y_n \in [\hat{x}, y_1]$ we must have $V(y_n) \leq V(y_0)$ for all $n \in \mathbb{N}$ large (when y_1 is not well defined we replaced by H and the argument still goes through). Furthermore, x_n converges to y_0 so the continuity of $V(\cdot)$ yields

$$V(y_0) = \lim_{n\to\infty} V(x_n) = \lim_{n\to\infty} V(y_n) - |y_n - x_n| \leq V(y_0) - \lim_{n\to\infty} (y_n - x_n) < V(y_0),$$

not possible. We conclude that $\exists\, h \in (0, \hat{x} - y_0)$ such that $\tau((y_0, y_0 + h) \times [\hat{x}, y_1]) = 0$. Note that the same must be true for some $h \in (0, (\hat{x} - y_0) \wedge \frac{(y_1 - y_0)}{2})$. We fix h in this interval with the property we just proved.

Next, note we also have that $\tau((y_0, y_0 + h) \times (y_1, H]) = 0$; otherwise, by Lemma A.5 we can find $(x, y) \in (y_0, y_0 + h) \times (y_1, H]$ such that $y \in \mathcal{IR}(x)$, which implies that $y \in \mathcal{IR}(y_1)$. Using the envelope result delivers $V(y_1) = V(y) - |y - y_1|$ and $V(x) = V(y) - |y - x|$. Since $V(y_1) = V(y_0)$ we have $(y_1 - x) = V(y_0) - V(x)$, but our choice of h implies that $y_1 - x > h$ thus

$$h < (y_1 - x) = V(y_0) - V(x) \leq |y_0 - x| \leq h,$$

again a contradiction. The last inequality comes from the Lipschitz property (see Lemma 1.1). In summary, we have that there exists $h \in (0, (\hat{x} - y_0) \wedge \frac{(y_1 - y_0)}{2})$ such that $\tau((y_0, y_0+h)\times[\hat{x}, H]) = 0$. To conclude the proof note the following

$$\begin{aligned}
0 &\overset{(a)}{<} \mu((y_0, y_0+h)) \\
&= \tau((y_0, y_0+h)\times\mathcal{C}) \\
&\overset{(b)}{=} \tau((y_0, y_0+h)\times[y_0, H]) \\
&= \tau((y_0, y_0+h)\times[y_0, \hat{x})) + \tau((y_0, y_0+h)\times[\hat{x}, H]) \\
&= \tau((y_0, y_0+h)\times[y_0, \hat{x})) \\
&\leq \tau_2([y_0, \hat{x}]) \\
&\overset{(c)}{=} \tau_2((y_0, \hat{x}]),
\end{aligned}$$

where (a) comes from the fact that the measure μ has full support in $\mathcal{C}$. The equality (b) holds because by construction no flow leaves $[y_0, H]$, and (c) is true because $\tau_2 \ll \Gamma$ and Γ does not have atoms in $[y_0, \hat{x}]$. This concludes the proof of Property 1.

Property 2. Both $\overline{y}_1$ and $\underline{y}_1$ are achieved in the set where they are defined. Furthermore, for any $z \in [\underline{y}_1, \overline{y}_1]$, $\overline{y}_1 \in \mathcal{IR}(z)$.

Proof of Property 2. First we show both

$$\exists\, y_q \in [y_0, y_1] \text{ such that } \overline{y}_1 \in \mathcal{IR}(y_q) \quad \text{and} \quad \exists\, x_q \in [y_1, H] \text{ such that } x_q \in \mathcal{IR}(\underline{y}_1). \tag{A.38}$$

Let us begin with the first statement. Let x^n be a sequence in A converging to $\overline{y}_1$, where

$$A = \{x \in [y_1, H] : \exists\, y \in [y_0, y_1] \text{ such that } x \in \mathcal{IR}(y)\}.$$

Then there exists a sequence $\{y^n\} \subset [y_0, y_1]$ such that $x^n \in \mathcal{IR}(y^n)$. Note that since $\{y^n\} \subset [y_0, y_1]$ and $x^n \in [y_1, H]$, Lemma A.3 implies that $x^n \in \mathcal{IR}(y_1)$. Fix $\epsilon > 0$ and $\delta > 0$ then we can find $n_0(\delta)$ such that for all $n \geq n_0(\delta)$ we have $B(x_n, \delta/2) \subset B(\overline{y}_1, \delta)$. This implies that $V_{B(x_n,\delta/2)}(y_1) \leq V_{B(\overline{y}_1,\delta)}(y_1)$ for all $n \geq n_0(\delta)$. Fix $n \geq n_0(\delta)$, because $x^n \in \mathcal{IR}(y_1)$ we know that

$$\exists\, \delta_0(\epsilon, n) \text{ such that } \forall \hat{\delta} \leq \delta_0(\epsilon, n)\;\; V_{B(x_n,\hat{\delta})}(y_1) \geq V(y_1) - \epsilon.$$

Let $r_0 = \delta_0(\epsilon, n) \wedge \frac{\delta}{2}$ then for all $\hat{\delta} \le r_0$ we have

$$V_{B(\overline{y}_1,\delta)}(y_1) \ge V_{B(x_n,\delta/2)}(y_1) \ge V_{B(x_n,\hat{\delta})}(y_1) \ge V(y_1) - \epsilon.$$

This shows that for any $\epsilon, \delta > 0$ we have $V_{B(\overline{y}_1,\delta)}(y_1) \ge V(y_1) - \epsilon$. That is, $\overline{y}_1 \in \mathcal{IR}(y_1)$.

Now we prove that $\underline{y}_1 \in A$ where

$$A = \{y \in [y_0, y_1] : \exists\, x \in [y_1, H] \text{ such that } x \in \mathcal{IR}(y)\}.$$

By the definition of $\underline{y}_1$ we can always construct a sequence $\{y^n\} \subset A$ converging to $\underline{y}_1$. From the definition of A there exists another sequence $\{x^n\} \subset [y_1, H]$ such that $x^n \in \mathcal{IR}(y^n)$ for all n. Fix $\epsilon > 0$ then we can always find $n_0(\epsilon)$ such that for all $n \ge n_0(\epsilon)$ we have $|y^n - \underline{y}_1| \le \epsilon/3$. Fix $n \ge n_0(\epsilon)$ then since $x^n \in \mathcal{IR}(y^n)$ we have

$$\exists\, \delta_0(\epsilon, n) \text{ such that } \forall \delta \le \delta_0(\epsilon, n)\ V_{B(x_n,\delta)}(y^n) \ge V(y^n) - \frac{\epsilon}{3}, \tag{A.39}$$

but from the Lipchitz property we can deduce that

$$V_{B(x_n,\delta)}(y^n) \le V_{B(x_n,\delta)}(\underline{y}_1) + \frac{\epsilon}{3} \quad \text{and} \quad V(y^n) \ge V(\underline{y}_1) - \frac{\epsilon}{3}.$$

Replacing this in Eq. (A.39) yields

$$\exists\, \delta_0(\epsilon, n) \text{ such that } \forall \delta \le \delta_0(\epsilon, n)\ V_{B(x_n,\delta)}(\underline{y}_1) \ge V(\underline{y}_1) - \epsilon,$$

that is, $x^n \in \mathcal{IR}(\overline{y}_1)$. This concludes the proof for Eq. (A.38).

Next, we show that for all $z \in [\underline{y}_1, \overline{y}_1]$, $\overline{y}_1 \in \mathcal{IR}(z)$. First, from our previos argument we know there exists y_q and x_q as in Eq. (A.38). Then Lemma A.3 implies $\overline{y}_1 \in \mathcal{IR}(z)$ for all $z \in [y_q, \overline{y}_1]$. Observe that this yields $\overline{y}_1 \in \mathcal{IR}(x_q)$ because $x_q \in [y_q, \overline{y}_1]$. Take $z \in [\underline{y}_1, y_q]$ then since $x_q \in \mathcal{IR}(\underline{y}_1)$ from Lemma A.3 we conclude that $x_q \in \mathcal{IR}(z)$. Using envelope result, Lemma 1.3, we have that $V(x_q) = V(z) + (x_q - z)$. Furthermore, fix $\epsilon > 0$ then since $\overline{y}_1 \in \mathcal{IR}(x_q)$ we have

$$\exists\, \delta_0(\epsilon) \text{ such that } \forall \delta \le \delta_0(\epsilon)\ V_{B(\overline{y}_1,\delta)}(x_q) + \epsilon \ge V(x_q) = V(z) + (x_q - z). \tag{A.40}$$

Thus for any $\delta \leq \delta_0(\epsilon)$, the Lipchitz property and Eq. (A.40) yield

$$V_{B(\overline{y}_1,\delta)}(z) \geq V_{B(\overline{y}_1,\delta)}(x_q) - (x_q - z) \geq V(z) + (x_q - z) - (x_q - z) - \epsilon = V(z) - \epsilon,$$

which implies that $\overline{y}_1 \in \mathcal{IR}(z)$. This concludes the proof of Property 2.

Property 3. Both Eq. (A.34) and Eq. (A.37) hold.

Proof of Property 3. Let us star with Eq. (A.34). In order to prove the first part in Eq. (A.34) consider the following set

$$K = \Big\{x \in [y_0, y^c] : \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(x) \leq \frac{d\tau_2}{d\Gamma}(x+h)\Big\}.$$

We want to show that $\Gamma(K^c) = 0$ (the complement is taken with respect to $[y_0, y^c]$). If this is not true then $\Gamma(K^c) > 0$ and we have

$$\tau_2^{\mathcal{B}^c}(K^c) = \int_{K^c} \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(x)\, d\Gamma(x) > \int_{K^c} \frac{d\tau_2}{d\Gamma}(x+h)\, d\Gamma(x) = \tau_2(K^c + h). \tag{A.41}$$

However,

$$\begin{aligned}
\tau_2^{\mathcal{B}^c}(K^c) &= \tau\Big(([y_0, y^c] \times K^c)_h\Big) + \mu(\pi_1(([y^c, \overline{y}_1] \times K^c) \cap \mathcal{D})) \\
&= \tau\Big(([y_0, y^c] \times K^c)_h\Big) \\
&= \tau\Big([y_0 + h, y^c + h] \times (K^c + h)\Big) \\
&\leq \tau\Big(\mathcal{C} \times (K^c + h)\Big) \\
&= \tau_2(K^c + h).
\end{aligned}$$

This together with Eq. A.41 yield a contradiction. To prove the second part of Eq. (A.34) consider any $\mathcal{R} \subset [y^c, \overline{y}_1]$, and observe that

$$\tau_2^{\mathcal{B}^c}(\mathcal{R}) = \tau\Big([y_1, \overline{y}_1] \times (\mathcal{R} + h)\Big) + \mu(\mathcal{R}) = \mu(\mathcal{R}) = \int_{\mathcal{R}} \mu_1 d\Gamma(x),$$

where the second equality comes from $\mathcal{R}+h \subset [\overline{y}_1, \overline{y}_1+h]$ and $\tau([y_1, \overline{y}_1] \times [\overline{y}_1, \overline{y}_1+h]) = 0$.

Finally, we provide a proof for Eq. (A.37). Let $Z(x) \triangleq \min\{V(y_0) + (x - y_0), V(\overline{y}_1)\}$. We verify that for all $x \in \mathcal{B}^c$

$$Z(x) \geq U\Big(w, p^{\mathcal{B}^c}(w), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(w)\Big) - |w - x|, \quad \Gamma - a.e.\ w \text{ in } \mathcal{B}^c, \tag{A.42}$$

and that $Z(x)$ is the smallest with such property. First, fix $x \in [y^c, \overline{y}_1]$ so $Z(x) = V(\overline{y}_1)$. Note that from our choice of prices in $[y^c, \overline{y}_1]$ we have

$$Z(x) = V(\overline{y}_1) \geq V(\overline{y}_1) - |w-x| = U\Big(w, p^{\mathcal{B}^c}(w), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(w)\Big) - |w-x|, \quad \Gamma - a.e. w \text{ in} [y^c, \overline{y}_1].$$

So we only need to show the same inequality but this time for $[y_0, y^c]$. From the definition of $\underline{y}_1$ and $\overline{y}_1$, Lemma A.3 and Lemma 1.3 we have that $V(\overline{y}_1) - |\overline{y}_1 - y_1|$ equals $V(y_1|\, p, \tau)$ and, therefore,

$$\begin{aligned} V(\overline{y}_1) &\geq U(w, p(w), s^\tau(w)) - |w - y_1| + |\overline{y}_1 - y_1|, \quad \Gamma - a.e.\ w \text{ in } [y_1, \overline{y}_1] \\ &\geq U(w, p(w), s^\tau(w)), \quad \Gamma - a.e.\ w \text{ in } [y_1, \overline{y}_1]. \end{aligned}$$

We can use this together with the fact that $[y_0, y^c] + h = [y_1, \overline{y}_1]$ to obtain

$$\begin{aligned} Z(x) = V(\overline{y}_1) &\overset{(a)}{\geq} U\Big(w + h, p(w + h), s^\tau(w + h)\Big), \quad \Gamma - a.e.\ w \text{ in } [y_0, y^c] \\ &\geq U\Big(w + h, p(w + h), s^\tau(w + h)\Big) - |w - x|, \quad \Gamma - a.e.\ w \text{ in } [y_0, y^c] \\ &\overset{(b)}{=} U\Big(w, p^{\mathcal{B}^c}(w), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(w)\Big) - |w - x|, \quad \Gamma - a.e.\ w \text{ in } [y_0, y^c], \end{aligned}$$

Inequality (a) comes from the fact that Γ in the interval under consideration is invariant under a shift; (b) comes from Eq. (A.47). That is, for $x \in [y^c, \overline{y}_1]$ Eq. (A.42) is satisfied. It is left to verify that $Z(x)$ is the smallest value satisfying Eq. (A.42). For any $\epsilon > 0$, since $x \in [y^c, \overline{y}_1]$ we have

$$\begin{aligned} 0 &< \Gamma(B(x, \epsilon) \cap [y^c, \overline{y}_1]) \\ &= \Gamma\Big(w \in [y^c, \overline{y}_1] : V(\overline{y}_1) - |w - x| > V(\overline{y}_1) - \epsilon\Big) \\ &= \Gamma\Big(w \in [y^c, \overline{y}_1] : U\Big(w, p^{\mathcal{B}^c}(w), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(w)\Big) - |w - x| > V(\overline{y}_1) - \epsilon\Big), \end{aligned}$$

hence $V(\overline{y}_1)$ is the smallest value satisfying Eq. (A.42).

Now we show Eq. (A.42) for $x \in [y_0, y^c]$. Fix $x \in [y_0, y^c]$ so $Z(x) = V(y_0) + (x - y_0)$. Note that $V(y_0)$ equals $V(y_1)$, and from the definition of $\overline{y}_1$ and the envelope result we have that $V(y_1)$ equals $V(\overline{y}_1) - (\overline{y}_1 - y_1)$. Therefore,

$$\begin{aligned} Z(x) &= V(\overline{y}_1) - (\overline{y}_1 - y_1) + (x - y_0) \\ &\overset{(a)}{\geq} V(\overline{y}_1) - (w - x), \quad \Gamma - a.e. \ w \ \text{ in } [y^c, \overline{y}_1] \\ &\overset{(b)}{=} U\Big(w, p^{\mathcal{B}^c}(w), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(w)\Big) - |w - x|, \quad \Gamma - a.e. \ w \ \text{ in } [y^c, \overline{y}_1], \end{aligned}$$

where (a) follows from $w \geq y_c$ and $y^c - y_0 = \overline{y}_1 - y_1$. Line (b) holds from our choice of prices in $[y^c, \overline{y}_1]$. Hence, $Z(x)$ upper bounds (almost surely) the desire quantity in $[y^c, \overline{y}_1]$, so we just need to prove the same bound for $[y_0, y^c]$. Note that from the definition of $\underline{y}_1$ and $\overline{y}_1$ we have that

$$V(x + h) = V(y_1) + (x + h - y_1) = V(y_1) + (x - y_0) = Z(x),$$

and thus

$$\begin{aligned} Z(x) &= V(x + h|\, p, \tau) \\ &\overset{(a)}{\geq} U(w, p(w), s^\tau(w)) - |w - (x + h)|, \quad \Gamma - a.e. \ w \ \text{ in } [y_1, \overline{y}_1] \\ &\overset{(b)}{=} U(w + h, p(w + h), s^\tau(w + h)) - |w + h - (x + h)|, \quad \Gamma - a.e. \ w \ \text{ in } [y_0, y^c] \\ &\overset{(c)}{=} U(w, p^{\mathcal{B}^c}(w), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(w)) - |w - x)|, \quad \Gamma - a.e. \ w \ \text{ in } [y_0, y^c], \end{aligned}$$

where (a) comes from the definition of $V(x + h|\, p, \tau)$, (b) from the invariance under translation of Γ. Line (c) follows from Eq. (A.47). Therefore, $Z(x)$ satisfies Eq. (A.42). To see why $Z(x)$ is the smallest value satisfying this equation observe that

$$\begin{aligned} 0 &< \Gamma(B(y^c, \epsilon) \cap [y^c, \overline{y}_1]) \\ &\overset{(a)}{=} \Gamma\Big(w \in [y^c, \overline{y}_1] : V(\overline{y}_1) - (w - x) > V(\overline{y}_1) - (\overline{y}_1 - y_1) + (x - y_0) - \epsilon\Big) \\ &= \Gamma\Big(w \in [y^c, \overline{y}_1] : U\Big(w, p^{\mathcal{B}^c}(w), \frac{d\tau_2^{\mathcal{B}^c}}{d\Gamma}(w)\Big) - |w - x| > Z(x) - \epsilon\Big), \end{aligned}$$

where in (a) we use that $y^c - y_0 = \overline{y}_1 - y_1$. This implies that $Z(x)$ is the smallest value satisfying Eq. (A.42), completing the proof.

□

Proof of Proposition 1.9. If $X_r = H$ there is nothing to prove, so assume $X_r < H$. Let (p, τ) be a feasible solution such that $V(\cdot|p, \tau)$ is non-decreasing. Due to Proposition 1.8 we can always restrict attention to this type of solution. We proceed by contradiction. Assume that there exists $\tilde{x} \in (X_r, H]$ such that

$$V(\tilde{x}) < \min\{V(X_r) + (\tilde{x} - X_r), \psi_1\} \triangleq Z(\tilde{x}). \tag{A.43}$$

First, we construct an interval $\tilde{I}$ such that $\tau_2(\tilde{I}) > 0$ and $V(x) < Z(x)$ for all $x \in \tilde{I}$. Then, we show that $Z(x)$ can be achieved in a feasible manner by appropriately creating a price-equilibrium pair $(\tilde{p}, \tilde{\tau})$ that mimics the flow generated by τ in $(X_r, H]$. The final step of the proof is to use the interval $\tilde{I}$ and the flow $\tilde{\tau}$ to show an strict objective improvement.

Interval construction. From Eq. (A.43) and the continuity of $V(\cdot)$ we can deduce the existence of an interval $[\tilde{a}, \tilde{b}] \subset (X_r, H]$ such $V(x) < Z(x)$ for all $x \in [\tilde{a}, \tilde{b}]$. Furthermore, the Lipchitz property (see Lemma 1.1) and Lemma 1.6 imply that $V(x) < Z(x)$ for all $x \in [\tilde{a}, \tilde{c}]$ where $\tilde{c}$ is the minimum between H and the value c such that $V(\tilde{a}) + (c - \tilde{a}) = \psi_1$. Also, Proposition 1.8 and Lemma A.5 together with Lemma 1.3 imply that $\tau([\tilde{a}, \tilde{c}] \times \mathcal{C}) = \tau([\tilde{a}, \tilde{c}] \times [\tilde{a}, \tilde{c}])$. Putting all of this together we conclude that there exists an interval $\tilde{I} = (\tilde{a}, \tilde{c})$ such that $\tau_2(\tilde{I}) > 0$ and $V(x) < Z(x)$ for all $x \in \tilde{I}$.

Flow mimicking. Define the collection of intervals

$$\mathcal{I} \triangleq \{I \subset (X_r, H] : I = [a, b], a < b, b \in \mathcal{IR}(a), a \text{ is minimal and } b \text{ is maximal}\}.$$

There are two cases: $\mathcal{I} = \emptyset$ and $\mathcal{I} \neq \emptyset$. We only do the latter because its treatment contains the former.

Suppose $\mathcal{I} \neq \emptyset$, then there exists $X_r < a < b$ such that $b \in \mathcal{IR}(a)$, where a and b are minimal and maximal with this property, respectively. We first look at some properties of the equilibrium in each element of $\mathcal{I}$ and then we look at its complement.

Note that from the minimality of a we have that for any $x < a$, $a \notin \mathcal{IR}(x)$. Similarly, for any $x > b$ we have $x \notin \mathcal{IR}(b)$. This, together with Proposition 1.8 and Lemma A.5 imply that $[a, b]$ is a flow-separated region, that is, there is no flow coming in nor flow going out of $[a, b]$, $\tau([a, b] \times [a, b]^c) = 0$ and $\tau([a, b]^c \times [a, b]) = 0$. Observe that our flow separation result in Lemma A.9 implies that in each interval $I \in \mathcal{I}$ we have an equilibrium. Furthermore, from Lemma 1.3 we must have

$$V(x) = V(a) + (x - a), \quad \forall x \in [a, b].$$

From the previous discussion we infer that the elements in the collection $\mathcal{I}$ are disjoint intervals and, since V is non-decreasing, the collection is at most countable.

For any a, b such that $[a, b] \in \mathcal{I}$ we define

$$t(a) \triangleq V(a) - V(X_r) + X_r, \quad \text{and} \quad t(b) \triangleq V(b) - V(X_r) + X_r.$$

Note that since V is non-decreasing we have $V(a) \geq V(X_r)$ and, therefore, $t(b) >$

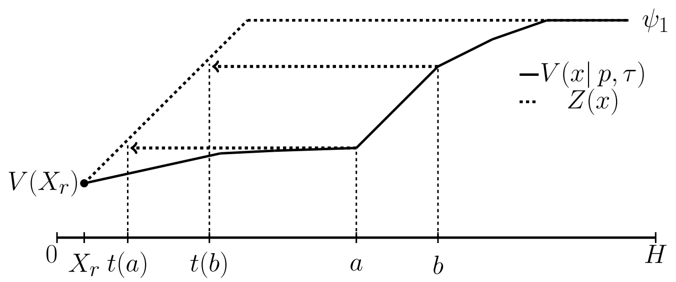


Figure A.2: Graphical representation of $t(a)$ and $t(b)$.

$t(a) \geq X_r$. Also, for any such b we have $t(b) < Y_r$. The points $t(a), t(b)$ are the corresponding points to a, b in the interval $[X_r, Y_r]$ (see Figure A.2). Furthermore, $t(\cdot)$ is a non-decreasing mapping.

We denote by $\mathcal{I}^c$ the collection of intervals whose elements are the intervals that do not belong to $\mathcal{I}$. Observe that the elements in $\mathcal{I}$ and $\mathcal{I}^c$ alternate in a consecutive manner. That is, if we have an interval $(c,d) \in \mathcal{I}^c$ then it can only be followed by and interval $[a,b] \in \mathcal{I}$ with $a = d$. In the case that $I = (c,d) \in \mathcal{I}^c$ is not followed by an interval in $\mathcal{I}$ then I equals $(c, H]$. Define the sets

$$\mathcal{K} \triangleq \bigcup_{I \in \mathcal{I}} I \text{ and } \mathcal{K}^c \triangleq \bigcup_{I \in \mathcal{I}^c} I.$$

Note that $(X_r, H] = \mathcal{K} \cup \mathcal{K}^c$ up to a set of Γ measure zero. Also, for each interval $I \in \mathcal{I}^c$ we must have that for all measurable sets $A \subset I$, $\tau(A \times A) = \mu(A) = \tau_2(A)$; otherwise, by Lemma A.5 we would get a contradiction with the definition of $\mathcal{I}$. In turn, this implies that $\frac{d\tau_2}{d\Gamma}(x) = \mu_1$, $\Gamma - a.e.\ x$ in $\mathcal{K}^c$.

We denote by $\mathcal{I}_t$ the collection of intervals $\{[t(a), t(b)]\}_{[a,b] \in \mathcal{I}}$, and $\mathcal{I}_t^c$ is defined in analogous manner. Also, $\mathcal{K}_t$ and $\mathcal{K}_t^c$ are defined similarly to $\mathcal{K}$ and $\mathcal{K}^c$ replacing $\mathcal{I}$ with $\mathcal{I}_t$ and $\mathcal{I}^c$ with $\mathcal{I}_t^c$, respectively.

The idea now is to construct a solution $(\tilde{p}, \tilde{\tau})$ in $(X_r, H]$ and then paste it with the old solution (p, τ) restricted to $[-H, X_r)$. To construct $(\tilde{p}, \tilde{\tau})$ we will make use of the collections $\mathcal{I}_t$ and $\mathcal{I}_t^c$. For each element in these collections we will create a price-equilibrium. For intervals $[t(a), t(b)] \in \mathcal{I}_t$ the idea is that the solution $(\tilde{p}, \tilde{\tau})$ has the same equilibrium than (p, τ) in $[a, b]$. For the interval in $\mathcal{I}_t^c$ we choose prices such that no drivers will have an incentive to move. Finally, using Lemma A.9 we will paste the equilibria generated in all the intervals.

First, we show how to construct prices and an equilibrium in some $[t(a), t(b)]$. Fix $[a,b] = I \in \mathcal{I}$ and denote the mimicking set $[t(a), t(b)]$ by I_t. Choose prices $p^{I_t}(x)$ equal to $p(x + (a - t(a)))$ for all $x \in I_t$. For the flows, we define τ^{I_t} for any $\mathcal{L} \subseteq I_t \times I_t$ by

$$\tau^{I_t}(\mathcal{L}) = \tau\Big(\mathcal{L} + (a - t(a), a - t(a))\Big),$$

that is, τ^{I_t} mimics τ in $I \times I$. It can be shown that (see Property 1 at the end of this proof) (p^{I_t}, τ^{I_t}) forms a price-equilibrium pair in I_t such that $\tau^{I_t} \in \mathcal{F}_{I_t}(\mu|_{I_t})$. Also,

$V(x|\, p^{I_t}, \tau^{I_t})$ equals $V(x+a-t(a)|\, p, \tau)$ for all $x \in I_t$, and

$$\frac{d\tau_2^{I_t}}{d\Gamma}(x) = \frac{d\tau_2}{d\Gamma}(x+a-t(a)), \quad \Gamma - a.e. \ x \ \text{ in } \ I_t. \tag{A.44}$$

Furthermore, because $I \in \mathcal{I}$ we have

$$V(x|p^{I_t}, \tau^{I_t}) = V(x+a-t(a)|p, \tau) = V(a)+(x-t(a)) = V(X_r)+(x-X_r) = Z(x), \quad \forall x \in I_t,$$

that is, for all intervals I_t the associated solution (p^{I_t}, τ^{I_t}) achieves the upper bound $Z(x)$.

Second, we show how to set the prices and construct an equilibrium everywhere else. Consider any two consecutive sets in $\mathcal{I}$, $I_1 = [a_1, b_1]$ and $I_2 = [a_2, b_2]$. The corresponding mimicking sets are $[t(a_1), t(b_1)]$ and $[t(a_2), t(b_2)]$. We need to set prices and define the flow in the interval $J_t = (t(b_1), t(a_2))$. We choose the prices p^{J_t} to be such that

$$U\Big(x, p^{J_t}(x), \mu_1\Big) = Z(x), \quad \forall x \in J_t.$$

Since $Z(x) \leq \psi_1$ these prices are guaranteed to exist. We define the measure τ^{J_t} for any measurable set $\mathcal{L} \subseteq J_t \times J_t$ by

$$\tau^{J_t}(\mathcal{L}) = \mu(\pi_1(\mathcal{L} \cap \mathcal{D})).$$

This measure has $d\tau_2^{J_t}/d\Gamma = \mu_1$, $\Gamma - a.e$ in J_t. It can be shown that (see Property 2 at the end of this proof) (p^{J_t}, τ^{J_t}) forms a price-equilibrium pair in J_t such that $\tau^{J_t} \in \mathcal{F}_{J_t}(\mu|_{J_t})$ and $V(x|\, p^{J_t}, \tau^{J_t})$ equals $Z(x)$ for all $x \in J_t$.

Third, the solutions $\{(p^{I_t}, \tau^{I_t})\}_{I_t \in \mathcal{I}_t}$ and $\{(p^{J_t}, \tau^{J_t})\}_{J_t \in \mathcal{I}_t^c}$ cover the whole interval $(X_r, H]$. Moreover they are defined in disjoint interval, and are such that the respective $V(\cdot)$ functions coincide at the boundaries of the interval (these functions coincide with $Z(\cdot)$). Thus, we can apply Lemma A.9 to paste all these solutions and obtain a new solution $(\tilde{p}, \tilde{\tau})$ in $(X_r, H]$. As mentioned before we can use the same lemma to paste this solution with the old solution restricted to $[-H, X_r]$. This would yield a solution in the entire city.

Objective improvement. Consider the revenue under (p,τ) in $(X_r, H]$, it easy to observe that

$$\begin{aligned}
\mathbf{Rev}_{(X_r,H]}(p,\tau) &= \int_{(X_r,H]} V(x|p,\tau)\cdot s^{\tau}(x)\,d\Gamma(x)\\
&= \int_{\mathcal{K}} V(x|p,\tau)\cdot s^{\tau}(x)\,d\Gamma(x) + \int_{\mathcal{K}^c} V(x|p,\tau)\cdot s^{\tau}(x)\,d\Gamma(x)\\
&= \underbrace{\sum_{I\in\mathcal{I}}\int_I V(x|p,\tau)\cdot s^{\tau}(x)\,d\Gamma(x)}_{=(a)} + \underbrace{\int_{\mathcal{K}^c} V(x|p,\tau)\cdot s^{\tau}(x)\,d\Gamma(x)}_{=(b)}.
\end{aligned}$$

Let us develop the integral of the term (a). Let I be equal to $[a,b]$ and I_t equal to $[t(a),t(b)]$ then

$$\begin{aligned}
\int_{[a,b]} V(x|p,\tau)\cdot s^{\tau}(x)\,d\Gamma(x) &= \int_{[t(a),t(b)]} V(x+a-t(a)|p,\tau)\cdot s^{\tau}(x+a-t(a))\,d\Gamma(x)\\
&= \int_{[t(a),t(b)]} V(x|p^{I_t},\tau^{I_t})\cdot s^{\tau^{I_t}}(x)\,d\Gamma(x),
\end{aligned}$$

where in the first line we use the invariance under translation of Γ, and in the second line we use that $V(x|\,p^{I_t},\tau^{I_t})$ equals $V(x+a-t(a)|\,p,\tau)$ for all $x\in I_t$ and Eq. (A.44). Thus,

$$\begin{aligned}
\mathbf{Rev}_{(X_r,H]}(p,\tau) &= \sum_{I_t\in\mathcal{I}_t}\int_{I_t} V(x|p^{I_t},\tau^{I_t})\cdot s^{\tau^{I_t}}(x)\,d\Gamma(x) + (b)\\
&= \int_{\mathcal{K}_t} Z(x)\cdot s^{\tilde{\tau}}(x)\,d\Gamma(x) + (b).
\end{aligned}$$

Thus, to conclude the proof we only need to show that

$$(b) = \int_{\mathcal{K}^c} V(x|p,\tau)\cdot s^{\tau}(x)\,d\Gamma(x) < \int_{\mathcal{K}_t^c} Z(x)\cdot s^{\tilde{\tau}}(x)\,d\Gamma(x). \tag{A.45}$$

Define the following functions

$$V_e(x) = \begin{cases} V(x|p,\tau) & \text{if } x\in\mathcal{K}^c,\\ V(a|p,\tau) & \text{if } x\in[a,b],\ \text{some } [a,b]\in\mathcal{I}, \end{cases}$$

$$Z_e(x) = \begin{cases} Z(x) & \text{if } x\in\mathcal{K}_t^c,\\ Z(t(a)) & \text{if } x\in[t(a),t(b)],\ \text{some } [t(a),t(b)]\in\mathcal{I}_t. \end{cases}$$

We verify that $V_e(x) \leq Z_e(x)$ for all $x \in (X_r, H]$, and the we use this inequality to prove the objective improvement. Let $x \in \mathcal{K}^c$ then there exists an interval $(c, d) \in \mathcal{I}^c$ with $x \in (c, d)$. If $x \in \mathcal{K}_t^c$ then the upper bound is trivial. If $x \notin \mathcal{K}_t^c$ then $x \in [t(a), t(b)]$ for some $[t(a), t(b)] \in \mathcal{I}_t$. We must have that $a \geq d$; otherwise, since $(c, d) \in \mathcal{I}$, it must be the case that $b \leq c$. In turn, this implies that $[t(a), t(b)] \cap (c, d) = \emptyset$ which contradiction our current assumption. Therefore,

$$V_e(x) = V(x|p, \tau) \leq V(d|p, \tau) \leq V(a|p, \tau) = Z(t(a)) = Z_e(x).$$

Let $x \in [a, b]$ for some $[a, b] \in \mathcal{I}$. If $x \in \mathcal{K}_t^c$, $t(b) < x$ otherwise we would have that $t(a) \leq a \leq a \leq t(b)$, that is, $x \in [t(a), t(b)] \in \mathcal{I}_t$. Under our current assumption this is not possible. Then,

$$V_e(x) = V(a|p, \tau) < V(b|p, \tau) = Z(t(b)) \leq Z(x) = Z_e(x), \tag{A.46}$$

that is, when $x \in \mathcal{K} \cap \mathcal{K}_t^c$ we have $V_e(x) < Z_e(x)$. If $x \in [t(\hat{a}), t(\hat{b})]$ for some $[t(\hat{a}), t(\hat{b})] \in \mathcal{I}_t$. Using similar arguments as before we can show that $\hat{a} \geq a$ and, therefore,

$$V_e(x) = V(a|p, \tau) = Z(t(a)) \leq Z(t(\hat{a})) = Z_e(x).$$

Now, recall that in the **Interval construction** part of the proof we defined an interval $\tilde{I} = [\tilde{a}, \tilde{c}]$ in which the function $V(\cdot|p, \tau)$ is uniformly strictly bounded by $Z(\cdot)$. Now we relate this interval to $\mathcal{K}_t^c$ by showing that there exists $\epsilon > 0$ such that $(\tilde{c} - \epsilon, \tilde{c}) \subseteq I_t^c$ with $I_t^c \in \mathcal{I}_t^c$. The idea is to use that $(\tilde{c} - \epsilon, \tilde{c}) \subset \tilde{I}$ and $(\tilde{c} - \epsilon, \tilde{c}) \subset \mathcal{K}_t^c$ together with Eq. (A.46) to show an strict objective improvement.

Note that if $\tilde{c} = H$ then

$$
\begin{aligned}
\sup_{[t(a),t(b)]\in\mathcal{I}_t} t(b) &\overset{(1)}{\leq} t(\tilde{c}) \\
&= V(\tilde{c}) - V(X_r) + X_r \\
&= (V(\tilde{c}) - V(\tilde{a})) + (V(\tilde{a}) - V(X_r)) + X_r \\
&\overset{(2)}{<} (V(\tilde{c}) - V(\tilde{a})) + (Z(\tilde{a}) - Z(X_r)) + X_r \\
&\overset{(3)}{\leq} (\tilde{c} - \tilde{a}) + (\tilde{a} - X_r) + X_r \\
&= \tilde{c},
\end{aligned}
$$

where (1) comes from the fact that $t(\cdot)$ is non-decreasing and $\tilde{c} = H$, line (2) follows from the $V(\tilde{a}) < Z(\tilde{a})$ and $V(X_r) = Z(X_r)$. Inequality, (3) holds because both V and Z are $1-$Lipschitz functions. In the case that $\tilde{c} < H$ we have $V(\tilde{a}) + (\tilde{c} - \tilde{a}) = \psi_1$. Also, we always have that $t(b) \leq Y_r$ where Y_r is such that $V(X_r) + (Y_r - X_r) = \psi_1$. From this we deduce that $Y_r < \tilde{c}$ and, therefore, we have that $\sup_{[t(a),t(b)]\in\mathcal{I}_t} t(b) < \tilde{c}$. Either way we can always find $\epsilon \in (0, \tilde{c} - \tilde{a})$ such that the interval $(\tilde{c} - \epsilon, \tilde{c})$ does not intersect with any interval in $\mathcal{I}_t$. Hence, since $\mathcal{I}_t^c$ are all the intervals that do not belong to $\mathcal{I}_t$ we must have that $(\tilde{c} - \epsilon, \tilde{c}) \subseteq I_t^c$ for some $I_t^c \in \mathcal{I}_t^c$.

Because $(\tilde{c} - \epsilon, \tilde{c})$ is a subset of both $\mathcal{K}_t^c$ and $(\tilde{a}, \tilde{c})$, for $x \in (\tilde{c} - \epsilon, \tilde{c}) \cap \mathcal{K}^c$ we have $V_e(x) < Z_e(x)$. Also, for $x \in (\tilde{c} - \epsilon, \tilde{c}) \cap \mathcal{K}$ from equation Eq. (A.46) we have

$V_e(x) < Z_e(x)$. That is, $V_e(x) < Z_e(x)$ for all $x \in (\tilde{c} - \epsilon, \tilde{c})$ and, therefore,

$$\begin{aligned}
\int_{\mathcal{K}^c} V(x|p,\tau) \cdot s^\tau(x)\, d\Gamma(x) &= \int_{(X_r,H]} V_e(x|p,\tau) \cdot \mu_1 d\Gamma(x) \\
&\quad - \sum_{[a,b]\in\mathcal{I}} \int_{[a,b]} V(a|p,\tau) \cdot \mu_1 d\Gamma(x) \\
&< \int_{(X_r,H]} Z_e(x) \cdot \mu_1 d\Gamma(x) - \sum_{[a,b]\in\mathcal{I}} \int_{[a,b]} V(a|p,\tau) \cdot \mu_1 d\Gamma(x) \\
&= \int_{(X_r,H]} Z_e(x) \cdot \mu_1 d\Gamma(x) - \sum_{[a,b]\in\mathcal{I}} V(a|p,\tau)\mu([a,b]) \\
&= \int_{(X_r,H]} Z_e(x) \cdot \mu_1 d\Gamma(x) - \sum_{[t(a),t(b)]\in\mathcal{I}_t} Z(t(a))\mu([t(a),t(b)]) \\
&= \int_{\mathcal{K}_t^c} Z(x) \cdot \mu_1 d\Gamma(x),
\end{aligned}$$

which proves Eq. (A.45). To conclude, we provide a proof for both Property 1 and Property 2.

Property 1. (p^{I_t}, τ^{I_t}) forms a price-equilibrium pair in I_t such that $\tau^{I_t} \in \mathcal{F}_{I_t}(\mu|_{I_t})$. Also, $V(x|\, p^{I_t}, \tau^{I_t})$ equals $V(x + a - t(a)|\, p, \tau)$ for all $x \in I_t$, and

$$\frac{d\tau_2^{I_t}}{d\Gamma}(x) = \frac{d\tau_2}{d\Gamma}(x + a - t(a)), \quad \Gamma - a.e.\ x \text{ in } I_t.$$

Proof of Property 1. We first show that $\tau^{I_t} \in \mathcal{F}_{I_t}(\mu|_{I_t})$. It is clear that $\tau^{I_t} \in \mathcal{M}(I_t \times I_t)$, and that $\tau_2^{I_t} \ll \Gamma$. To see why $\tau_1^{I_t}$ coincides with μ_{I_t} consider a set $K \subset I_t$ then $\tau_1^{I_t}(K)$ equals

$$\begin{aligned}
\tau_1^{I_t}(K \times I_t) = \tau((K + a - t(a)) \times (I_t + a - t(a))) &= \tau((K + a - t(a)) \times [a,b]) \\
&= \tau((K + a - t(a)) \times \mathcal{C}) \\
&= \mu(K + a - t(a)) \\
&= \mu(K),
\end{aligned}$$

where the fourth line holds because the set $K + a - t(a)$ is contained in $[a,b]$, and we know there is no flow leaving this interval. Next, using a similar argument we show the property for $d\tau_2^{I_t}/d\Gamma$, let K be a measurable subset of I_t then

$$\begin{aligned}\int_K \frac{d\tau_2^{I_t}}{d\Gamma}(x)\, d\Gamma(x) &= \tau^{I_t}(I_t \times K)\\ &= \tau([a,b] \times (K + a - t(a)))\\ &= \int_{(K+a-t(a))} \frac{d\tau_2}{d\Gamma}(x)\, d\Gamma(x)\\ &= \int_K \frac{d\tau_2}{d\Gamma}(x + a - t(a))\, d\Gamma(x).\end{aligned}$$

Using this last property and the prices definition is easy to see that

$$\begin{aligned}V(x|\, p^{I_t}, \tau^{I_t}) &= \inf\{u \in \mathbb{R} : \Gamma(y \in I_t : U(y, p^{I_t}(y), \frac{d\tau_2^{I_t}}{d\Gamma}(y)) - |y - x| > u) = 0\}\\ &= \inf\{u \in \mathbb{R} : \Gamma(y \in I_t : U(y, p(y + a - t(a)), \frac{d\tau_2}{d\Gamma}(y + a - t(a)))\\ &- |y - x| > u) = 0\}\\ &= \inf\{u \in \mathbb{R} : \Gamma(y \in I : U(y, p(y), \frac{d\tau_2}{d\Gamma}(y))\\ &- |y - (x + a - t(a))| > u) = 0\}\\ &= V_I(x + a - t(a)|\, p, \tau),\end{aligned}$$

but from out flow separation result (see Lemma A.9) we have that $V_I(x + a - t(a)|\, p, \tau) = V(x + a - t(a)|\, p, \tau)$. Using this same approach, the definition of τ^{I_t} and the fact that τ is an equilibrium in $[a, b]$ it is easy to verify the equilibrium condition.

Property 2. The pair (p^{J_t}, τ^{J_t}) forms a price-equilibrium pair in J_t such that $\tau^{J_t} \in \mathcal{F}_{J_t}(\mu|_{J_t})$ and $V(x|\, p^{J_t}, \tau^{J_t})$ equals $Z(x)$ for all $x \in J_t$.

Proof of Property 2. From the definition of τ^{J_t} it is clear that $\tau^{J_t} \in \mathcal{F}_{J_t}(\mu|_{J_t})$. Also, $d\tau_2^{J_t}/d\Gamma = \mu_1$, $\Gamma - a.e$ in J_t. To see why $V(x|\, p^{J_t}, \tau^{J_t})$ equals $Z(x)$ for all $x \in J_t$, note that for fixed $x \in J_t$

$$\Gamma(y \in J_t : U(y, p^{J_t}(y), \frac{d\tau_2^{J_t}}{d\Gamma}(y)) - |y - x| > Z(x)) = \Gamma(y \in J_t : Z(y) - |x - y| > Z(x)) = 0,$$

where in the first equality we use the definition of $p^{J(t)}$ together with $d\tau_2^{J_t}/d\Gamma = \mu_1$, $\Gamma - a.e$ in J_t. In the second equality we use the Lipschitz property of the function

$Z(\cdot)$. That is, $Z(x) \geq V(x|\, p^{J_t}, \tau^{J_t})$. This upper bound ($\Gamma-$a.e) is tight. Let $\epsilon > 0$ then

$$\begin{aligned}
0 &< \Gamma(B(x, \epsilon/2) \cap J_t) \\
&\leq \Gamma(y \in B(x, \epsilon/2) \cap J_t) : \epsilon > |x - y| + (Z(x) - Z(y))) \\
&= \Gamma(y \in B(x, \epsilon/2) \cap J_t) : Z(y) - |y - x| > Z(x) - \epsilon) \\
&= \Gamma(y \in B(x, \epsilon/2) \cap J_t) : U(y, p^{J_t}(y), \frac{d\tau_2^{J_t}}{d\Gamma}(y)) - |y - x| > Z(x) - \epsilon),
\end{aligned}$$

thus $Z(x)$ is the smallest upper bound ($\Gamma-$a.e) and we have $Z(x) = V(x|\, p^{J_t}, \tau^{J_t})$. It is not hard to verify that the equilibrium condition reduces to

$$\tau^{J_t}((x, y) \in J_t \times J_t : Z(y) - |y - x| = Z(x)) = \mu(J_t),$$

and by the definition of τ^{J_t} this is immediately satisfied.

□

Proof of Theorem 1.2. The result follows directly from Proposition 1.9, and the fact that $[X_l, X_r]$ is an attraction region where $V(\cdot)$ is pinned down.

□

Proof of Theorem 1.3. We separate the proof in several steps. First, we argue that there are at most three attraction regions in the any optimal solution. Then we show that any optimal solution does not have drivers moving to the interval $[W_r, X_r]$ and $[X_l, W_l]$; otherwise, the platform can incentivize the movement of a positive fraction of drivers outside of the center and make strictly larger revenue. After this we put into practice Theorem 1.1 which prescribes what are the optimal prices and post-relocation supply in each attraction region. In the final main step of the proof we argue that the optimal solution has to be symmetric. We present the proof of two properties that we will use during the main arguments, Property 1 and Property 2, after the main proof.

Attraction regions identification: Lemma 1.5 establishes that at an optimal solution the attraction region of the origin is well defined with $X_l < 0 < X_r$. So Our first attraction region is the interval $[X_l, X_r]$.

The second and third attraction regions correspond to the intervals $[Y_l, X_l]$ and $[X_r, Y_r]$ with Y_l and Y_r being sink locations. WLOG consider only the right interval, if $Y_r = X_r$ we do not identify any attraction region to the right of X_r. Assume that $X_r < Y_r$, we will show that $A(Y_r) = [X_r, Y_r]$ and $Y_r \notin A(z)$ for any $z \neq Y_r$. In order, to show this we first show that $Y_r \in \mathcal{IR}(X_r|\, p, \tau)$. From Theorem 1.2 we know that $V(x)$ equals $V(X_r) + (x - X_r)$ for all $x \in [X_r, Y_r]$. Fix $\epsilon > 0$ and $\delta_0 \in (0, Y_r - X_r)$ then for any $\delta \leq \delta_0$ define the set

$$K^\delta \triangleq \{y \in B(Y_r, \delta) \cap [X_r, Y_r] : U(y) = V(y)\}.$$

Since $\mu((Y_r - \delta, Y_r]) > 0$ and $\tau((Y_r - \delta, Y_r] \times (\mathcal{C} \setminus (Y_r - \delta, Y_r])) = 0$ (otherwise we would obtain a contradiction with Theorem 1.2), we must have that $\tau_2((Y_r - \delta, Y_r]) > 0$. This together with Lemma A.2 and $\tau_2 \ll \Gamma$ imply that $\Gamma(K^\delta) > 0$. Hence,

$$\begin{aligned}
0 &< \Gamma(K^\delta) \\
&= \Gamma(y \in K^\delta : \epsilon > 0) \\
&= \Gamma(y \in K^\delta : V(X_r) > V(X_r) - \epsilon) \\
&= \Gamma(y \in K^\delta : V(y) - |y - X_r| > V(X_r) - \epsilon) \\
&= \Gamma(y \in K^\delta : U(y) - |y - X_r| > V(X_r) - \epsilon) \\
&\leq \Gamma(y \in B(Y_r, \delta) : U(y) - |y - X_r| > V(X_r) - \epsilon)
\end{aligned}$$

This implies that $V_{B(Y_r,\delta)}(X_r) \geq V(X_r) - \epsilon$. By the choice of ϵ and δ we conclude that $\lim_{\delta \downarrow 0} V_{B(Y_r,\delta)}(X_r)$ is $V(X_r)$. In other words, $Y_r \in \mathcal{IR}(X_r|\, p, \tau)$. Now, Y_r cannot belong to any other attraction region; otherwise, by the Lemma 1.3 the value function would not be as in Theorem 1.2. Therefore, Y_r is a sink location and $[X_r, Y_r] \subseteq A(Y_r)$. If there existed $x \in A(Y_r)$ but $x \notin [X_r, Y_r]$, the value function would not be as in Theorem 1.2. In conclusion, $A(Y_r) = [X_r, Y_r]$ and $Y_r \notin A(z)$ for any $z \neq Y_r$.

No supply in $[W_r, X_r]$**:** Next we argue that at an optimal solution (p, τ) we must have that $\tau_2([W_r, X_r]) = 0$, the same is true for the left side. Suppose by contradiction that $\tau_2([W_r, X_r]) > 0$ and denote this amount of supply by q_r, we construct a new solution $(\tilde{p}, \tilde{\tau})$ that yields an strict objective improvement. Observe that,

$$0 < q_r = \tau(\mathcal{C} \times [W_r, X_r]) = \tau([W_r, X_r] \times [W_r, X_r]) \leq \mu([W_r, X_r]) = \mu_1 \cdot (X_r - W_r).$$

That is, from the total amount of initial supply in $[W_r, X_r]$ we have that q_r units stay within $[W_r, X_r]$ and a total of $\mu_1 \cdot (X_r - W_r) - q_r$ units travel to $[0, W_r]$. Note that for this q_r units of mass their V is bounded by ψ_1 and, therefore, what the platform can make from them is strictly bounded by $\psi_1 \cdot q_r$ (times a scaling factor). Let $\tilde{X}_r \in [W_r, X_r)$ be such that $q_r = \mu_1 \cdot (X_r - \tilde{X}_r)$. In the new solution, we will modify the attraction region $[X_l, X_r]$ to be $[X_l, \tilde{X}_r]$. We will maintain the same prices and post-relocation supply in the origin's attraction region. However, to the right side of $\tilde{X}_r$ we will set new prices that will be consistent with a new value function and flows that upper bound those of the old solution, see Figure A.3.

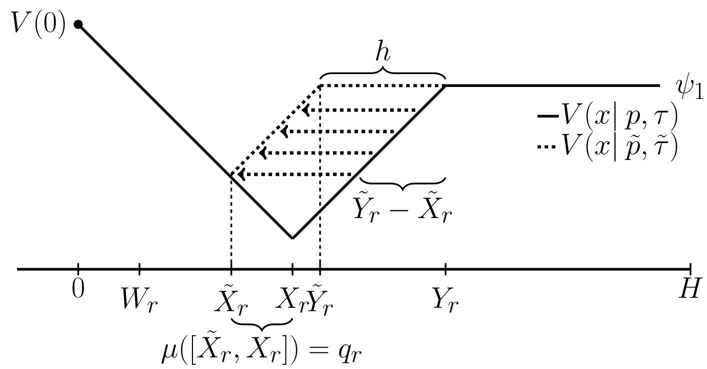


Figure A.3: **No supply in** $[W_r, X_r]$. The new solution moves the right end of the attraction region from X_r to $\tilde{X}_r$, so now a mass q_r of drivers can travel towards the periphery. From this mass the platform now makes ψ_1 instead of $V(x)$ with $V(x) < \psi_1$.

We begin our construction of $(\tilde{p}, \tilde{\tau})$ with the interval $I_r^1 = [\tilde{X}_r, \tilde{Y}_r]$, where $\tilde{Y}_r$ is such that $\psi_1 = V(\tilde{X}_r) + (\tilde{Y}_r - \tilde{X}_r)$. Let $h \triangleq 2 \cdot (X_r - \tilde{X}_r)$, we define flows for any

$\mathcal{L} \subseteq I_r^1 \times I_r^1$ by

$$\tau^{I_r^1}(\mathcal{L}) = \tau\Big(\mathcal{L} + (h,h)\Big).$$

Consider the set $K \triangleq \{x \in I_r^1 : \frac{d\tau_2^{I_r^1}}{d\Gamma}(x) \leq \frac{d\tau_2}{d\Gamma}(x+h)\}$. We set prices to be such that

$$U\Big(x, p^{I_r^1}(x), \frac{d\tau_2^{I_r^1}}{d\Gamma}(x)\Big) = U\Big(x+h, p(x+h), s^{\tau}(x+h)\Big), \quad \forall x \in K, \tag{A.47}$$

and zero otherwise. We prove, in Property 1 (see end of present proof), that $(p^{I_r^1}, \tau^{I_r^1})$ is a price-equilibrium pair in I_r^1 such that $V(x|\, p^{I_r^1}, \tau^{I_r^1}) = V(\tilde{X}_r) + (x - \tilde{X}_r)$ and $\Gamma(K^c) = 0$.

In the interval $I_r^2 = (\tilde{Y}_r, H]$ we can achieve the optimal solution when there is no demand shock. As in the optimal solution in the pre-demand shock environment (see Proposition 1.6) we set prices equal to ρ_1 and the flows are such that $d\tau^{I_r^2}/d\Gamma$ equals μ_1, $\Gamma - a.e$ in I_r^2.

The interval $I_r^0 = [X_l, \tilde{X}_r]$ is more involved. Observe that all the initial flow to the right of the origin that we have to allocate in $[0, \tilde{X}_r]$ equals $\mu_1 \cdot X_r - q_r$. This is exactly the same amount of drivers in $[0, X_r]$ that travels to $[0, W_r]$ according to τ. Our new solution will generate the same post-relocation supply than τ in $[0, W_r]$ but this time only using drivers from $[0, \tilde{X}_r]$.

We use the same prices, that is $p^{I_r^0}(x) = p(x)$ for all $x \in [X_l, \tilde{X}_r]$. For the flows we define them through two measures: the flow that goes from $[X_l, 0]$ to $[X_l, 0]$ and the flow that goes from $[0, \tilde{X}_r]$ to $[0, \tilde{X}_r]$. For the first flow we use $\tau^{\ell} = \tau|_{[X_l,0]}$, for the second measure τ^r we will use a monotone coupling (see e.g, [60] for details). Define the initial supply to the right measure μ^r to be equal to $\mu|_{[0,\tilde{X}_r]}$, and the final supply S^r to be

$$S^r(\mathcal{B}) \triangleq \tau([0, X_r] \times \mathcal{B}), \qquad \text{for any measurable set } \mathcal{B} \subseteq [0, \tilde{X}_r].$$

Note that $S^r([0, W_r])$ equals $\mu^r([0, \tilde{X}_r])$. Given this we define τ^r by

$$\tau^r(\mathcal{L}) \triangleq (F_{\mu^r}^{[-1]}, F_{S^r}^{[-1]})_{\#} m(\mathcal{L}), \quad \text{for any measurable set } \mathcal{L} \subseteq [0, \tilde{X}_r] \times [0, \tilde{X}_r],$$

where # correspond to the push-forward operator. For any measure ν defined in $[0, \tilde{X}_r]$ we define its cumulative function and pseudo-inverse by

$$F_\nu(y) \triangleq \nu([0, y]), \quad \forall\, y \geq 0 \quad \text{and} \quad F_\nu^{[-1]}(t) \triangleq \inf\{y \geq 0 : F_\nu(y) \geq t\},$$

$\forall t \in [0, \mu^r([0, \tilde{X}_r])]$. Effectively, τ^r transports the initial mass in $[0, \tilde{X}_r]$ to the final supply distribution (considering only drivers that come from the right) in $[0, W_r]$ as prescribed by τ. The final flow measure $\tau^{I_r^0}$ correspond to $\tau^\ell + \tau^r|_{[0,\tilde{X}_r]}$. In Property 2 below we show that $(p^{I_r^0}, \tau^{I_r^0})$ is a price-equilibrium pair such that $\mathbf{Rev}_{[X_l, W_r]}(p^{I_r^0}, \tau^{I_r^0}) = \mathbf{Rev}_{[X_l, W_r]}(p, \tau)$.

The solution $(\tilde{p}, \tilde{\tau})$ is constructed by pasting (see Lemma A.9) the old solution is $[-H, X_l)$ with the new solution in I^0, I_r^1 and I_r^2. The pasting is possible because the equilibrium utility function coincide in the boundaries of these intervals. This new solution preserves the platform's revenue in $[-H, W_r] \cup [Y_r, H]$ but it strictly improves it in $[W_r, Y_r]$. Indeed, note that

$$\begin{aligned} q_r &= \int_{[X_r, Y_r]} s^\tau(x)\, dx - \int_{[\tilde{X}_r, \tilde{Y}_r]} s^{\tilde{\tau}}(x)\, dx \\ &= \int_{[\tilde{X}_r, \tilde{Y}_r]} \underbrace{(s^\tau(x+h) - s^{\tilde{\tau}}(x))}_{\geq 0 \ \Gamma - a.e} \, dx + \int_{[X_r, X_r + (X_r - \tilde{X}_r)]} s^\tau(x)\, dx, \end{aligned} \tag{A.48}$$

thus

$$\begin{aligned}
\frac{1}{\gamma} \cdot \mathbf{Rev}_{[W_r,Y_r]}(\tilde{p},\tilde{\tau}) &= \int_{[W_r,\tilde{X}_r]} V(x|\tilde{p},\tilde{\tau}) \cdot s^{\tilde{\tau}}(x)\,dx + \int_{[\tilde{X}_r,Y_r]} V(x|\tilde{p},\tilde{\tau}) \cdot s^{\tilde{\tau}}(x)\,dx \\
&\overset{(a)}{=} \int_{[\tilde{X}_r,Y_r]} V(x|\tilde{p},\tilde{\tau}) \cdot s^{\tilde{\tau}}(x)\,dx \\
&\overset{(b)}{=} \int_{[\tilde{X}_r,\tilde{Y}_r]} V(x|\tilde{p},\tilde{\tau}) \cdot s^{\tilde{\tau}}(x)\,dx + \psi_1 \cdot 2 \cdot q_r \\
&\overset{(c)}{>} \int_{[\tilde{X}_r,\tilde{Y}_r]} V(x|\tilde{p},\tilde{\tau}) \cdot s^{\tilde{\tau}}(x)\,dx + \psi_1 \cdot q_r + \int_{[W_r,X_r]} V(x) \cdot s^{\tau}(x)\,dx \\
&\overset{(d)}{\geq} \int_{[\tilde{X}_r,\tilde{Y}_r]} V(x|\tilde{p},\tilde{\tau}) \cdot s^{\tau}(x+h)\,dx \\
&+ \int_{[W_r,X_r+(X_r-\tilde{X}_r)]} V(x) \cdot s^{\tau}(x)\,dx \\
&\overset{(e)}{=} \int_{[W_r,Y_r]} V(x) \cdot s^{\tau}(x)\,dx = \frac{1}{\gamma} \cdot \mathbf{Rev}_{[W_r,Y_r]}(p,\tau),
\end{aligned}$$

where (a) follows because $\tilde{\tau}$ does not put mass in $[W_r, \tilde{X}_r]$, (b) because $Y_r - \tilde{Y}_r$ equals $2 \cdot (X_r - \tilde{X}_r)$. Using the fact that $\tau_2([W_r, X_r]) = q_r$ we obtain (c), while (d) follows from Eq. (A.48) and (e) from $V(x|\,\tilde{p},\tilde{\tau})$ being equal to $V(x+h)$ for all $x \in [\tilde{X}_r, \tilde{Y}_r]$.

In conclusion, any optimal solution both $\tau_2([W_r, X_r])$ and $\tau_2([X_l, W_l])$ must equal zero.

Using Theorem 1.1: All the conditions in Theorem 1.1 are met. So, for any of the three attraction regions if (p, τ) is not already as in the statement of the theorem we can find at least a weak improvement. That is, we can restrict to solution as in Theorem 1.1. Therefore, the prices are as stated in the present theorem, and there exists $\beta_c^l \in [W_l, 0]$, $\beta_c^r \in [0, W_r]$, $\beta_p^l \in [Y_l, X_l]$ and $\beta_p^r \in [X_r, Y_r]$ such that

$$s^{\tau}(x) = \begin{cases} 0 & \text{if } x \in (\beta_c^r, \beta_p^r) \cup (\beta_p^l, \beta_c^l), \\ \psi_x^{-1}(V(x|\,p,\tau)) & \text{otherwise,} \end{cases}$$

with

$$\int_{\beta_c^l}^{\beta_c^r} \psi_x^{-1}(V(x|\,p,\tau))\,d\Gamma(x) = \mu_1 \cdot (X_r - X_l)$$

and

$$\int_{\beta_p^r}^{Y_r} \psi_x^{-1}(V(x|p,\tau))\,d\Gamma(x) = \mu_1\cdot(Y_r - X_r), \quad \int_{Y_l}^{\beta_p^l} \psi_x^{-1}(V(x|p,\tau))\,d\Gamma(x) = \mu_1\cdot(X_l - Y_l).$$

Note that the fact that $\beta_c^l \in [W_l, 0]$ and $\beta_c^r \in [0, W_r]$, does not come directly from Theorem 1.1 but rather is a consequence of that any optimal solution must satisfy both $\tau_2([W_r, X_r]) = 0$ and $\tau_2([X_l, W_l]) = 0$. Also, observe that Theorem 1.1 only gives us a solution in each attraction but above we have stated the solution for the entire city. The only missing interval are $[-H, Y_l]$ and $[Y_r, H]$. In this intervals, as in the pre-shock environment, the solution set prices equal to ρ_1 and the supply at every location is μ_1, in turn, the V equals ψ_1 in this region. This gives a complete solution to the platform's problem up to three values: $V(0), X_l, X_r$.

Symmetry: In the last main step of the proof we argue that the solution is symmetric. After proving this, the solution will take the exact form in the statement of the present theorem.

Note that given a value for $V(0)$ and an central attraction region characterize by X_l and X_r we can characterize the optimal solution as we did in **Using Theorem 1.1**. So fix these three values and the optimal solution associated to them. We now proceed to construct a new solution that yields a strict objective improvement when the solution is not symmetric. WLOG assume that $|X_l| > X_r$ and let $\delta = (|X_l| - X_r)/2$. Consider the solution $(\tilde{p}, \tilde{\tau})$ associated to the values

$$\tilde{V}(0) = V(0), \quad \tilde{X}_l = X_l + \delta, \quad \tilde{X}_r = X_r + \delta.$$

Note that with this values we have $|\tilde{X}_l|, \tilde{W}_r \geq W_r$ and $\tilde{Y}_i = Y_i + 2\cdot\delta$ for $i \in \{l, r\}$. We next show that this new solution yields a weak objective improvement in the center, and a strict objective improvement in the periphery.

Note that given $\tilde{V}(0), \tilde{X}_l$ and $\tilde{X}_r$ Theorem 1.2 characterizes $V(\cdot|\,\tilde{p}, \tilde{\tau})$. It has the same shape than $V(\cdot|\,p, \tau)$ except that now the dip in $[\tilde{Y}_l, W_l]$ is smaller, while the dip in $[W_r, Y_r]$ is larger. See Figure A.4 for a graphical representation. Consider

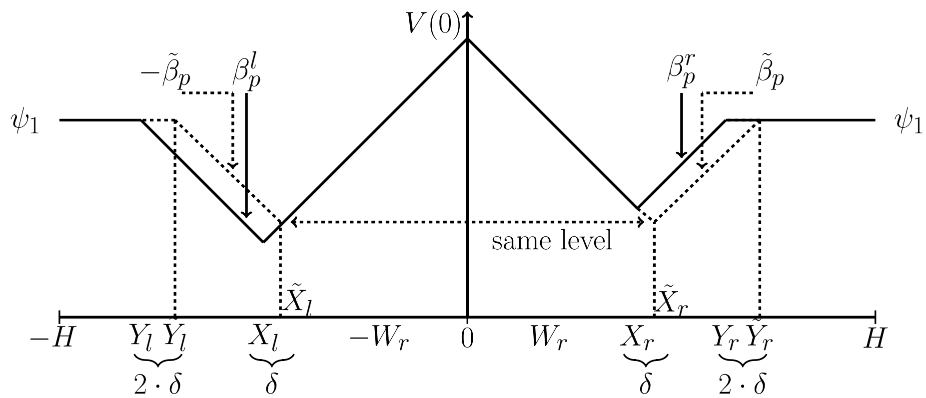


Figure A.4: **Symmetry** argument.

first the solution in the center, $[\tilde{X}_l, \tilde{X}_r]$. This interval contains the same amount of drivers that the old attraction region. The difference is that it lost a mass of $\mu_1 \cdot \delta$ drivers to the left and gain the same mass to the right. As in the discussion that follows Theorem 1.1 the optimal solution in $[\tilde{X}_l, \tilde{X}_r]$ can be obtained using a knapsack argument. This new attraction region is symmetric, $|\tilde{X}_l| = \tilde{X}_r$, with equal mass of drivers at both sides of the origin. Therefore the knapsack solution must be symmetric, with $\tilde{\beta}_c \in [0, W_r]$ such that

$$s^{\tilde{\tau}}(x) = \psi_x^{-1}(V(x|\, \tilde{p}, \tilde{\tau})) = \psi_x^{-1}(V(x|\, p, \tau)), \quad \forall x \in [-\tilde{\beta}_c, \tilde{\beta}_c],$$

and equals zero otherwise, and

$$\int_{-\tilde{\beta}_c}^{\tilde{\beta}_c} \psi_x^{-1}(V(x|\, \tilde{p}, \tilde{\tau})\, d\Gamma(x) = \mu_1 \cdot (\tilde{X}_r - \tilde{X}_l) = \mu_1 \cdot (X_r - X_l).$$

Note that $\tilde{\beta}_c \in [0, W_r]$ is a consequence of the having $\beta_c^l \in [W_l, 0]$ and $\beta_c^r \in [0, W_r]$ in the old solution. Theorem 1.1 prescribes how to formally implement this solution through prices and flows. We omit the details of how to construct the flows, but we note that the optimal prices are given $\tilde{p}(x) = \rho_x^{loc}(s^{\tilde{\tau}}(x))$. In the case that $\tilde{\beta} = 0$ then $s^{\tilde{\tau}}(0) = \mu_1 \cdot (X_r - X_l)$ and $\tilde{p}(0)$ is such that $U(0, p(0), s^{\tilde{\tau}}(0)) = V(0)$. The platform's revenue in the new center is then

$$\frac{1}{\gamma} \cdot \mathbf{Rev}_{[\tilde{X}_l, \tilde{X}_r]}(\tilde{p}, \tilde{\tau}) = \int_{\tilde{X}_l}^{\tilde{X}_r} V(x|\, \tilde{p}, \tilde{\tau}) \cdot s^{\tilde{\tau}}(x)\, dx = \int_{-\tilde{\beta}_c}^{\tilde{\beta}_c} V(x) \cdot \psi_x^{-1}(V(x))\, dx.$$

This expression is an upper bound for the platform's revenue under (p,τ) in $[X_l, X_r]$. In fact, WLOG assume $\beta_c^r \geq |\beta_c^l|$ which implies that $\tilde{\beta}_c \in [|\beta_c^l|, \beta_c^r]$ and we must have

$$\begin{aligned}
\frac{1}{\gamma} \cdot \mathbf{Rev}_{[X_l,X_r]}(p,\tau) &= \int_{\beta_c^l}^{\beta_c^r} V(x) \cdot \psi_x^{-1}(V(x))\, dx \\
&= \int_{\beta_c^l}^{|\beta_c^l|} V(x) \cdot \psi_x^{-1}(V(x))\, dx + \int_{|\beta_c^l|}^{\beta_c^r} V(x) \cdot \psi_x^{-1}(V(x))\, dx \\
&= \frac{1}{\gamma} \cdot \mathbf{Rev}_{[\tilde{X}_l,\tilde{X}_r]}(\tilde{p},\tilde{\tau}) \\
&- 2 \cdot \int_{|\beta_c^l|}^{\tilde{\beta}_c} V(x) \cdot \psi_x^{-1}(V(x))\, dx + \int_{|\beta_c^l|}^{\beta_c^r} V(x) \cdot \psi_x^{-1}(V(x))\, dx \\
&= \frac{1}{\gamma} \cdot \mathbf{Rev}_{[\tilde{X}_l,\tilde{X}_r]}(\tilde{p},\tilde{\tau}) \\
&- \int_{|\beta_c^l|}^{\tilde{\beta}_c} V(x) \cdot \psi_x^{-1}(V(x))\, dx + \int_{\tilde{\beta}_c}^{\beta_c^r} V(x) \cdot \psi_x^{-1}(V(x))\, dx \\
&\leq \frac{1}{\gamma} \cdot \mathbf{Rev}_{[\tilde{X}_l,\tilde{X}_r]}(\tilde{p},\tilde{\tau}) \\
&+ V(\tilde{\beta}_c) \cdot \Big(- \int_{|\beta_c^l|}^{\tilde{\beta}_c} \psi_x^{-1}(V(x))\, dx + \int_{\tilde{\beta}_c}^{\beta_c^r} \psi_x^{-1}(V(x))\, dx \Big) \\
&= \frac{1}{\gamma} \cdot \mathbf{Rev}_{[\tilde{X}_l,\tilde{X}_r]}(\tilde{p},\tilde{\tau}).
\end{aligned}$$

That is, the new solution in the center is a weakly improvement over the old solution.

Now let us consider the periphery. Since $|\tilde{X}_l| = \tilde{X}_r$ both right and left periphery are symmetric. Thus the optimal solution as given by Theorem 1.1 is the symmetric at both sides. The post-relocation supply is characterize by $\tilde{\beta}_p \in [\tilde{X}_r, \tilde{Y}_r]$ such that

$$s^{\tilde{\tau}}(x) = \psi_x^{-1}(V(x|\,\tilde{p},\tilde{\tau})) = \psi_x^{-1}(V(X_r) + (x - X_r) - 2\cdot\delta), \quad \forall x \in [\tilde{\beta}_p, \tilde{Y}_r],$$

and equals zero otherwise, and

$$\int_{\tilde{\beta}_p}^{\tilde{Y}_r} \psi_x^{-1}(V(x|\,\tilde{p},\tilde{\tau})\, d\Gamma(x) = \mu_1 \cdot (\tilde{Y}_r - \tilde{X}_r) = \mu_1 \cdot (Y_r - X_r) + \mu_1 \cdot \delta.$$

The optimal prices are $\tilde{p}(x) = \rho_x^{loc}(s^{\tilde{\tau}}(x))$. As before we omit the characterization of the equilibrium flow as their existence is guaranteed by Theorem 1.1. The platforms revenue in the periphery is

$$\frac{1}{\gamma} \cdot \mathbf{Rev}_{[-H,\tilde{X}_l]\cup[\tilde{X}_r,H]}(\tilde{p},\tilde{\tau}) = 2 \cdot \int_{\tilde{\beta}_p}^{\tilde{Y}_r} V(x|\tilde{p},\tilde{\tau}) \cdot \psi^{-1}(V(x|\tilde{p},\tilde{\tau}))\, dx + 2\cdot\psi_1\cdot\mu_1\cdot(H - \tilde{Y}_r),$$

where we have dropped the subindex x from ψ_x^{-1} to stress the fact that in this part of the city this subindex does not change the congestion function. We need to compare this revenue with the revenue of the old solution in the periphery. Not that since $|X_l| > X_r$ we must have

$$Y_r - \beta_p^r < \tilde{Y}_r - \tilde{\beta}_p < \beta_p^l - Y_l.$$

Thus,

$$\begin{aligned}
\frac{1}{\gamma}\cdot\mathbf{Rev}_{[-H,X_l]\cup[X_r,H]}(p,\tau) &= \int_{Y_l}^{\beta_p^l} V(x)\cdot\psi^{-1}(V(x))\,dx + \int_{\beta_p^r}^{Y_r} V(x)\cdot\psi^{-1}(V(x))\,dx \\
&\quad + \psi_1\cdot\mu_1\cdot(H - Y_r + Y_l + H) \\
&= \int_{Y_l+(Y_r-\beta_p^r)}^{\beta_p^l} V(x)\cdot\psi^{-1}(V(x))\,dx + 2\cdot\int_{\beta_p^r}^{Y_r} V(x)\cdot\psi^{-1}(V(x))\,dx \\
&\quad + \psi_1\cdot\mu_1\cdot(H - Y_r + Y_l + H) \\
&= \int_{Y_l+(Y_r-\beta_p^r)}^{\beta_p^l} V(x)\cdot\psi^{-1}(V(x))\,dx + 2\cdot\int_{\beta_p^r+2\cdot\delta}^{\tilde{Y}_r} V(x|\,\tilde{p},\tilde{\tau})\cdot\psi^{-1}(V(x|\,\tilde{p},\tilde{\tau}))\,dx \\
&\quad + 2\cdot\psi_1\cdot\mu_1\cdot(H - \tilde{Y}_r) \\
&= \underbrace{\int_{Y_l+(Y_r-\beta_p^r)}^{\beta_p^l} V(x)\cdot\psi^{-1}(V(x))\,dx - 2\cdot\int_{\tilde{\beta}_p}^{\beta_p^r+2\cdot\delta} V(x|\,\tilde{p},\tilde{\tau})\cdot\psi^{-1}(V(x|\,\tilde{p},\tilde{\tau}))\,dx}_{(a)} \\
&\quad + \frac{1}{\gamma}\cdot\mathbf{Rev}_{[-H,\tilde{X}_l]\cup[\tilde{X}_r,H]}(\tilde{p},\tilde{\tau}),
\end{aligned}$$

So if we show that the term (a) is strictly negative we will be done. Not that

$$\begin{aligned}
(a) \quad &= \int_{Y_l+(Y_r-\beta_p^r)}^{\beta_p^l} V(x)\cdot\psi^{-1}(V(x))\,dx - 2\cdot\int_{Y_l+(Y_r-\beta_p^r)}^{Y_l+(\tilde{Y}_r-\tilde{\beta}_p)} V(x)\cdot\psi^{-1}(V(x))\,dx \\
&= \int_{Y_l+(\tilde{Y}_r-\tilde{\beta}_p)}^{\beta_p^l} V(x)\cdot\psi^{-1}(V(x))\,dx - \int_{Y_l+(Y_r-\beta_p^r)}^{Y_l+(\tilde{Y}_r-\tilde{\beta}_p)} V(x)\cdot\psi^{-1}(V(x))\,dx \\
&< V(Y_l+(\tilde{Y}_r-\tilde{\beta}_p))\cdot\Big(\int_{Y_l+(\tilde{Y}_r-\tilde{\beta}_p)}^{\beta_p^l} \psi^{-1}(V(x))\,dx - \int_{Y_l+(Y_r-\beta_p^r)}^{Y_l+(\tilde{Y}_r-\tilde{\beta}_p)} \psi^{-1}(V(x))\,dx\Big) \\
&= 0.
\end{aligned}$$

In conclusion, we have constructed a new symmetric solution that yields an strict revenue improvement over the old solution. Therefore, any optimal solution ought to be symmetric.

Property 1. $(p^{I_r^1}, \tau^{I_r^1})$ forms a price-equilibrium pair in I_r^1 such that $V(x|p^{I_r^1}, \tau^{I_r^1})$ equals $V(\tilde{X}_r) + (x - \tilde{X}_r)$ and $\Gamma(K^c) = 0$.

Proof of Property 1. We first show that $\tau^{I_r^1} \in \mathcal{F}_{I_r^1}(\mu|_{I_r^1})$. It is clear that $\tau^{I_r^1} \in \mathcal{M}(I_r^1 \times I_r^1)$, and that $\tau_2^{I_r^1} \ll \Gamma$. To see why $\tau_1^{I_r^1}$ coincides with $\mu_{I_r^1}$ consider a set $I \subset I_r^1$ then $\tau_1^{I_r^1}(K)$ equals

$$\begin{aligned}
\tau_1^{I_r^1}(K \times I_r^1) &= \tau((I+h) \times (I_r^1 + h)) \\
&= \tau((I+h) \times [\tilde{X}_r + h, Y_r]) \\
&= \tau((I+h) \times \mathcal{C}) \\
&= \mu(I+h) \\
&= \mu(I),
\end{aligned}$$

where the fourth line holds because the set $I + h$ is contain in $[\tilde{X}_r + h, Y_r]$, and we know there is no flow leaving this interval. Next, using a similar argument we show the property for $d\tau_2^{I_r^1}/d\Gamma$, let I be a measurable subset of I_r^1 then

$$\begin{aligned}
\int_I \frac{d\tau_2^{I_r^1}}{d\Gamma}(x)\, d\Gamma(x) = \tau^{I_r^1}(I_r^1 \times I) &= \tau([\tilde{X}_r + h, Y_r] \times (I+h)) \\
&\leq \tau([X_r, Y_r] \times (I+h)) \\
&= \int_{(I+h)} \frac{d\tau_2}{d\Gamma}(x)\, d\Gamma(x) \\
&= \int_I \frac{d\tau_2}{d\Gamma}(x+h)\, d\Gamma(x),
\end{aligned}$$

that is, $\Gamma(K^c) = 0$. As for the equilibrium utility function let $x \in [\tilde{X}_r, \tilde{Y}_r)$ we have

$$\begin{aligned}
V(x|\, p^{I_r^1}, \tau^{I_r^1}) &= \inf\{u \in \mathbb{R} : \Gamma(y \in I_r^1 : U(y, p^{I_r^1}(y), \frac{d\tau_2^{I_r^1}}{d\Gamma}(y)) - |y - x| > u) = 0\} \\
&= \inf\{u \in \mathbb{R} : \Gamma(y \in I_r^1 : U(y, p(y+h), \frac{d\tau_2}{d\Gamma}(y+h)) - |y - x| > u) = 0\} \\
&= \inf\{u \in \mathbb{R} : \Gamma(y \in [\tilde{X}_r + h, Y_r] : U(y, p(y), \frac{d\tau_2}{d\Gamma}(y)) \\
&\quad - |y - (x+h)| > u) = 0\} \\
&\leq V(x+h|\, p, \tau).
\end{aligned}$$

Actually this upper bound is tight. Indeed, Fix any $\epsilon > 0$ and consider $\delta > 0$ small enough such that $(x+h) \notin B(Y_r, \delta)$. We have $\tau_2(\{y \in B(y,\delta) \cap [\tilde{X}_r + h, Y_r] : U(y) = V(y)\}) > 0$ which implies that $\Gamma(\{y \in B(Y_r, \delta) \cap [\tilde{X}_r + h, Y_r] : U(y) = V(y)\}) > 0$ and, therefore,

$$\begin{aligned}
0 &< \Gamma(\{y \in B(Y_r,\delta) \cap [\tilde{X}_r + h, Y_r] : U(y) = V(y), \epsilon + y - (x+h) > |y - (x+h)|\}) \\
&= \Gamma(\{y \in B(Y_r,\delta) \cap [\tilde{X}_r + h, Y_r] : U(y) = V(y), U(y) - |y - (x+h)| \\
&> V(x+h) - \epsilon\}) \\
&\leq \Gamma(\{y \in [\tilde{X}_r + h, Y_r] : U(y) - |y - (x+h)| > V(x+h) - \epsilon\}) \\
&= \Gamma(\{y \in I_r^1 : U(y, p^{I_r^1}(y), \frac{d\tau_2^{I_r^1}}{d\Gamma}(y)) - |y - x| > V(x+h) - \epsilon\}),
\end{aligned}$$

therefore $V(x|\, p^{I_r^1}, \tau^{I_r^1})$ equals $V(x+h)$ for all $x \in [\tilde{X}_r, \tilde{Y}_r)$, and by continuity for all $x \in I_r^1$. Since $V(x+h)$ equals $V(\tilde{X}_r) + (x - \tilde{X}_r)$ we obtain the desired result.

Now we need to verify that this selection of prices and flows yields an equilibrium. That is, we need show that the set

$$\mathcal{E}_{I_r^1} = \left\{(x,y) \in I_r^1 \times I_r^1 : \Pi(x, y, p^{I_r^1}(y), \frac{d\tau_2^{I_r^1}}{d\Gamma}(y)) = V(x|\, p^{I_r^1}, \tau^{I_r^1})\right\},$$

has $\tau^{I_r^1}$ measure equal to $\mu(I_r^1)$. Observe that $\tau(\mathcal{E}_{I_r^1})$ equals

$$\tau\left(\left\{(x,y) \in [\tilde{X}_r + h, Y_r] \times [\tilde{X}_r + h, Y_r] : \Pi(x-h, y-h, p^{I_r^1}(y-h), \frac{d\tau_2^{I_r^1}}{d\Gamma}(y-h)) = V(x)\right\}\right),$$

using that $\Gamma(K^c) = 0$ and the way we chose the prices one can verify that this expression equals

$$\tau\Big(\Big\{(x,y) \in [\tilde{X}_r + h, Y_r] \times [\tilde{X}_r + h, Y_r] : \Pi(x,y,p(y),s^\tau(y)) = V(x|\, p,\tau)\Big\}\Big).$$

There is no τ flow of drivers leaving $[\tilde{X}_r + h, Y_r]$ so the fact that τ is an equilibrium flow implies that this last expression equals $\mu([\tilde{X}_r + h, Y_r])$, which equals $\mu(I_r^1)$.

Property 2. $(p^{I_r^0}, \tau^{I_r^0})$ is a price-equilibrium pair such that $\mathbf{Rev}_{[X_l,W_r]}(p^{I_r^0}, \tau^{I_r^0}) = \mathbf{Rev}_{[X_l,W_r]}(p,\tau)$.

Proof of Property 2. First a couple of observations, note that for any $y \in [0, \tilde{X}_r]$ and the set $[0, y]$ then

$$\begin{aligned}
\tau_1^r([0,y]) = \tau^r([0,y] \times [0,\tilde{X}_r]) &= m\Big(t \in [0, \mu^r([0,\tilde{X}_r])] : F_{\mu^r}^{[-1]}(t) \in [0,y]\Big) \\
&= m\Big(t \in [0, \mu^r([0,\tilde{X}_r])] : 0 \le t \le F_{\mu^r}(y)\Big) \\
&= F_{\mu^r}(y),
\end{aligned}$$

and the same argument holds for τ_2^r and S^r, this characterizes the first and second marginals of τ^r. Furthermore, it's not difficult to see that for $y_1, y_2 \in [0, \tilde{X}_r]$ we have

$$\tau^r([0,y_1]\times[0,y_2]) = m\Big(t \in [0,\mu^r([0,\tilde{X}_r])] : t \le F_{\mu^r}(y_1), t \le F_{S^r}(y_2)\Big) = F_{\mu^r}(y_1) \wedge F_{S^r}(y_2). \tag{A.49}$$

Next, we show that $\tau^{I_r^0} \in \mathcal{F}_{I_r^0}(\mu|_{I_r^0})$ is an equilibrium in I_r^0. In order to do so we first show that$\tau^{I_r^0} \in \mathcal{F}_{I_r^0}(\mu|_{I_r^0})$. Second, we compute the supply density of $\tau_2^{I_r^0}$ and corroborate they coincide with s^τ. Third, we compute $V_{I_r^0}(\cdot|\, p^{I_r^0}, \tau^{I_r^0})$ and verify is coincides with $V(\cdot|\, p,\tau)$ in I_r^0. Finally, we check the equilibrium condition.

Clearly $\tau^{I_r^0}$ is a non-negative measure in $I_r^0 \times I_r^0$ because is the sum of non-negative measures. Now we check that $\tau_1^{I_r^0} = \mu|_{I_r^0}$. Consider a measurable set $\mathcal{B} \subseteq I_r^0$ then

$$\begin{aligned}
\tau_1^{I_r^0}(\mathcal{B}) &= \tau((\mathcal{B} \cap [X_l, 0]) \times [X_l, 0]) + \tau^r((\mathcal{B} \cap [0, \tilde{X}_r]) \times [0, \tilde{X}_r]) \\
&= \tau((\mathcal{B} \cap [X_l, 0]) \times \mathcal{C}) + \mu^r(\mathcal{B} \cap [0, \tilde{X}_r]) \\
&= \mu(\mathcal{B} \cap [X_l, 0]) + \mu(\mathcal{B} \cap [0, \tilde{X}_r]) \\
&= \mu|_{I_r^0}(\mathcal{B})
\end{aligned}$$

and thus we also have $\tau_1^{I_r^0} \ll \Gamma$. For the second marginal of $\tau^{I_r^0}$ we have

$$\begin{aligned}
\tau_2^{I_r^0}(\mathcal{B}) &= \tau([X_l, 0] \times (\mathcal{B} \cap [X_l, 0])) + \tau^r([0, \tilde{X}_r] \times (\mathcal{B} \cap [0, \tilde{X}_r])) \\
&= \tau([X_l, 0] \times (\mathcal{B} \cap [X_l, 0])) + S^r(\mathcal{B} \cap [0, \tilde{X}_r]) \\
&= \tau([X_l, 0] \times (\mathcal{B} \cap [X_l, 0])) + \tau([0, X_r] \times (\mathcal{B} \cap [0, \tilde{X}_r])) \\
&= \tau_2(\mathcal{B} \cap [X_l, 0)) + \tau_2(\mathcal{B} \cap (0, \tilde{X}_r]) + \tau_2(\mathcal{B} \cap \{0\}) \\
&= \tau_2|_{I_r^0}(\mathcal{B}),
\end{aligned}$$

and thus $\tau_2^{I_r^0} \ll \Gamma$. We conclude that $\tau^{I_r^0} \in \mathcal{F}_{I_r^0}(\mu|_{I_r^0})$. From this we can also conclude that

$$\frac{d\tau_2^{I_r^0}}{d\Gamma}(x) = s^\tau(x), \quad \Gamma - a.e. \ x \ \text{ in } \ I_r^0.$$

Next we compute the equilibrium utilities. We show that $V(x|p^{I_r^0}, \tau^{I_r^0})$ equals $V(x|p, \tau)$ for all $x \in I_r^0$. Observe that $\Gamma - a.e.$ y in I_r^0 we have $U(y, p^{I_r^0}(y), s^{\tau^{I_r^0}}(y)) = U(y, p(y), s^\tau(y))$, and, therefore, $V(x|\, p, \tau) \geq V(x|\, p^{I_r^0}, \tau^{I_r^0})$. Using the same argument that we used for the proof of Property 1 we can argue that this upper bound is tight, that is, $V(x|\, p, \tau) = V(x|\, p^{I_r^0}, \tau^{I_r^0})$.

Now the equilibrium condition. Consider the equilibrium set

$$\mathcal{E}_{I_r^0} \triangleq \Big\{(x, y) \in I_r^0 \times I_r^0 : U(y, p^{I_r^0}(y), s^{\tau^{I_r^0}}(y)) - |y - x| = V(x|\, p^{I_r^0}, \tau^{I_r^0})\Big\},$$

we need to verify that $\tau^{I_r^0}(\mathcal{E}_{I_r^0})$ equals $\mu(I_r^0)$. First, for $\tau^l(\mathcal{E}_{I_r^0})$ we have

$$\begin{aligned}\tau^l(\mathcal{E}_{I_r^0}) &= \tau\Big(\Big\{(x,y)\in[X_l,0]\times[X_l,0] : U(y,p(y),s^\tau(y)) - |y-x| = V(x|\,p,\tau)\Big\}\Big)\\ &= \tau([X_l,0]\times[X_l,0])\\ &= \tau([X_l,0]\times\mathcal{C})\\ &= \mu([X_l,0])\end{aligned}$$

where we have used our choice of prices, the relation between $d\tau_2^{I_r^0}/d\Gamma$ and s^τ, and the fact that τ is an equilibrium flow that does not setnd flow out of $[X_l,0]$. For $\tau^r|_{[0,\tilde{X}_r]}$, note that its second marginal is S^r and, therefore, Lemma A.2 implies that

$$\tau^r|_{[0,\tilde{X}_r]}(\mathcal{E}_{I_r^0}) = \tau^r\Big(\Big\{(x,y)\in[0,\tilde{X}_r]\times[0,\tilde{X}_r] : V(y|\,p,\tau) - |y-x| = V(x|\,p,\tau)\Big\}\Big),$$

and because $V(z|\,p,\tau)$ equals $V(0)-z$ for any $z\in[0,\tilde{X}_r]$ we have

$$\begin{aligned}\tau^r|_{[0,\tilde{X}_r]}(\mathcal{E}_{I_r^0}) &= \tau^r\Big(\Big\{(x,y)\in[0,\tilde{X}_r]\times[0,\tilde{X}_r] : -y-|y-x| = -x\Big\}\Big)\\ &= \tau^r\Big(\Big\{(x,y)\in[0,\tilde{X}_r]\times[0,\tilde{X}_r] : x\geq y\Big\}\Big)\\ &= \mu^r([0,\tilde{X}_r]) - \tau^r\Big(\Big\{(x,y)\in[0,\tilde{X}_r]\times[0,\tilde{X}_r] : x< y\Big\}\Big),\end{aligned}$$

but

$$\begin{aligned}\tau^r\left(\left\{(x,y) \in [0,\tilde{X}_r] \times [0,\tilde{X}_r] : x < y\right\}\right) &\leq \sum_{q \in \cap [0,\tilde{X}_r]} \tau^r([0,q] \times (q,\tilde{X}_r]) \\ &= \sum_{q \in \mathbb{Q} \cap [0,\tilde{X}_r]} \Big\{\tau^r([0,q] \times [0,\tilde{X}_r]) \\ &- \tau^r([0,q] \times [0,q])\Big\} \\ &= \sum_{q \in \mathbb{Q} \cap [0,\tilde{X}_r]} \Big\{\mu^r([0,q]) \wedge S^r([0,\tilde{X}_r]) \\ &- \mu^r([0,q]) \wedge S^r([0,q])\Big\} \\ &= \sum_{q \in \mathbb{Q} \cap [0,\tilde{X}_r]} \Big\{\mu^r([0,q]) \wedge S^r([0,\tilde{X}_r]) \\ &- \mu^r([0,q]) \wedge S^r([0,q])\Big\} \\ &= 0,\end{aligned}$$

where in the last line we used that $\mu^r([0,q]) \leq S^r([0,q])$. Adding up $\tau^l(\mathcal{E}_{I_r^0})$ with $\tau^r|_{[0,\tilde{X}_r]}(\mathcal{E}_{I_r^0})$, yields that $\tau^{I_r^0}(\mathcal{E}_{I_r^0})$ equals $\mu(I_r^0)$, and the equilibrium condition is satisfied. Finally, the revenue condition in the statement of the Property is immediately satisfied as $d\tau_2^{I_r^0}/d\Gamma$ coincide with s^τ in I_0^r, and the same is true for the equilibrium utilities.

□

Appendix B

Spatial Capacity Planning

B.1 Proofs for Section 2.3.2

Proof of Lemma 2.1. Let x_0 be in the interior of $\mathcal{C}$, a bounded subset of $\mathbb{R}^2$ with area denoted by $|\mathcal{C}|$. It is enough to prove that the following limit exists

$$\lim_{k\to\infty} \sqrt{k} \cdot \mathbf{E}\left[\min_{i=1,\dots,k} \|X_i - x_0\|\right].$$

Let $Z_k \triangleq \min_{i=1,\dots,k} \|X_i - x_0\|$. First, note that since x_0 is in the interior of the bounded region we can always find a ball $B(x_0, \epsilon)$ that is contain in $\mathcal{C}$ (below we take ϵ small enough). From this and the fact that the points X_i are drawn uniformly at random in $\mathcal{C}$, we have the following lower and upper bounds for any $i = 1, \dots, k$

$$\frac{\pi \cdot (z \wedge \epsilon)^2}{|\mathcal{C}|} = \mathbf{P}[\|X_i - x_0\| \leq z \wedge \epsilon] \leq \mathbf{P}[\|X_i - x_0\| \leq z] \leq \frac{\pi \cdot z^2}{|\mathcal{C}|}.$$

Second, from these bounds and the fact that the points X_i are IID we deduce

$$\left(1 - \frac{\pi \cdot z^2}{|\mathcal{C}|}\right)^k \vee 0 \leq \mathbf{P}[Z_k > z] \leq \left(1 - \frac{\pi \cdot (z \wedge \epsilon)^2}{|\mathcal{C}|}\right)^k \vee 0.$$

This yields the following bound for $\mathbf{E}[Z_k]$

$$\int_0^{\sqrt{|\mathcal{C}|/\pi}} \left(1 - \frac{\pi \cdot z^2}{|\mathcal{C}|}\right)^k dz \leq \mathbf{E}[Z_k] \leq \int_0^{\epsilon} \left(1 - \frac{\pi \cdot z^2}{|\mathcal{C}|}\right)^k dz + \int_{\epsilon}^{R_{\mathcal{C}}} \left(1 - \frac{\pi \cdot \epsilon^2}{|\mathcal{C}|}\right)^k dz,$$

where $R_{\mathcal{C}} = \max_{x,y\in\mathcal{C}} \|x - y\|$ and we are assuming that $\epsilon < R_{\mathcal{C}}$. Note that

$$\lim_{k\to\infty} \sqrt{k} \cdot \int_{\epsilon}^{R_{\mathcal{C}}} \left(1 - \frac{\pi \cdot \epsilon^2}{|\mathcal{C}|}\right)^k dz = \lim_{k\to\infty} \sqrt{k} \cdot \left(1 - \frac{\pi \cdot \epsilon^2}{|\mathcal{C}|}\right)^k \cdot (R_{\mathcal{C}} - \epsilon) = 0,$$

where we are using that ϵ is small enough such that $\pi \cdot \epsilon^2/|\mathcal{C}| < 1$. Therefore, we have that

$$\sqrt{k}\cdot\int_0^{\epsilon}\Big(1-\frac{\pi\cdot z^2}{|\mathcal{C}|}\Big)^k dz \leq \sqrt{k}\cdot\mathbf{E}[Z_k] \leq \sqrt{k}\cdot\int_0^{\epsilon}\Big(1-\frac{\pi\cdot z^2}{|\mathcal{C}|}\Big)^k dz+\sqrt{k}\cdot\int_{\epsilon}^{R_{\mathcal{C}}}\Big(1-\frac{\pi\cdot \epsilon^2}{|\mathcal{C}|}\Big)^k dz,$$

where the last term on the RHS above converges to zero. To complete the proof note that

$$\lim_{k\to\infty}\sqrt{k}\cdot\int_0^{\epsilon}\Big(1-\frac{\pi\cdot z^2}{|\mathcal{C}|}\Big)^k dz = \sqrt{\frac{|\mathcal{C}|}{\pi}}\cdot\lim_{k\to\infty}\sqrt{k}\cdot\int_0^{\epsilon\cdot\sqrt{\pi/|\mathcal{C}|}}\Big(1-z^2\Big)^k dz \approx 0.886\cdot\sqrt{\frac{|\mathcal{C}|}{\pi}},$$

where in the last step we use that for any $0<\delta<1$ the limit as $k\uparrow\infty$ of $\sqrt{k}\int_0^{\delta}(1-z^2)^k dz$ is approximately 0.886. □

B.2 Proofs for Section 2.4

Proof of Theorem 2.1. We make use of Proposition B.1 which we state and prove after the proof of this theorem. We prove each statements in the theorem.

(i) First we show that $\overline{q}_n$ as given in the statement is always an stable equilibrium. We have that $\overline{q}_n = n + z_n^2$ with $z_n = \rho_n/(1-\rho_n)$. Any equilibrium solves $f_n(q)=0$, thus we just need to verify that

$$1+\frac{1}{z_n} = \frac{n}{\lambda_n\bar{s}} = \frac{1}{\rho_n},$$

which is clearly satisfied. To verify stability we proceed using the Lyapunov method. Let $V(q)=|q-\overline{q}_n|$, then $\dot{V}(q)=\mathrm{sgn}(q-\overline{q}_n)\cdot f_n(q)$. We need to verify that $\dot{V}(q)<0$ for $q\neq\overline{q}_n$ (for n large enough). By Proposition B.1 part *(i)*, if $q\in(\overline{q}_n,\overline{q}_n+\delta]$ we have that $\dot{V}(q)=f_n(q)<0$, and if $\in[\overline{q}_n-\delta,\overline{q}_n)$ $\dot{V}(q)=-f_n(q)<0$ for $\delta>0$ small enough. Hence, $\overline{q}_n$ is a locally asymptotically stable equilibrium.

If $\alpha>1/3$ or if $\alpha=1/3$ and $\beta<\beta_1^*$ by Proposition B.1 we have that $f_n(q)>0$ for all $q\in[0,\overline{q}_n)$. Therefore the same Lyapunov analysis as before leads to the conclusion that $\overline{q}_n$ is a globally asymptotically stable equilibrium.

(ii) Both equilibria $\underline{q}$ and $\tilde{q}_n$ can be found by equating $g_{1,n}(q)$ and $g_{2,n}(q)$. This turns out to be equivalent to solving the equation

$$(n-q)+\frac{n\cdot\rho_n}{\sqrt{n-q}}=n\cdot(1-\rho_n). \tag{B.1}$$

For the current values of α and β, Proposition B.1 part *(iii)*, we know the latter equation has two solutions: $\tilde{q}_n$ and $\underline{q}$. Let's start with $\tilde{q}_n$. From Proposition B.1 we know that in a vicinity to the left of $\tilde{q}_n$ we have $f_n(q)<0$, that is, in a vicinity to the left of $\tilde{q}_n$ we have $d\tilde{Q}_n(t)/dt<0$ and, therefore, the systems moves away from $\tilde{q}_n$. Similarly, in a vicinity to the right of $\tilde{q}_n$ we have $f_n(q)>0$ and, therefore, the system moves away from $\tilde{q}_n$. This shows that this equilibrium is unstable.

For $\underline{q}$ we can use the same Lyapunov analysis as before, together with Proposition B.1, to show that it is a locally asymptotically stable equilibrium.

To conclude we need to provide a closed form characterization the two equilibria. We transform the equation that defines them, Eq. (B.1), in to a cubic equation. Consider the change of variables $w=\sqrt{n-q}$, then the equation becomes

$$w^3-n\cdot(1-\rho_n)\cdot w+n\cdot\rho_n=0. \tag{B.2}$$

The solution to this equation can be found in [62]. When the term $-4n^3\cdot(1-\rho_n)^3+27n^2\cdot\rho_n^2$ is non-positive the three possible solutions to (B.2) are real and given by

$$w_i=2\sqrt{\frac{n\cdot(1-\rho_n)}{3}}\cdot\cos\Big(\frac{1}{3}\arccos\Big(-\sqrt{\frac{27\rho_n^2}{4n\cdot(1-\rho_n)^3}}\Big)-\frac{2\pi i}{3}\Big),\quad i=0,1,2.$$

In order to verify that $-4n^3\cdot(1-\rho_n)^3+27n^2\cdot\rho_n^2\le 0$, note that this is equivalent to $27\rho_n^2\le 4n^{1-3\alpha}\cdot(n^\alpha(1-\rho_n))^3$. For large n, this last inequality holds for $\alpha<1/3$. The same is true for $\alpha=1/3$ and $\beta>\beta_1^*$. Therefore, the solutions w_k are all real. Furthermore, it is possible to verify that they are ordered, $w_0\ge w_1\ge w_2$, and that w_2 satisfies

$$w_2=-2\sqrt{\frac{n\cdot(1-\rho_n)}{3}}\cdot\cos\Big(\frac{1}{3}\arccos\Big(\sqrt{\frac{27\rho_n^2}{4n\cdot(1-\rho_n)^3}}\Big)\Big)<0,$$

and $w_1 \geq 0$ for large n. Since we are using the change of variables $w = \sqrt{n-q}$, we can disregard w_2 as a solution and take w_0 and w_1 to compute the solutions of our original equation. Because $\underline{q} \leq \tilde{q}_n$ we obtain that $\underline{q} = n - w_0^2$ and $\tilde{q}_n = n - w_1^2$.

□

Proposition B.1 *Suppose* $\lim_{n\to\infty}(1-\rho_n)n^{\alpha} = \beta$ *and that* $\rho_n \uparrow 1$. *Let* $\beta_1^* = 3/4^{1/3}$ *then*

(i) there exists n_0 *such that for all* $n \geq n_0$ *there exists* $\overline{q}_n > n$ *for which*

$$f_n(q) \begin{cases} = 0 & \text{if } q = \overline{q}_n \\ < 0 & \text{if } q > \overline{q}_n \\ > 0 & \text{if } q \in [n, \overline{q}_n). \end{cases}$$

(ii) if $\alpha > 1/3$, *or if* $\alpha = 1/3$ *and* $\beta < \beta_1^*$, *there exists* n_0 *such that for all* $n \geq n_0$ *we have* $f_n(q) > 0$ *for all* $q \in [0, \overline{q}_n)$.

(iii) if $\alpha < 1/3$, *or if* $\alpha = 1/3$ *and* $\beta > \beta_1^*$ *then there exists* n_0 *such that for all* $n \geq n_0$ *there exist* $\underline{q}$ *and* $\tilde{q}_n$ *with* $0 \leq \underline{q} < n - (\frac{n\cdot\rho_n}{2})^{2/3} < \tilde{q}_n < n-1$ *such that*

$$f_n(q) \begin{cases} = 0 & \text{if } q \in \{\underline{q}, \tilde{q}_n\} \\ < 0 & \text{if } q \in (\underline{q}, \tilde{q}_n) \\ > 0 & \text{if } q \in [0, \underline{q}) \cup (\tilde{q}_n, \overline{q}_n). \end{cases}$$

Proof of Proposition B.1. First note that from the definition of f_n we have

$$f_n(q) = \lambda_n - \frac{1}{\frac{\bar{s}}{\sqrt{|q-n| \vee 1}} + \bar{s}} \cdot \min(n, q). \tag{B.3}$$

Next prove each part of the statement separately.

(i) Consider $q \geq n+1$ then $f_n(q) = 0$ if and only if

$$\left(1 + \frac{1}{\sqrt{q-n}}\right) = \frac{1}{\rho_n}.$$

The left hand side is a decreasing function of q with maximum value equal to 2 for $q \geq n+1$. Also, since $\rho_n < 1$ we have that $1/\rho_n > 1$. If n is large enough so that $1/\rho_n < 2$, we can always find a solution $\overline{q}_n > n$ such that $f_n(\overline{q}_n) = 0$. Moreover, $f_n(\overline{q}_n) < 0$ for $q > \overline{q}_n$, and $f_n(\overline{q}_n) > 0$ for $q \in [0, \overline{q}_n)$.

(ii) First suppose that $q \in [n, \overline{q}_n)$, from what we did in the proof of *(i)* we can conclude that $f_n(q) > 0$ for n large enough. For $q \in [n-1, n)$, $f_n(q) > 0$ if and only if $2 > q/(n\rho_n)$. Since $\rho_n \uparrow 1$ and q is at most n this last inequality holds for all n large enough.

Next, suppose that $q < n-1$. Note that $f_n(q) > 0$ if and only if

$$\Big(1 + \frac{1}{\sqrt{n-q}}\Big) > \frac{q}{n\rho_n}.$$

We can rewrite the previous equation in the following equivalent form

$$\underbrace{x_n + \frac{n \cdot \rho_n}{\sqrt{x_n}}}_{g_n(x_n)} > n \cdot (1-\rho_n),$$

where $x_n = n - q$. Hence, $f_n(q) > 0$ if and only if $g_n(x_n) > n \cdot (1-\rho)$. Note that

$$\frac{dg_n(x)}{dx} = 1 - \frac{n \cdot \rho}{2x^{3/2}}, \quad \text{and} \quad \frac{d^2 g_n(x)}{dx^2} = \frac{3n \cdot \rho}{4x^{5/2}}.$$

Hence, $g_n(x)$ is a convex function with minimum at $x_n^* = (\frac{n \cdot \rho_n}{2})^{2/3}$. Thus, whenever $g_n(x_n^*) > n \cdot (1-\rho_n)$ we have that $f_n(q) > 0$. Observe that

$$g_n(x_n^*) > n \cdot (1-\rho_n) \Leftrightarrow (n \cdot \rho_n)^{2/3} \underbrace{(\frac{1}{2^{2/3}} + 2^{1/3})}_{\beta_1^*} > n \cdot (1-\rho_n),$$

which is equivalent to $\rho_n^{2/3} \beta_1^* > n^{1/3-\alpha} \cdot (1-\rho_n) n^{\alpha}$. If $\alpha > 1/3$ then, because $(1-\rho_n) \cdot n^{1/3} \to \beta$, the last inequality above holds for all n sufficiently large. If $\alpha = 1/3$ the last inequality above becomes $\rho_n^{2/3} \cdot \beta_1^* > (1-\rho_n) \cdot n^{1/3}$, and if $\beta < \beta_1^*$, since $(1-\rho_n) \cdot n^{1/3} \to \beta$ and $\rho_n \uparrow 1$, we would have $g_n(x_n^*) > n \cdot (1-\rho_n)$

for all n sufficiently large. Therefore in both cases we have that $f_n(q) > 0$ for all $q < n - 1$.

(iii) Similarly, we can argue that if $\alpha < 1/3$, or if $\alpha = 1/3$ and $\beta > \beta_1^*$ then $g_n(x_n^*) < n \cdot (1 - \rho_n)$ for n sufficiently large. When $g_n(x_n^*) < n \cdot (1 - \rho_n)$ the function $g_n(x)$ (recall this is a convex function) crosses $n \cdot (1 - \rho_n)$ at two points: $\underline{x}_{1,n}$ and $\overline{x}_{1,n}$, with $1 < \underline{x}_{1,n} < x_n^* < \overline{x}_{1,n} \leq n$. Defining $\underline{q} = n - \overline{x}_{1,n}$ and $\tilde{q}_n = n - \underline{x}_{1,n}$ we conclude the result.

□

B.3 Proofs for Section 2.5

Proof of Proposition 2.1. We make use of Eq. (2.9) and Proposition B.1.

(i) Note that from Proposition B.1 part *(i)* we have that $f_n(k) \geq 0$ for all $k \in [n, \overline{q}_n]$, since $\lfloor \overline{q}_n \rfloor \leq \overline{q}_n$ from Eq. (2.9) we deduce that $\pi_n(k)$ is increasing for all $k \in [n, \lfloor \overline{q}_n \rfloor] \cap \mathbb{N}$. Moreover, because $f_n(k) < 0$ for $k > \overline{q}_n$ and $\overline{q}_n < \lfloor \overline{q}_n \rfloor + 1$ from Eq. (2.9) we have that $\pi_n(k)$ decreases for all $k \in (\lfloor \overline{q}_n \rfloor, \infty) \cap \mathbb{N}$. Finally, using a similar argument and Proposition B.1 part *(ii)*, we deduce that $\pi_n(k)$ is increasing for all $k \in [0, n] \cap \mathbb{N}$.

(ii) Note that from Proposition B.1 part *(iii)* we have that $f_n(k) \geq 0$ for all $k \in [0, \underline{q}]$, $f_n(k) < 0$ for all $k \in (\underline{q}, \tilde{q}_n)$, and $f_n(k) \geq 0$ for all $k \in [\tilde{q}_n, \overline{q}_n)$. Eq. (2.9) then implies that $\pi_n(k)$ increases for $k \in [0, \lfloor \underline{q} \rfloor] \cap \mathbb{N}$, it decreases for $k \in (\lfloor \underline{q} \rfloor, \lfloor \tilde{q}_n \rfloor] \cap \mathbb{N}$, and it increases for $k \in (\lfloor \tilde{q}_n \rfloor, \lfloor \overline{q}_n \rfloor) \cap \mathbb{N}$.

□

Proof of Theorem 2.2. This result relies on Proposition 2.2 which is stated in the main text in the Proof sketch of Theorem 2.2 discussion. We provide a proof for Proposition 2.2 after the present proof.

We prove each statement in the theorem separately.

(i) We analyze different cases. First we consider $\alpha \in (1/3, 1)$. In this case from Proposition 2.1 part *(ii)* we now that $\pi_n(k) \leq \pi_n(n)$ for all k, for all n large enough. Moreover, from Proposition 2.2 part *(i)* we have that for $\epsilon \in (0, 1/\beta)$ for all n large enough the following inequality holds

$$\frac{\pi_n(n)}{\pi_n(\lfloor \overline{q}_n \rfloor)} \leq \exp\Big(- n^{\alpha}(\frac{1}{\beta} - \epsilon)\Big).$$

Therefore,

$$\begin{aligned}
\mathbf{P}[Q_n(\infty) < n] &= \sum_{k=0}^{n-1} \pi_n(k) \leq n \cdot \pi_n(n) \\
&= n \cdot \frac{\pi_n(n)}{\pi_n(\lfloor \overline{q}_n \rfloor)} \cdot \pi_n(\lfloor \overline{q}_n \rfloor) \\
&\leq n \cdot \exp\Big(- n^{\alpha}(\frac{1}{\beta} - \epsilon)\Big) \cdot \pi_n(\lfloor \overline{q}_n \rfloor) \to 0.
\end{aligned}$$

Next, consider $\alpha = 1/3$ and $\beta < \beta_2^*$. Let $\pi_n(k|\beta)$ be the steady-state probability when λ_n is such that $(1-\rho_n)n^{1/3} = \beta$. For notational clarity we use $\lambda_n(\beta), \overline{q}_n(\beta)$ and $\underline{q}(\beta)$ instead of $\lambda_n, \overline{q}_n$ and $\underline{q}$. It is possible to show that for $\beta < \beta'$ and n large enough we must have that

$$\frac{\pi_n(k|\beta)}{\pi_n(\lfloor \overline{q}_n(\beta') \rfloor | \beta)} \leq \frac{\pi_n(k|\beta')}{\pi_n(\lfloor \overline{q}_n(\beta') \rfloor | \beta')}, \quad \forall k \leq n-1. \tag{B.4}$$

Before we show Eq. (B.4), we will use to conclude this part of the proof. Fix $\beta < \beta_2^*$ then we can find $\beta' \in (\max\{\beta_1^*, \beta\}, \beta_2^*)$ for which Eq. (B.4) holds and, therefore, from Proposition 2.2 we can take $\epsilon \in (0, g(\beta'))$ such that for n large enough we have

$$\begin{aligned}
\mathbf{P}[Q_n(\infty) < n] &= \sum_{k=0}^{n-1} \pi_n(k|\beta) \\
&\leq \sum_{k=0}^{n-1} \frac{\pi_n(k|\beta)}{\pi_n(\lfloor \overline{q}_n(\beta') \rfloor | \beta)} \\
&\leq \sum_{k=0}^{n-1} \frac{\pi_n(k|\beta')}{\pi_n(\lfloor \overline{q}_n(\beta') \rfloor | \beta')} \\
&\leq n \cdot \frac{\pi_n(\lfloor \underline{q}(\beta') \rfloor | \beta')}{\pi_n(\lfloor \overline{q}_n(\beta') \rfloor | \beta')} \\
&= n \cdot \exp\Big(- n^{1/3}(g(\beta') - \epsilon)\Big) \to 0,
\end{aligned}$$

as $n \to \infty$. Next, we verify Eq. (B.4). Note that for $k < \lfloor \overline{q}_n(\beta') \rfloor$

$$\frac{\pi_n(\lfloor \overline{q}_n(\beta') \rfloor | \beta)}{\pi_n(k|\beta)} = \prod_{m=k+1}^{\lfloor \overline{q}_n(\beta') \rfloor} \frac{\lambda_n(\beta)\bar{s}}{\min\{m,n\}} \cdot \Big(1 + \frac{1}{\sqrt{|n-m| \vee 1}}\Big)$$

and

$$\frac{\pi_n(\lfloor \overline{q}_n(\beta') \rfloor | \beta')}{\pi_n(k|\beta')} = \prod_{m=k+1}^{\lfloor \overline{q}_n(\beta') \rfloor} \frac{\lambda_n(\beta')\bar{s}}{\min\{m,n\}} \cdot \Big(1 + \frac{1}{\sqrt{|n-m| \vee 1}}\Big).$$

Hence, Eq. (B.4) is satisfied if and only if

$$\lambda_n(\beta')^{\lfloor \overline{q}_n(\beta') \rfloor - k} \leq \lambda_n(\beta)^{\lfloor \overline{q}_n(\beta') \rfloor - k} \Leftrightarrow \lambda_n(\beta') \leq \lambda_n(\beta)$$

which is equivalent to

$$n^{1/3}\Big(1 - \frac{\lambda_n(\beta')\bar{s}}{n}\Big) \geq n^{1/3}\Big(1 - \frac{\lambda_n(\beta)\bar{s}}{n}\Big),$$

since both expression in the last inequality above converge to β' and β (respectively) and $\beta' > \beta$, we can always find n large enough so that the inequality is true. This shows Eq. (B.4).

(ii) Consider first $\alpha \in (0, 1/3)$. Write

$$\mathbf{P}[Q_n(\infty) \geq n] = \sum_{k=n}^{\lfloor \overline{q}_n \rfloor} \pi_n(k) + \sum_{k=\lfloor \overline{q}_n \rfloor + 1}^{\infty} \pi_n(k). \tag{B.5}$$

We next bound both terms and then show they converge to zero. The first term in Eq. (B.5) is bounded above

$$\begin{aligned}\sum_{k=n}^{\lfloor \overline{q}_n \rfloor} \pi_n(k) &\leq \pi_n(\lfloor \overline{q}_n \rfloor) \cdot (\lfloor \overline{q}_n \rfloor - n + 1) \\ &= \pi_n(\lfloor \overline{q}_n \rfloor) \cdot (\lfloor \overline{q}_n \rfloor - \overline{q}_n + \overline{q}_n - n + 1) \\ &\leq \pi_n(\lfloor \overline{q}_n \rfloor) \cdot (\frac{\rho_n^2}{(1-\rho_n)^2} + 1),\end{aligned}$$

where in the last inequality we used that $\lfloor \overline{q}_n \rfloor \leq \overline{q}_n$, and Theorem 2.1 part *(i)* to obtain an expression for $\overline{q}_n$. In order to bound the second term in Eq. (B.5), first note that

$$\frac{\pi_n(k)}{\pi_n(\lfloor \overline{q}_n \rfloor)} = \prod_{\ell=\lfloor \overline{q}_n \rfloor+1}^{k} \rho_n \cdot \Big(1 + \frac{1}{\sqrt{\ell - n}}\Big), \quad \forall k > \lfloor \overline{q}_n \rfloor.$$

Let

$$a_n = \rho_n \cdot \Big(1 + \frac{1}{\sqrt{\lfloor \overline{q}_n \rfloor + 1 - n}}\Big),$$

which satisfies $a_n < 1$ for all n. Indeed,

$$\rho_n \cdot \Big(1 + \frac{1}{\sqrt{\lfloor \overline{q}_n \rfloor + 1 - n}}\Big) < 1 \;\; \Leftrightarrow \;\; \frac{\rho_n^2}{(1-\rho_n)^2} < \lfloor \overline{q}_n \rfloor + 1 - n$$

which is equivalent to

$$\frac{\rho_n^2}{(1-\rho_n)^2} < 1 - (\overline{q}_n - \lfloor \overline{q}_n \rfloor) + \overline{q}_n - n,$$

from Theorem 2.1 part *(i)*, the last inequality becomes $(\overline{q}_n - \lfloor \overline{q}_n \rfloor) < 1$, which is always true. then

$$\begin{aligned}
\sum_{k=\lfloor \bar{q}_n \rfloor+1}^{\infty} \pi_n(k) &= \pi_n(\lfloor \bar{q}_n \rfloor) \cdot \sum_{k=\lfloor \bar{q}_n \rfloor+1}^{\infty} \prod_{\ell=\lfloor \bar{q}_n \rfloor+1}^{k} \rho_n \cdot \Big(1 + \frac{1}{\sqrt{\ell - n}}\Big) \\
&\overset{(a)}{\leq} \pi_n(\lfloor \bar{q}_n \rfloor) \cdot \sum_{k=\lfloor \bar{q}_n \rfloor+1}^{\infty} \prod_{\ell=\lfloor \bar{q}_n \rfloor+1}^{k} a_n \\
&= \pi_n(\lfloor \bar{q}_n \rfloor) \cdot \sum_{k=\lfloor \bar{q}_n \rfloor+1}^{\infty} a_n^{k-\lfloor \bar{q}_n \rfloor} \\
&= \pi_n(\lfloor \bar{q}_n \rfloor) \cdot a_n^{-\lfloor \bar{q}_n \rfloor} \cdot \frac{a_n^{\lfloor \bar{q}_n \rfloor+1}}{1 - a_n} \\
&< \pi_n(\lfloor \bar{q}_n \rfloor) \cdot \frac{1}{1 - a_n},
\end{aligned}$$

where (a) holds because the term $1 + 1/\sqrt{\ell - n}$ is decreasing in ℓ. Putting the upper bounds for Eq. (B.5) together yields

$$\mathbf{P}[Q_n(\infty) \geq n] \leq \pi_n(\lfloor \bar{q}_n \rfloor) \cdot \Big(\frac{\rho_n^2}{(1 - \rho_n)^2} + 1 + \frac{1}{1 - a_n}\Big).$$

Observe that the term in brackets is $O(n^\gamma)$ for some $\gamma > 0$. Also, we can always consider $\epsilon > 0$ such that $\beta^2/2 > \epsilon$ and then we can use Theorem 2.2 to find n_0 such that for all $n \geq n_0$

$$\pi_n(\lfloor \bar{q}_n \rfloor) \leq \pi_n(\lfloor \underline{q} \rfloor) \cdot \exp\Big(- (\frac{\beta^2}{2} - \epsilon) \cdot n^{1-2\alpha}\Big).$$

Since $\pi_n(\lfloor \underline{q} \rfloor) \leq 1$ and $1 - 2\alpha > 0$ we conclude that

$$\mathbf{P}[Q_n(\infty) \geq n] \leq \exp\Big(- (\frac{\beta^2}{2} - \epsilon) \cdot n^{1-2\alpha}\Big) \cdot O(n^\gamma) \longrightarrow 0, \quad \text{as } n \to \infty.$$

Note that for $\alpha = 1/3$ and $\beta > \beta_2^*$ the same argument holds, we only need to chose $\epsilon > 0$ such that $|g(\beta)| > \epsilon$. This is always possible since for $\beta > \beta_2^*$ Theorem 2.2 establishes that $g(\beta) < 0$. This concludes the proof.

□

Proof of Proposition 2.2. We prove each part separately. First, note that

$$\frac{\pi_n(\lfloor \bar{q}_n \rfloor)}{\pi_n(m)} = \prod_{k=m+1}^{\lfloor \bar{q}_n \rfloor} \frac{\lambda_n}{\mu_n(k) \cdot \min\{k, n\}} = \prod_{k=m+1}^{\lfloor \bar{q}_n \rfloor} \frac{\lambda_n \bar{s}}{\min\{k, n\}} \cdot \Big(1 + \frac{1}{\sqrt{|n - k| \vee 1}}\Big),$$

for any $m < \lfloor \overline{q}_n \rfloor$. Then

$$\log\left(\frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(m)}\right) = (\lfloor \overline{q}_n \rfloor - m)\log(\rho_n) + \sum_{k=m+1}^{\lfloor \overline{q}_n \rfloor} \log\left[\frac{n}{\min\{k,n\}} \cdot \left(1 + \frac{1}{\sqrt{|n-k| \vee 1}}\right)\right] \tag{B.6}$$

(i) For $m = n$: Let $x_n = \lfloor \overline{q}_n \rfloor - n$, then equation (B.6) becomes

$$\begin{aligned}
\log\left(\frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(n)}\right) &= x_n \log(\rho_n) + \sum_{k=n+1}^{\lfloor \overline{q}_n \rfloor} \log\left[1 + \frac{1}{\sqrt{k-n}}\right] \\
&= x_n \cdot \log(\rho_n) + \int_1^{x_n} \log\left[1 + \frac{1}{\sqrt{x}}\right] dx + O(1) \\
&= x_n \cdot \log(\rho_n) + \left[\sqrt{x} + x\log(1 + \frac{1}{\sqrt{x}}) - \log(1+\sqrt{x})\right]\Big|_1^{x_n} + O(1) \\
&= \sqrt{x_n} - \log(1+\sqrt{x_n}) + x_n \cdot \left(\log(\rho_n) + \log\left(1 + \frac{1}{\sqrt{x_n}}\right)\right) + O(1).
\end{aligned}$$

In the expression above we can use that $x_n \to \infty$, $x_n = \lfloor \overline{q}_n \rfloor - \overline{q}_n + \frac{\rho^2}{(1-\rho)^2}$ and Taylor expansions to conclude that

$$\sqrt{x_n} = \frac{\rho_n}{(1-\rho_n)} + o(1),$$

and that

$$x_n \cdot \left(\log(\rho_n) + \log\left(1 + \frac{1}{\sqrt{x_n}}\right)\right) = -\frac{\rho^2}{(1-\rho)^2} + \sqrt{x_n} + O(1) = O(1).$$

Since $(1-\rho_n)n^\alpha \to \beta$ we have

$$\lim_{n\to\infty} \frac{1}{n^\alpha} \log\left(\frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(n)}\right) = \lim_{n\to\infty} \frac{1}{n^\alpha} \frac{\rho_n}{(1-\rho_n)} = \frac{1}{\beta}.$$

(ii) We assume that $\alpha < 1/3$ and we take $m = \lfloor \underline{q} \rfloor$. Note that since $\alpha < 1/3$ we have

$$\frac{27\rho_n^2}{4n \cdot (1-\rho_n)^3} \to 0, \quad as \quad n \to \infty.$$

Then, we can use Theorem 2.1 and do a Taylor expansion to deduce that

$$r_{0,n}(\rho_n) = 1 - \frac{2}{3\sqrt{3}}\sqrt{x} - \frac{2}{27}x - \frac{5}{81\sqrt{3}}x^{3/2} + O(x^2)\Big|_{x = \frac{27\rho_n^2}{4n\cdot(1-\rho_n)^3}}.$$

Hence, since $\alpha < 1/3$ we deduce that

$$n - \underline{q} = n \cdot (1 - \rho_n) + O(n^{(1+\alpha)/2}). \tag{B.7}$$

In order to prove the result for this part of the proposition we need to analyze the term

$$\begin{aligned}
\log\left(\frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)}\right) &= (\lfloor \overline{q}_n \rfloor - \lfloor \underline{q} \rfloor)\log(\rho_n) + \sum_{k=\lfloor \underline{q} \rfloor + 1}^{n-1} \log\left[\frac{n}{k} \cdot \left(1 + \frac{1}{\sqrt{n-k}}\right)\right] \\
&+ \sum_{k=1}^{\lfloor \overline{q}_n \rfloor - n} \log\left[1 + \frac{1}{\sqrt{k}}\right] + \log(2) \\
&= \underbrace{(n - \lfloor \underline{q} \rfloor)\log(\rho_n)}_{A} + \underbrace{\sum_{k=\lfloor \underline{q} \rfloor + 1}^{n-1} \log\left[\frac{n}{k} \cdot \left(1 + \frac{1}{\sqrt{n-k}}\right)\right]}_{B} \\
&+ \underbrace{(\lfloor \overline{q}_n \rfloor - n)\log(\rho_n) + \sum_{k=1}^{\lfloor \overline{q}_n \rfloor - n} \log\left[1 + \frac{1}{\sqrt{k}}\right]}_{C} + \log(2).
\end{aligned}$$

Let's look at each one of the terms A, B and C. For A, using Eq. (B.8), we have that

$$(n - \lfloor \underline{q} \rfloor)\log(\rho_n) = n \cdot (1-\rho_n)\log(\rho_n) + O(n^{(1-\alpha)/2}) = -n \cdot (1-\rho_n)^2 + O(n^{1-3\alpha}) + O(n^{(1-\alpha)/2}),$$

and because $\alpha < 1/3$, we have that $A/n^{1-2\alpha} \to -\beta^2$. So we only need to case analyze B and C. From the proof of part *(i)* we have

$$C = \frac{\rho_n}{(1-\rho_n)} + \log(1-\rho_n) + O(1) = o(n^{1-2\alpha}),$$

where the last equality comes from $\alpha < 1/3$. For B,

$$B = \int_{\lfloor \underline{q} \rfloor}^{n-1} \log\Big[\frac{n}{x}\cdot\Big(1+\frac{1}{\sqrt{n-x}}\Big)\Big]dx + o(n^{1-2\alpha})$$

$$= \Big[x\log(\frac{n}{x}) + x - \sqrt{n-x} - (n-x)\log(1+\frac{1}{\sqrt{n-x}}) + \log(1+\sqrt{n-x})\Big]\Big|_{\lfloor \underline{q} \rfloor}^{n-1}$$

$$+ o(n^{1-2\alpha})$$

$$= n-1-\Big[\lfloor \underline{q} \rfloor\log(\frac{n}{\lfloor \underline{q} \rfloor}) + \lfloor \underline{q} \rfloor - \sqrt{n-\lfloor \underline{q} \rfloor} - (n-\lfloor \underline{q} \rfloor)\log(1+\frac{1}{\sqrt{n-\lfloor \underline{q} \rfloor}})$$

$$+\log(1+\sqrt{n-\lfloor \underline{q} \rfloor})\Big] + o(n^{1-2\alpha})$$

$$= n - \lfloor \underline{q} \rfloor\log(\frac{n}{\lfloor \underline{q} \rfloor}) - \lfloor \underline{q} \rfloor + o(n^{1-2\alpha})$$

$$= n - \lfloor \underline{q} \rfloor - \lfloor \underline{q} \rfloor\cdot\Big(\frac{(n-\lfloor \underline{q} \rfloor)}{\lfloor \underline{q} \rfloor} - \frac{(n-\lfloor \underline{q} \rfloor)^2}{2\lfloor \underline{q} \rfloor^2}\Big) + o(n^{1-2\alpha})$$

$$= \frac{(n-\lfloor \underline{q} \rfloor)^2}{2\lfloor \underline{q} \rfloor} + o(n^{1-2\alpha}),$$

using that $\alpha < 1/3$ it follows that this last expression, when scaled by $1/n^{1-2\alpha}$, converges to $\beta^2/2$. Therefore,

$$\frac{1}{n^{1-2\alpha}}\log\Big(\frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)}\Big) \to -\beta^2 + \frac{\beta^2}{2} + 0 = -\frac{\beta^2}{2}, \quad \text{as} \quad n\to\infty,$$

as required.

(iii) We assume that $\alpha = 1/3$ and we take $m = \lfloor \underline{q} \rfloor$. Note that

$$r_{0,n}(\rho_n) \to \frac{4}{3}\cdot\cos\Big(\frac{1}{3}\arccos\Big(-\sqrt{\Big(\frac{\beta_1^*}{\beta}\Big)^3}\Big)\Big)^2 \triangleq r(\beta), \quad as \quad n\to\infty. \quad \text{(B.8)}$$

Observe that since we are considering $\beta \geq \beta_1^*$ the $\arccos(\cdot)$ term is well defined and, therefore, so is $r(\beta)$. We need to analyze the following expression

$$\log\Big(\frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)}\Big) = \underbrace{(n-\lfloor \underline{q} \rfloor)\log(\rho_n)}_{A} + \underbrace{\sum_{k=\lfloor \underline{q} \rfloor+1}^{n-1}\log\Big[\frac{n}{k}\cdot\Big(1+\frac{1}{\sqrt{n-k}}\Big)\Big]}_{B}$$

$$+ \underbrace{(\lfloor \overline{q}_n \rfloor - n)\log(\rho_n) + \sum_{k=1}^{\lfloor \overline{q}_n \rfloor - n}\log\Big[1+\frac{1}{\sqrt{k}}\Big]}_{C} + \log(2).$$

Let's look at each one of the terms A, B and C. For A, using Theorem 2.1 we have that

$$A = (n - \lfloor \underline{q} \rfloor) \log(\rho_n) = n \cdot (1-\rho_n) \cdot r_{0,n}(\rho_n) \log(\rho_n) + o(1) = -n(1-\rho_n)^2 r_{0,n}(\rho_n) + o(n^{1/3}).$$

Similarly to part *(ii)* above, for C we deduce

$$C = \frac{\rho_n}{(1-\rho_n)} + \log(1-\rho_n) + O(1) = \frac{\rho_n}{(1-\rho_n)} + o(n^{1/3}).$$

Finally, for B (similarly to part *(ii)* above)

$$\begin{aligned} B &= n - \Big[\lfloor \underline{q} \rfloor \log(\frac{n}{\lfloor \underline{q} \rfloor}) + \lfloor \underline{q} \rfloor - \sqrt{n - \lfloor \underline{q} \rfloor} - (n - \lfloor \underline{q} \rfloor) \log(1 + \frac{1}{\sqrt{n - \lfloor \underline{q} \rfloor}}) \\ &\quad + \log(1 + \sqrt{n - \lfloor \underline{q} \rfloor}) \Big] + o(n^{1/3}) \\ &= n - \lfloor \underline{q} \rfloor - \lfloor \underline{q} \rfloor \log(\frac{n}{\lfloor \underline{q} \rfloor}) + 2\sqrt{n - \lfloor \underline{q} \rfloor} + o(n^{1/3}) \\ &= \frac{(n - \lfloor \underline{q} \rfloor)^2}{2\lfloor \underline{q} \rfloor} + 2\sqrt{n - \lfloor \underline{q} \rfloor} + o(n^{1/3}), \end{aligned}$$

and, therefore, using that $n^{1/3}(1-\rho_n) \to \beta$, $n - \lfloor \underline{q} \rfloor = n(1-\rho_n) r_{0,n}(\rho_n)$ and Eq. (B.8) we can compute the limit

$$\lim_{n\to\infty} \frac{B}{n^{1/3}} = \lim_{n\to\infty} \frac{1}{n^{1/3}} \cdot \frac{(n - \lfloor \underline{q} \rfloor)^2}{2\lfloor \underline{q} \rfloor} + 2\frac{\sqrt{n - \lfloor \underline{q} \rfloor}}{n^{1/3}} = \frac{\beta^2 r(\beta)^2}{2} + 2\sqrt{\beta r(\beta)},$$

where $r(\beta)$ is defined in Eq. (B.8). From this we can deduce that

$$\frac{1}{n^{1/3}} \log\left(\frac{\pi_n(\overline{q}_n)}{\pi_n(\underline{q})}\right) \to -\beta^2 r(\beta) + \frac{\beta^2 r(\beta)^2}{2} + 2\sqrt{\beta r(\beta)} + \frac{1}{\beta} \triangleq g(\beta), \quad \text{as} \quad n \to \infty. \tag{B.9}$$

It is possible to verify that $g(\beta)$ satisfies $g(\beta_1^*) > 0$ and it is strictly decreasing for $\beta \geq \beta_1^*$, with $\lim_{\beta\to\infty} g(\beta) = -\infty$, see Figure B.1. Therefore, there exists $\beta_2^* > \beta_1^*$ such that $g(\beta_2^*) = 0$. Thus we have verified that $g(\beta)$ is such that if $\beta_1^* < \beta < \beta_2^*$ then $g(\beta) > 0$, whereas if $\beta > \beta_2^*$ then $g(\beta) < 0$.

□

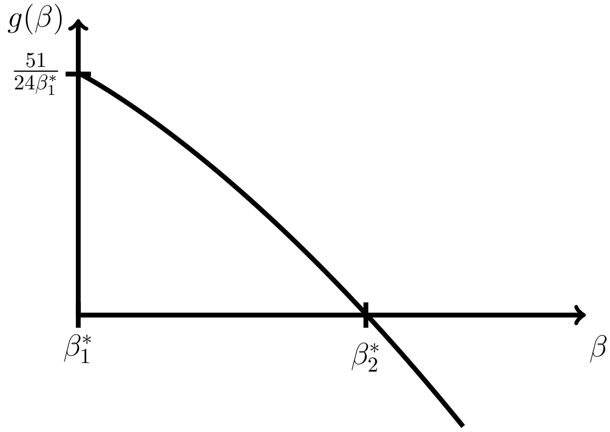


Figure B.1: Function $g(\beta)$ as defined in Eq. (B.9), $g(\beta)$ is strictly decreasing and it crosses zero at β_2^*.

Proof of Theorem 2.3. We make use of the lemmata B.1 and B.2 which we first state and then prove after the proof of this theorem. We also make use of Proposition 2.3 which is stated in the main text and proven in this appendix.

In order to simplify notation let $p_n^+ = \mathbf{P}[Q_n(\infty) \geq n]$. Let $\beta = \beta_2^*$ and $\alpha = 1/3$ then from Lemma B.1 and Lemma B.2 there exists n_1 such that

$$\underbrace{\frac{\frac{1-\exp\left(-\frac{C^2\beta^3}{2}\right)}{\frac{C^2\beta^3}{2}}}{1+\frac{\exp\left(-\frac{C^2}{2}\left(1-\frac{1}{2(\beta\cdot r(\beta))^{3/2}}\right)\right)}{\frac{C^2}{2}\left(1-\frac{1}{2(\beta\cdot r(\beta))^{3/2}}\right)}}}_{A(C)} \cdot \frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)} \leq \frac{p_n^+}{(1-p_n^+)}$$

$$\leq \underbrace{\frac{1+1\cdot\frac{\exp\left(-\frac{C^2\beta^3}{4}\right)}{\frac{C^2\beta^3}{4}}}{\frac{1-\exp\left(-C^2\cdot\left(1-\frac{1}{2(\beta\cdot r(\beta))^{3/2}}\right)\right)}{C^2\cdot\left(1-\frac{1}{2(\beta\cdot r(\beta))^{3/2}}\right)}}}_{B(C)} \cdot \frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)},$$

$\forall n \geq n_1$. Next, fix $\epsilon > 0$ then by Proposition 2.3 we have that there exists n_2 such that

$$\exp(-\epsilon + c) \leq \frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)} \leq \exp(\epsilon + c), \quad \forall n \geq n_2.$$

Therefore, for all $n \geq \max\{n_1, n_2\}$

$$A(C) \cdot \exp(-\epsilon + c) \;\leq\; \frac{p_n^+}{(1-p_n^+)} \;\leq\; B(C) \cdot \exp(\epsilon + c),$$

or, alternatively, (letting $\epsilon \to 0$)

$$\frac{A(C)}{e^{-c} + A(C)} \;\leq\; \liminf_{n\to\infty} p_n^+ \leq \limsup_{n\to\infty} p_n^+ \;\leq\; \frac{B(C)}{e^{-c} + B(C)}.$$

Now we want to find the tightest upper and lower bound. To do this it is enough to maximize the LHS and minimize the RHS above as a function of C. Since all the parameters are known ($\beta_2^* \approx 2.6030$ and $r(\beta_2^*) \approx 0.7192$) we can obtain numerical values,

$$\max_{C>0}\left\{\frac{A(C)}{e^{-c} + A(C)}\right\} \approx \frac{0.0524}{e^{-c} + 0.0524}, \quad \text{and} \quad \min_{C>0}\left\{\frac{B(C)}{e^{-c} + B(C)}\right\} \approx \frac{1.3173}{e^{-c} + 1.3173}.$$

So if we fix $p_H \in (0,1)$ then there exists $c^* \in \mathbb{R}$ such that

$$\frac{1.3173}{e^{-c^*} + 1.3173} = p_H,$$

and c^* increases with p_H. Therefore if we let

$$p_L(p_H) = \frac{0.0524}{e^{-c^*} + 0.0524},$$

we have that $p_L(p_H) \in (0,1)$ increases with p_H. In particular, $\lim_{p_H \to 1} p_L(p_H) = 1$ and $\lim_{p_H \to 0} p_L(p_H) = 0$, as desired.

□

Lemma B.1 *Fix $\alpha \in (0, 1/3)$ and $\beta > 0$, or $\alpha = 1/3$ and $\beta > \beta_1^*$. Suppose that $\lim_{n\to\infty} n^\alpha(1-\rho_n) = \beta$ and let $C > 0$ be a constant then*

$$2 \cdot \frac{1 - \exp\left(-C^2 \cdot \left(1 - \frac{\mathbf{1}_{\{\alpha=1/3\}}}{2(\beta \cdot r(\beta))^{3/2}}\right)\right)}{C^2 \cdot \left(1 - \frac{\mathbf{1}_{\{\alpha=1/3\}}}{2(\beta \cdot r(\beta))^{3/2}}\right)} \leq \liminf_{n\to\infty} \frac{1}{C\sqrt{n}} \cdot \frac{\mathbf{P}[Q_n(\infty) < n]}{\pi_n(\lfloor \underline{q} \rfloor)},$$

and

$$\limsup_{n\to\infty} \frac{1}{C\sqrt{n}} \cdot \frac{\mathbf{P}[Q_n(\infty) < n]}{\pi_n(\lfloor \underline{q} \rfloor)} \leq 2 \;+\; 2 \cdot \frac{\exp\left(-\frac{C^2}{2}\left(1 - \frac{\mathbf{1}_{\{\alpha=1/3\}}}{2(\beta \cdot r(\beta))^{3/2}}\right)\right)}{\frac{C^2}{2}\left(1 - \frac{\mathbf{1}_{\{\alpha=1/3\}}}{2(\beta \cdot r(\beta))^{3/2}}\right)},$$

where $r(\beta) = \lim_{n\to\infty} r_{0,n}(\rho_n)$.

Lemma B.2 *Fix* $\alpha \in (0,1)$ *and* $\beta > 0$. *Suppose that* $\lim_{n\to\infty} n^{\alpha}(1-\rho_n) = \beta$ *and let* $C > 0$ *be a constant then*

$$2 \cdot \frac{1-\exp\Big(-\frac{C^2\beta^3}{2}\Big)}{\frac{C^2\beta^3}{2}} \leq \liminf_{n\to\infty} \frac{1}{Cn^{\frac{3}{2}\alpha}} \cdot \frac{\mathbf{P}[Q_n(\infty) \geq n]}{\pi_n(\lfloor \overline{q}_n \rfloor)},$$

and

$$\limsup_{n\to\infty} \frac{1}{Cn^{\frac{3}{2}\alpha}} \cdot \frac{\mathbf{P}[Q_n(\infty) \geq n]}{\pi_n(\lfloor \overline{q}_n \rfloor)} \leq 2 \;+\; 2 \cdot \frac{\exp\Big(-\frac{C^2\beta^3}{4}\Big)}{\frac{C^2\beta^3}{4}}.$$

Proof of Lemma B.1. We start we the lower bound. Let $b_n = C\sqrt{n}$ and note that

$$\begin{aligned}
\frac{\mathbf{P}[Q_n(\infty) < n]}{\pi_n(\lfloor \underline{q} \rfloor) b_n} &= \frac{1}{b_n} \sum_{k=0}^{n-1} \frac{\pi_n(k)}{\pi_n(\lfloor \underline{q} \rfloor)} \\
&\geq \frac{1}{b_n} \sum_{k=\lfloor \underline{q} \rfloor - b_n}^{\lfloor \underline{q} \rfloor + b_n} \frac{\pi_n(k)}{\pi_n(\lfloor \underline{q} \rfloor)} \\
&= \frac{1}{b_n} \sum_{k=\underline{q}-b_n}^{\lfloor \underline{q} \rfloor} \prod_{\ell=k+1}^{\lfloor \underline{q} \rfloor} \frac{1}{\rho_n} \frac{\ell}{n} \frac{1}{\Big(1+\frac{1}{\sqrt{n-\ell}}\Big)} + \frac{1}{b_n} \sum_{k=\lfloor \underline{q} \rfloor+1}^{\lfloor \underline{q} \rfloor + b_n} \prod_{\ell=\lfloor \underline{q} \rfloor+1}^{k} \rho_n \frac{n}{\ell}\Big(1+\frac{1}{\sqrt{n-\ell}}\Big) \\
&\overset{(a)}{\geq} \frac{1}{b_n} \sum_{k=\lfloor \underline{q} \rfloor - b_n}^{\lfloor \underline{q} \rfloor} \Big(\underbrace{\frac{1}{\rho_n} \frac{\lfloor \underline{q} \rfloor - b_n}{n} \frac{1}{\Big(1+\frac{1}{\sqrt{n-\lfloor \underline{q} \rfloor + b_n}}\Big)}}_{s_{1n}} \Big)^{\lfloor \underline{q} \rfloor - k} \\
&+ \frac{1}{b_n} \sum_{k=\lfloor \underline{q} \rfloor+1}^{\lfloor \underline{q} \rfloor + b_n} \Big(\underbrace{\rho_n \frac{n}{\lfloor \underline{q} \rfloor + b_n} \Big(1+\frac{1}{\sqrt{n-\lfloor \underline{q} \rfloor - b_n}}\Big)}_{s_{2n}} \Big)^{k-\lfloor \underline{q} \rfloor} \\
&= \frac{1}{b_n} \cdot \frac{1-s_{1n}^{b_n+1}}{1-s_{1n}} + \frac{1}{b_n} \cdot \frac{s_{2n}-s_{2n}^{b_n+1}}{1-s_{2n}}, \qquad \text{(B.10)}
\end{aligned}$$

where (a) comes from the fact that the function

$$h_n(x) = \frac{1}{x} \cdot (1 + \frac{1}{\sqrt{n-x}}),$$

is decreasing in $[0, \underline{q} + b_n]$ for n large, we show this at the end of the proof. Next we show that both terms in Eq. (B.10) above converge to a constant. First note that from Theorem 2.1 we have that $\underline{q} = n - z_n^2$ where z_n^2 is given by $n \cdot (1-\rho_n) \cdot r_{0,n}(\rho_n)$.

Note that $1 - r_{0,n}(\rho_n)$ is of order $O(n^{-(1-3\alpha)/2})$ if $\alpha < 1/3$ and $r_{0,n}(\rho_n)$ converges to a function of β, $r(\beta)$, for $\alpha = 1/3$

$$r(\beta) = \frac{4}{3} \cdot \cos\Big(\frac{1}{3} \arccos\Big(- \sqrt{\Big(\frac{\beta_1^*}{\beta}\Big)^3}\Big)\Big)^2.$$

For the rest of the proof we will use $\tilde{b}_n$ to denote $b_n + (\underline{q} - \lfloor \underline{q} \rfloor)$. Note that $|(\underline{q} - \lfloor \underline{q} \rfloor)| \leq 1$. Let $\ell_n = (n - z_n^2 - \tilde{b}_n)/n$, for s_{1n} we have that

$$\begin{aligned}
s_{1n} &= \frac{1}{\rho_n} \ell_n \frac{1}{\Big(1 + \frac{1}{\sqrt{z_n^2 + \tilde{b}_n}}\Big)} \\
&= \frac{1}{\rho_n} \ell_n \Big(1 - \frac{1}{\sqrt{z_n^2 + \tilde{b}_n}} + O(\frac{1}{z_n^2 + \tilde{b}_n})\Big) \\
&= \frac{1}{\rho_n} \ell_n \Big(1 - \frac{1}{z_n} \frac{1}{\sqrt{1 + \frac{\tilde{b}_n}{z_n^2}}} + O(\frac{1}{z_n^2 + \tilde{b}_n})\Big) \\
&= \frac{1}{\rho_n} \ell_n \Big(1 - \frac{1}{z_n} + \frac{\tilde{b}_n}{2 z_n^3} + O(\frac{\tilde{b}_n^2}{z_n^5}) + O(\frac{1}{z_n^2 + \tilde{b}_n})\Big) \\
&= \frac{\ell_n}{\rho_n} - \frac{\ell_n}{\rho_n z_n} + \frac{\ell_n \tilde{b}_n}{2 \rho_n z_n^3} + O(\frac{\ell_n \tilde{b}_n^2}{\rho_n z_n^5}) + O(\frac{\ell_n}{\rho_n (z_n^2 + \tilde{b}_n)}),
\end{aligned}$$

the last two terms above times b_n converge to zero. Hence,

$$b_n \cdot (1 - s_{1n}) = b_n - \frac{b_n \ell_n}{\rho_n} + \frac{b_n \ell_n}{\rho_n z_n} - \frac{\ell_n b_n \tilde{b}_n}{2 \rho_n z_n^3} + o(1).$$

The expression above converges to $C^2 \cdot \Big(1 - \frac{\mathbf{1}_{\{\alpha = 1/3\}}}{2(\beta \cdot r(\beta))^{3/2}}\Big)$. Indeed, the fourth term above is $O(n)/O(n^{\frac{3}{2}(1-\alpha)})$ which is $o(1)$ when $\alpha < 1/3$ and converges to $-C^2/(2\beta^{3/2} \cdot r(\beta)^{3/2})$ when $\alpha = 1/3$. The first three terms converge to C^2. Indeed, recall that $\underline{q}$ solves the equation

$$(n - \underline{q}) + \frac{n \rho_n}{\sqrt{n - \underline{q}}} = n(1 - \rho_n), \quad \text{or equivalently,} \quad z_n^2 + \frac{n \rho_n}{z_n} = n(1 - \rho_n). \quad \text{(B.11)}$$

Hence

$$
\begin{aligned}
b_n \cdot (1 - s_{1n}) &= b_n - \frac{b_n \ell_n}{\rho_n} + \frac{b_n \ell_n}{\rho_n z_n} - \frac{\ell_n b_n \tilde{b}}{2\rho_n z_n^3} + o(1) \\
&= b_n - \Big(1 - \frac{z_n^2}{n} - \frac{\tilde{b}_n}{n}\Big) \cdot \frac{b_n}{\rho_n} + \Big(1 - \frac{z_n^2}{n} - \frac{\tilde{b}_n}{n}\Big) \cdot \frac{b_n}{\rho_n z_n} - \frac{\ell_n b_n \tilde{b}}{2\rho_n z_n^3} + o(1) \\
&= b_n - \Big(1 - \frac{z_n^2}{n}\Big) \cdot \frac{b_n}{\rho_n} + \frac{b_n}{\rho_n z_n} + \frac{b_n \tilde{b}_n}{\rho_n n} - \frac{\ell_n b_n \tilde{b}}{2\rho_n z_n^3} + o(1) \\
&\overset{\text{Eq.(B.11)}}{=} b_n - \Big(1 - \frac{z_n^2}{n}\Big) \cdot \frac{b_n}{\rho_n} + \frac{b_n}{\rho_n^2}\Big((1-\rho_n) - \frac{z_n^2}{n}\Big) + \frac{b_n \tilde{b}_n}{\rho_n n} - \frac{\ell_n b_n \tilde{b}}{2\rho_n z_n^3} + o(1) \\
&= \frac{b_n \tilde{b}_n}{\rho_n n} - \frac{\ell_n b_n \tilde{b}}{2\rho_n z_n^3} + o(1) \\
&\to C^2 - \mathbf{1}_{\{\alpha = 1/3\}} \cdot \frac{C^2}{2(\beta \cdot r(\beta))^{3/2}}.
\end{aligned}
$$

Given this, we have

$$
\begin{aligned}
\frac{1}{b_n} \cdot \frac{1 - s_{1n}^{b_n+1}}{1 - s_{1n}} &= \frac{1 - \exp\Big((b_n + 1)\log(s_{1n})\Big)}{b_n(1 - s_{1n})} \\
&= \frac{1 - \exp\Big(-b_n(1 - s_{1n}) + o(1)\Big)}{b_n(1 - s_{1n})} \\
&\to \frac{1 - \exp\Big(-C^2 \cdot \Big(1 - \frac{\mathbf{1}_{\{\alpha=1/3\}}}{2(\beta \cdot r(\beta))^{3/2}}\Big)\Big)}{C^2 \cdot \Big(1 - \frac{\mathbf{1}_{\{\alpha=1/3\}}}{2(\beta \cdot r(\beta))^{3/2}}\Big)},
\end{aligned}
$$

note that the function $(\beta r(\beta))^{3/2}$ is strictly increasing and equal to $1/2$ at $\beta = \beta_1^*$. Because we are considering $\beta > \beta_1^*$, the last expression above is positive. Finally, since this limit is a lower bound we obtain the desired lower bound for the lim inf.

A similar argument shows that

$$
\frac{1}{b_n} \cdot \frac{s_{2n} - s_{2n}^{b_n+1}}{1 - s_{2n}} \to \frac{1 - \exp\Big(-C^2 \cdot \Big(1 - \frac{\mathbf{1}_{\{\alpha=1/3\}}}{2(\beta \cdot r(\beta))^{3/2}}\Big)\Big)}{C^2 \cdot \Big(1 - \frac{\mathbf{1}_{\{\alpha=1/3\}}}{2(\beta \cdot r(\beta))^{3/2}}\Big)}
$$

Next we move to the upper bound. We first note that

$$
\sum_{k=\lfloor \underline{q} \rfloor - b_n}^{\lfloor \underline{q} \rfloor + b_n} \pi_n(k) \le \pi_n(\lfloor \underline{q} \rfloor) \cdot (2 \cdot b_n + 1).
$$

Now we bound the terms in $[0, \lfloor \underline{q} \rfloor - b_n - 1]$ and $[\lfloor \underline{q} \rfloor + b_n + 1, n-1]$ separately.

$$
\begin{aligned}
\frac{1}{b_n \cdot \pi_n(\lfloor \underline{q} \rfloor)} \cdot \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \pi_n(k) &= \frac{1}{b_n} \cdot \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \prod_{\ell=k+1}^{\lfloor \underline{q} \rfloor} \frac{1}{\rho_n} \cdot \frac{\ell}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-\ell}}\right)} \\
&\overset{(a)}{\leq} \frac{1}{b_n} \cdot \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \prod_{\ell=k+1}^{\lfloor \underline{q} \rfloor - 1} \frac{1}{\rho_n} \cdot \frac{\ell}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-\ell}}\right)} \\
&\overset{(b)}{\leq} \frac{1}{b_n} \cdot \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \left\{ \frac{1}{\lfloor \underline{q} \rfloor - k - 1} \cdot \sum_{\ell=k+1}^{\lfloor \underline{q} \rfloor - 1} \frac{1}{\rho_n} \cdot \frac{\ell}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-\ell}}\right)} \right\}^{\lfloor \underline{q} \rfloor - k - 1} \\
&\overset{(c)}{\leq} \frac{1}{b_n} \cdot \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \Bigg\{ \underbrace{\frac{1}{b_n} \cdot \sum_{\ell=\lfloor \underline{q} \rfloor - b_n}^{\lfloor \underline{q} \rfloor - 1} \frac{1}{\rho_n} \cdot \frac{\ell}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-\ell}}\right)}}_{s_{1n}} \Bigg\}^{\lfloor \underline{q} \rfloor - k - 1} \\
&= \frac{s_{1n}^{b_n} - s_{1n}^{\lfloor \underline{q} \rfloor - 1}}{b_n \cdot (1 - s_{1n})},
\end{aligned}
$$

where in (a) we use that

$$
\frac{1}{\rho_n} \cdot \frac{\lfloor \underline{q} \rfloor}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n - \lfloor \underline{q} \rfloor}}\right)} \leq 1,
$$

in (b) the inequality of arithmetic and geometric means, and in (c) the fact that $h_n(x)$ is decreasing for $x \leq \underline{q} + b_n$. In order to simplify notation let $\tilde{z}_n^2 = n - \lfloor \underline{q} \rfloor$. Let us analyze s_{1n},

$$
\begin{aligned}
s_{1n} &= \frac{1}{b_n} \cdot \sum_{\ell=\lfloor q \rfloor - b_n}^{\lfloor q \rfloor - 1} \frac{1}{\rho_n} \cdot \frac{\ell}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-\ell}}\right)} \\
&\le \frac{1}{b_n} \cdot \int_{\lfloor q \rfloor - b_n}^{\lfloor q \rfloor} \frac{1}{\rho_n} \cdot \frac{x}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-x}}\right)} dx \\
&= \frac{1}{b_n n \rho_n} \cdot \Bigg[\frac{1}{6}\Big(-3n^2 + n(8\sqrt{n-x} + 6) + 4x\sqrt{n-x} - 12\sqrt{n-x} + 3x^2 - 6x \Big) \\
&\quad - 2(n-1)\log(\sqrt{n-x} + 1) \Bigg] \Bigg|_{\lfloor q \rfloor - b_n}^{\lfloor q \rfloor} \\
&= \frac{1}{b_n n \rho_n} \cdot \Bigg[\frac{12n - 4\tilde{z}_n^2 - 12}{6} \Big(\tilde{z}_n - \sqrt{\tilde{z}_n^2 + b_n} \Big) + \frac{4}{6} b_n \sqrt{\tilde{z}_n^2 + b_n} + (n - \tilde{z}_n^2) b_n - \frac{b_n^2}{2} \\
&\quad - 2(n-1)\log\Big(\frac{\tilde{z}_n + 1}{\sqrt{\tilde{z}_n^2 + b_n} + 1} \Big) \Bigg],
\end{aligned}
$$

If we denote this last expression $\tilde{s}_{1n}$ then for $b_n(1 - \tilde{s}_{1n})$ we have that

$$
\begin{aligned}
b_n(1 - \tilde{s}_{1n}) &= b_n - \frac{1}{n\rho_n} \left[\frac{12n - 4\tilde{z}_n^2 - 12}{6} \tilde{z}_n \Big(1 - \sqrt{1 + \frac{b_n}{\tilde{z}_n^2}} \Big) + (n - \tilde{z}_n^2) b_n - \frac{b_n^2}{2} \right] + o(1) \\
&= b_n - \frac{1}{n\rho_n} \left[\frac{12n - 4\tilde{z}_n^2 - 12}{6} \Big(-\frac{b_n}{2\tilde{z}_n} + \frac{b_n^2}{8\tilde{z}_n^3} \Big) + (n - \tilde{z}_n^2) b_n - \frac{b_n^2}{2} \right] + o(1) \\
&\overset{\text{Eq.(B.11)}}{=} b_n - \Big(1 - \frac{\tilde{z}_n^2}{n} \Big) \frac{b_n}{\rho_n} + \Big(\frac{(1-\rho_n)}{\rho_n} - \frac{z_n^2}{n\rho_n} \Big) \frac{b_n}{\rho_n} \cdot \frac{z_n}{\tilde{z}_n} - \frac{b_n^2}{4\rho_n \tilde{z}_n^3} + \frac{b_n^2}{2\rho_n n} \\
&\quad + o(1) \\
&= b_n \frac{(1-\rho_n)^2}{\rho_n^2} (1 - r_{0,n}) - \frac{b_n^2}{4\rho_n \tilde{z}_n^3} + \frac{b_n^2}{2\rho_n n} + o(1) \\
&= -\frac{b_n^2}{4\rho_n \tilde{z}_n^3} + \frac{b_n^2}{2\rho_n n} + o(1),
\end{aligned}
$$

where in the last equality we used that when $\alpha < 1/3$ then $(1 - r_{0,n}) = O(n^{-(1-3\alpha)/2})$. This last expression converges to $\frac{1}{2}(C^2 - \mathbf{1}_{\{\alpha=1/3\}} \cdot \frac{C^2}{2(\beta \cdot r(\beta))^{3/2}})$, which is a positive quantity. Therefore, for n large enough we have $\tilde{s}_{1n} \leq 1$ and, thus

$$\frac{1}{b_n \cdot \pi_n(\lfloor \underline{q} \rfloor)} \cdot \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \pi_n(k) \leq \frac{s_{1n}^{b_n} - s_{1n}^{\lfloor \underline{q} \rfloor + 1}}{b_n \cdot (1 - s_{1n})} \leq \frac{\tilde{s}_{1n}^{b_n}}{b_n \cdot (1 - \tilde{s}_{1n})}$$
$$\to \frac{\exp\Big(- \frac{1}{2}\Big(C^2 - \mathbf{1}_{\{\alpha = 1/3\}} \cdot \frac{C^2}{2(\beta \cdot r(\beta))^{3/2}}\Big)\Big)}{\frac{1}{2}\Big(C^2 - \mathbf{1}_{\{\alpha = 1/3\}} \cdot \frac{C^2}{2(\beta \cdot r(\beta))^{3/2}}\Big)},$$

where is the second inequality we used that for n large enough $\tilde{s}_{1n} \leq 1$.

Next we move to the range $[\lfloor \underline{q} \rfloor + b_n + 1, n - 1]$. First observe that

$$\sum_{k=\lfloor \tilde{q}_n \rfloor}^{n-1} \pi_n(k) \leq \pi_n(n) \cdot (n - \lfloor \tilde{q}_n \rfloor) \leq \frac{\pi_n(n)}{\pi_n(\lfloor \overline{q}_n \rfloor)} \cdot (n - \lfloor \tilde{q}_n \rfloor) \to 0,$$

where the limit follows from Proposition 2.2 part *i)*. Thus,

$$\begin{aligned}
\frac{1}{b_n \cdot \pi_n(\lfloor \underline{q} \rfloor)} \cdot \sum_{k=\lfloor \underline{q} \rfloor + b_n + 1}^{n-1} \pi_n(k) &= \frac{1}{b_n} \cdot \sum_{k=\lfloor \underline{q} \rfloor + b_n + 1}^{\lfloor \tilde{q}_n \rfloor} \prod_{\ell = \lfloor \underline{q} \rfloor + 1}^{k} \rho_n \cdot \frac{n}{\ell} \cdot \Big(1 + \frac{1}{\sqrt{n - \ell}}\Big) + o(1) \\
&\leq \frac{1}{b_n} \cdot \sum_{k=\lfloor \underline{q} \rfloor + b_n + 1}^{\lfloor \tilde{q}_n \rfloor} \left\{ \frac{1}{k - \lfloor \underline{q} \rfloor} \cdot \sum_{\ell = \lfloor \underline{q} \rfloor + 1}^{k} \rho_n \cdot \frac{n}{\ell} \cdot \Big(1 + \frac{1}{\sqrt{n - \ell}}\Big) \right\}^{k - \lfloor \underline{q} \rfloor} \\
&\leq \frac{1}{b_n} \cdot \sum_{k=\lfloor \underline{q} \rfloor + b_n + 1}^{\lfloor \tilde{q}_n \rfloor} \Bigg\{ \underbrace{\frac{1}{b_n} \cdot \sum_{\ell = \lfloor \underline{q} \rfloor + 1}^{\lfloor \underline{q} \rfloor + b_n} \rho_n \cdot \frac{n}{\ell} \cdot \Big(1 + \frac{1}{\sqrt{n - \ell}}\Big)}_{s_{2n}} \Bigg\}^{k - \lfloor \underline{q} \rfloor} \\
&\leq \frac{s_{2n}^{b_n + 1}}{b_n \cdot (1 - s_{2n})}.
\end{aligned}$$

Let us analyze s_{2n},

$$\begin{aligned}
s_{2n} &= \frac{1}{b_n} \cdot \sum_{\ell = \lfloor \underline{q} \rfloor + 1}^{\lfloor \underline{q} \rfloor + b_n} \rho_n \cdot \frac{n}{\ell} \cdot \Big(1 + \frac{1}{\sqrt{n - \ell}}\Big) \\
&\leq \frac{1}{b_n} \cdot \int_{\lfloor \underline{q} \rfloor}^{\lfloor \underline{q} \rfloor + b_n} \rho_n \cdot \frac{n}{x} \cdot \Big(1 + \frac{1}{\sqrt{n - x}}\Big) dx \\
&= \frac{\rho_n \cdot n}{b_n} \cdot \Big[\log(x) + \frac{1}{\sqrt{n}} \Big(\log(\sqrt{n} - \sqrt{n - x}) - \log(\sqrt{n - x} + \sqrt{n}) \Big) \Big] \Big|_{\lfloor \underline{q} \rfloor}^{\lfloor \underline{q} \rfloor + b_n} \\
&= \frac{\rho_n \cdot n}{b_n} \cdot \Big[\log(1 + \frac{b_n}{\lfloor \underline{q} \rfloor}) + \frac{1}{\sqrt{n}} \Big(\log \Big[\frac{1 - \frac{\tilde{z}_n}{\sqrt{n}} \cdot \sqrt{1 - \frac{b_n}{\tilde{z}_n^2}}}{1 + \frac{\tilde{z}_n}{\sqrt{n}} \cdot \sqrt{1 - \frac{b_n}{\tilde{z}_n^2}}} \Big] - \log \Big[\frac{1 - \frac{\tilde{z}_n}{\sqrt{n}}}{1 + \frac{\tilde{z}_n}{\sqrt{n}}} \Big] \Big) \Big].
\end{aligned}$$

Denoting this last expression by $\tilde{s}_{2n}$ we have that

$$
\begin{aligned}
b_n \cdot (1-\tilde{s}_{2n}) &= b_n - \rho_n \cdot n \cdot \log(1+\frac{b_n}{\lfloor \underline{q} \rfloor}) - \rho_n \cdot \sqrt{n} \cdot \Big(\log \Big[\frac{1-\frac{\tilde{z}_n}{\sqrt{n}} \cdot \sqrt{1-\frac{b_n}{\tilde{z}_n^2}}}{1+\frac{\tilde{z}_n}{\sqrt{n}} \cdot \sqrt{1-\frac{b_n}{\tilde{z}_n^2}}} \Big] \\
&\quad - \log \Big[\frac{1-\frac{\tilde{z}_n}{\sqrt{n}}}{1+\frac{\tilde{z}_n}{\sqrt{n}}} \Big] \Big) \\
&= b_n - \rho_n \cdot n \cdot \Big(\frac{b_n}{\lfloor \underline{q} \rfloor} - \frac{b_n^2}{2\lfloor \underline{q} \rfloor^2} \Big) - \rho_n \cdot \sqrt{n} \cdot \Big(\frac{b_n}{\sqrt{n}\tilde{z}_n} + \frac{b_n^2}{4\sqrt{n}\tilde{z}_n^3} \Big) + o(1) \\
&= \rho_n \cdot n \cdot \frac{b_n^2}{2\lfloor \underline{q} \rfloor^2} - \rho_n \cdot \frac{b_n^2}{4\tilde{z}_n^3} + b_n \cdot \Big(1 - \frac{\rho_n \cdot n}{\lfloor \underline{q} \rfloor} - \frac{\rho_n}{\tilde{z}_n} \Big) + o(1) \\
&\overset{\text{Eq.(B.11)}}{=} \rho_n \cdot n \cdot \frac{b_n^2}{2\lfloor \underline{q} \rfloor^2} - \rho_n \cdot \frac{b_n^2}{4\tilde{z}_n^3} + b_n \cdot (1-\rho_n)^2 \cdot (1-r_n) \cdot r_n + o(1) \\
&= \rho_n \cdot n \cdot \frac{b_n^2}{2\lfloor \underline{q} \rfloor^2} - \rho_n \cdot \frac{b_n^2}{4\tilde{z}_n^3} + o(1)
\end{aligned}
$$

where from the first to second equality we we did a Taylor expansion around zero of the functions $\log(1+x), \log((1-x)/(1+x))$ and $\sqrt{1-x}$, and collected the $o(1)$ terms. In the last equality we used that when $\alpha < 1/3$ then $(1-r_n) = O(n^{-(1-3\alpha)/2})$. As before we can argue that $\tilde{s}_{2n} \leq 1$ for n large. From this we have

$$
\begin{aligned}
\frac{1}{b_n \cdot \pi_n(\lfloor \underline{q} \rfloor)} \cdot \sum_{k=\lfloor \underline{q} \rfloor + b_n + 1}^{n-1} \pi_n(k) &\leq \frac{s_{2n}^{b_n+1}}{b_n \cdot (1-s_{2n})} \leq \frac{\tilde{s}_{2n}^{b_n}}{b_n \cdot (1-\tilde{s}_{2n})} \\
&\to \frac{\exp\Big(-\frac{1}{2}\Big(C^2 - \mathbf{1}_{\{\alpha=1/3\}} \cdot \frac{C^2}{2(\beta \cdot r(\beta))^{3/2}} \Big) \Big)}{\frac{1}{2}\Big(C^2 - \mathbf{1}_{\{\alpha=1/3\}} \cdot \frac{C^2}{2(\beta \cdot r(\beta))^{3/2}} \Big)}.
\end{aligned}
$$

Finally, since this limit is an upper bound we obtain the desired upper bound for the lim sup.

Remaining proofs. Let

$$h_n(x) = \frac{1}{x} \cdot (1 + \frac{1}{\sqrt{n-x}}),$$

we show is decreasing in $(0, \underline{q} + b_n]$ for n large. First,

$$\frac{dh_n}{dx}(x) = -\frac{1}{x^2} \cdot \Big(1 + \frac{1}{\sqrt{n-x}} \Big) + \frac{1}{2x}(n-x)^{-3/2},$$

so $h_n(x)$ is decreasing if and only if $x \leq 2((n-x)^{3/2} + n - x)$. Note that the LHS in the previous inequality is strictly increasing and the RHS is strictly decreasing. Also, at $x = 0$ the LHS is below the RHS, and for $x = n$ the converse is true. Therefore, there if for some y,

$$\frac{y}{2} \leq (n-y)^{3/2} \cdot \Big(1 + \frac{1}{\sqrt{n-y}}\Big) \tag{B.12}$$

then the same is true for all $x \leq y$. Consider $y = \underline{q} + b_n$ and let $\ell_n = 1 - \frac{b_n}{n-\underline{q}}$

$$\begin{aligned}(n - \underline{q} - b_n)^{3/2} \cdot \Big(1 + \frac{1}{\sqrt{n-\underline{q}-b_n}}\Big) &= (n-\underline{q})^{3/2}\ell_n^{3/2} \cdot \Big(1 + \frac{1}{\sqrt{\ell_n}}\frac{1}{\sqrt{n-\underline{q}}}\Big) \\ &\overset{\text{Eq.(B.11)}}{=} (n-\underline{q})^{3/2}\ell_n^{3/2} \cdot \Big(1 + \frac{1}{\sqrt{\ell_n}} \cdot \frac{(\underline{q} - n\rho_n)}{n\rho_n}\Big) \\ &= (n-\underline{q})^{3/2}\ell_n \cdot \Big(\frac{n\rho_n(\sqrt{\ell_n} - 1) + \underline{q}}{n\rho_n}\Big),\end{aligned}$$

note that for n large enough $n\rho_n(\sqrt{\ell_n} - 1) + \underline{q} > 0$. Then Eq. (B.12) is satisfied if and only if

$$\frac{\rho_n}{2} \cdot \underbrace{\Big[\frac{\underline{q} + b_n}{\ell_n(n\rho_n(\sqrt{\ell_n} - 1) + \underline{q})}\Big]}_{H_n} \leq \frac{(n-\underline{q})^{3/2}}{n} = n^{(1-3\alpha)/2}(n^{\alpha}(1-\rho_n)r_{0,n}(\rho_n))^{3/2}, \tag{B.13}$$

where we used that $n - \underline{q} = n(1-\rho_n)r_{0,n}(\rho_n)$. Since $\ell_n \to 1$, $H_n \to 1$. If $\alpha < 1/3$ then for n large enough the previous inequality hold. If $\alpha = 1/3$ and $\beta > \beta_1^*$, the LHS in Eq (B.13) converges to 1/2 and the RHS to $(\beta r(\beta))^{3/2}$. This last function is strictly increasing and equal to 1/2 at $\beta = \beta_1^*$. This implies that for n large enough Eq. (B.13) is satisfied, completing the proof.

□

Proof of Lemma B.2. We start we the lower bound, let $b_n = Cn^{\frac{3\alpha}{2}}$ and note that

$$
\begin{aligned}
\frac{\mathbf{P}[Q_n(\infty) \geq n]}{b_n \pi_n(\lfloor \overline{q}_n \rfloor)} &= \frac{1}{b_n} \sum_{k=n}^{\infty} \frac{\pi_n(k)}{\pi_n(\lfloor \overline{q}_n \rfloor)} \\
&\geq \frac{1}{b_n} \sum_{k=\lfloor \overline{q}_n \rfloor - b_n}^{\lfloor \overline{q}_n \rfloor + b_n} \frac{\pi_n(k)}{\pi_n(\lfloor \overline{q}_n \rfloor)} \\
&= \frac{1}{b_n} \sum_{k=\lfloor \overline{q}_n \rfloor - b_n}^{\lfloor \overline{q}_n \rfloor} \prod_{\ell=k+1}^{\lfloor \overline{q}_n \rfloor} \frac{1}{\rho_n} \frac{1}{\left(1 + \frac{1}{\sqrt{\ell - n}}\right)} + \frac{1}{b_n} \sum_{k=\lfloor \overline{q}_n \rfloor + 1}^{\lfloor \overline{q}_n \rfloor + b_n} \prod_{\ell=\lfloor \overline{q}_n \rfloor + 1}^{k} \rho_n \left(1 + \frac{1}{\sqrt{\ell - n}}\right) \\
&\geq \frac{1}{b_n} \sum_{k=\lfloor \overline{q}_n \rfloor - b_n}^{\lfloor \overline{q}_n \rfloor} \Bigg[\underbrace{\frac{1}{\rho_n} \frac{1}{\left(1 + \frac{1}{\sqrt{\lfloor \overline{q}_n \rfloor - b_n - n}}\right)}}_{s_{1n}} \Bigg]^{\lfloor \overline{q}_n \rfloor - k} \\
&\quad + \frac{1}{b_n} \sum_{k=\lfloor \overline{q}_n \rfloor + 1}^{\lfloor \overline{q}_n \rfloor + b_n} \Bigg[\underbrace{\rho_n \left(1 + \frac{1}{\sqrt{\lfloor \overline{q}_n \rfloor + b_n - n}}\right)}_{s_{2n}} \Bigg]^{k - \lfloor \overline{q}_n \rfloor} \qquad \text{(B.14)} \\
&= \frac{1}{b_n} \cdot \frac{1 - s_{1n}^{b_n+1}}{1 - s_{1n}} + \frac{1}{b_n} \cdot \frac{s_{2n} - s_{2n}^{b_n+1}}{1 - s_{2n}}.
\end{aligned}
$$

Next we compute limits for $b_n(1 - s_{1n})$ and $b_n(1 - s_{2n})$. Before we begin note that $\overline{q}_n = n + z_n^2$ where $z_n^2 = \rho_n^2/(1-\rho_n)^2$ and let $\tilde{b}_n = b_n + (\overline{q}_n - \lfloor \overline{q}_n \rfloor)$ then

$$
\begin{aligned}
b_n(1 - s_{1n}) &= \frac{b_n}{\tilde{b}_n} \left[\tilde{b}_n - \tilde{b}_n \cdot \frac{1}{\rho_n} \frac{1}{\left(1 + \frac{1}{\sqrt{\lfloor \overline{q}_n \rfloor - b_n - n}}\right)} \right] \\
&= \frac{b_n}{\tilde{b}_n} \left[\tilde{b}_n - \tilde{b}_n \cdot \frac{1}{\rho_n} \left(1 - \frac{1}{\sqrt{z_n^2 - \tilde{b}_n}}\right) + o(1) \right] \\
&= \frac{b_n}{\tilde{b}_n} \left[\tilde{b}_n - \tilde{b}_n \cdot \frac{1}{\rho_n} \left(1 - \frac{1}{z_n} \left\{ 1 + \frac{\tilde{b}_n}{2z_n^2} \right\} \right) + o(1) \right] \\
&= \frac{b_n}{\tilde{b}_n} \left[\tilde{b}_n \frac{(1-\rho_n)^2}{\rho_n^2} + \frac{\tilde{b}_n^2}{2\rho_n^4} \cdot (1-\rho_n)^3 + o(1) \right] \\
&\to \frac{C^2 \beta^3}{2}.
\end{aligned}
$$

Thus,

$$
\frac{1}{b_n} \cdot \frac{1 - s_{1n}^{b_n+1}}{1 - s_{1n}} \to \frac{1 - \exp\left(-\frac{C^2\beta^3}{2}\right)}{\frac{C^2\beta^3}{2}}.
$$

For $b_n(1-s_{2n})$ we have

$$\begin{aligned}
b_n(1-s_{2n}) &= \frac{b_n}{\tilde{b}_n}\Big[\tilde{b}_n - \tilde{b}_n \cdot \rho_n\Big(1+\frac{1}{\sqrt{z_n^2+\tilde{b}_n}}\Big)\Big] \\
&= \frac{b_n}{\tilde{b}_n}\Big[\tilde{b}_n - \tilde{b}_n \cdot \rho_n\Big(1+\frac{1}{z_n}\Big\{1-\frac{\tilde{b}_n}{2z_n^2}\Big\}\Big)+o(1)\Big] \\
&= \frac{b_n}{\tilde{b}_n}\Big[\frac{\tilde{b}_n^2}{2\rho_n^2}\cdot(1-\rho_n)^3+o(1)\Big] \\
&\to \frac{C^2\beta^3}{2}.
\end{aligned}$$

Thus,

$$\frac{1}{b_n}\cdot\frac{s_{2n}-s_{2n}^{b_n+1}}{1-s_{2n}} \to \frac{1-\exp\Big(-\frac{C^2\beta^3}{2}\Big)}{\frac{C^2\beta^3}{2}}.$$

Finally, since this limit is a lower bound we obtain the desired lower bound for the lim inf.

For the upper bound note that

$$\frac{\mathbf{P}[Q_n(\infty)\geq n]}{b_n\pi_n(\lfloor\bar{q}_n\rfloor)} = \frac{1}{b_n}\sum_{k=n}^{\infty}\frac{\pi_n(k)}{\pi_n(\lfloor\bar{q}_n\rfloor)} \leq 2+\frac{1}{b_n}\sum_{k=n}^{\lfloor\bar{q}_n\rfloor-b_n}\frac{\pi_n(k)}{\pi_n(\lfloor\bar{q}_n\rfloor)}+\frac{1}{b_n}\sum_{k=\lfloor\bar{q}_n\rfloor+b_n+1}^{\infty}\frac{\pi_n(k)}{\pi_n(\lfloor\bar{q}_n\rfloor)}, \tag{B.15}$$

so we just need to upper bound both summation on the right hand side of Eq. (B.15) and take the limit. For the first summation we have

$$\begin{aligned}
\frac{1}{b_n}\sum_{k=n}^{\lfloor\bar{q}_n\rfloor-b_n}\frac{\pi_n(k)}{\pi_n(\lfloor\bar{q}_n\rfloor)} &= \frac{1}{b_n}\sum_{k=n}^{\lfloor\bar{q}_n\rfloor-b_n}\prod_{\ell=k+1}^{\lfloor\bar{q}_n\rfloor}\frac{1}{\rho_n\cdot\Big(1+\frac{1}{\sqrt{\ell-n}}\Big)} \\
&\overset{(a)}{\leq} \frac{1}{b_n}\sum_{k=n}^{\lfloor\bar{q}_n\rfloor-b_n}\Bigg[\underbrace{\frac{1}{b_n}\cdot\sum_{\ell=\lfloor\bar{q}_n\rfloor-b_n+1}^{\lfloor\bar{q}_n\rfloor-1}\frac{1}{\rho_n\cdot\Big(1+\frac{1}{\sqrt{\ell-n}}\Big)}}_{s_{1n}}\Bigg]^{\lfloor\bar{q}_n\rfloor-k} \\
&\leq \frac{1}{b_n}\frac{s_{1n}^{b_n}}{1-s_{1n}},
\end{aligned}$$

where in (a) we used the inequality of arithmetic and geometric means, and the fact that the function inside the summation is increasing. For s_{1n} we have

$$\begin{aligned} s_{1n} &= \frac{1}{\rho_n \cdot b_n} \cdot \sum_{\ell=\lfloor \overline{q}_n \rfloor - b_n + 1}^{\lfloor \overline{q}_n \rfloor - 1} \frac{1}{\left(1 + \frac{1}{\sqrt{\ell - n}}\right)} \\ &\leq \frac{1}{\rho_n \cdot b_n} \cdot \int_{\overline{q}_n - b_n}^{\overline{q}_n} \frac{1}{\left(1 + \frac{1}{\sqrt{x-n}}\right)} dx \\ &= \frac{1}{\rho_n \cdot b_n} \cdot \Big[-2\sqrt{x-n} + 2\log(\sqrt{x-n} + 1) + x - n \Big] \Big|_{\overline{q}_n - b_n}^{\overline{q}_n} \\ &= \frac{1}{\rho_n \cdot b_n} \cdot \Big[-2z_n + 2\log(z_n + 1) + 2\sqrt{z_n^2 - b_n} - 2\log(\sqrt{z_n^2 - b_n} + 1) + b_n \Big], \end{aligned}$$

then denoting the last expression above by $\tilde{s}_{1n}$ we have

$$\begin{aligned} b_n \cdot (1 - \tilde{s}_{1n}) &= b_n - \frac{1}{\rho_n} \cdot \Big[-2z_n + 2\log(z_n + 1) + 2\sqrt{z_n^2 - b_n} \\ &\quad - 2\log(\sqrt{z_n^2 - b_n} + 1) + b_n \Big] \\ &= b_n + \frac{b_n}{\rho_n z_n} + \frac{b_n^2}{4\rho_n z_n^3} - \frac{b_n}{\rho_n} + o(1) \\ &\to \frac{C^2 \beta^3}{4}. \end{aligned}$$

Hence, since (for n large) $\tilde{s}_{1n} \leq 1$ we have

$$\frac{1}{b_n} \sum_{k=n}^{\overline{q}_n - b_n} \frac{\pi_n(k)}{\pi_n(\overline{q}_n)} \leq \frac{1}{b_n} \frac{s_{1n}^{b_n}}{1 - s_{1n}} \leq \frac{1}{b_n} \frac{\tilde{s}_{1n}^{b_n}}{1 - \tilde{s}_{1n}} \to \frac{\exp\left(-\frac{C^2\beta^3}{4}\right)}{\frac{C^2\beta^3}{4}}.$$

Now let us consider the second summation in Eq. (B.15),

$$\begin{aligned} \frac{1}{b_n} \sum_{k=\lfloor \overline{q}_n \rfloor + b_n + 1}^{\infty} \frac{\pi_n(k)}{\pi_n(\lfloor \overline{q}_n \rfloor)} &= \frac{1}{b_n} \sum_{k=\lfloor \overline{q}_n \rfloor + b_n + 1}^{\infty} \prod_{\ell=\lfloor \overline{q}_n \rfloor + 1}^{k} \rho_n \Big(1 + \frac{1}{\sqrt{\ell - n}}\Big) \\ &\leq \frac{1}{b_n} \sum_{k=\lfloor \overline{q}_n \rfloor + b_n}^{\infty} \Bigg[\underbrace{\frac{1}{b_n} \cdot \sum_{\ell=\lfloor \overline{q}_n \rfloor + 1}^{\lfloor \overline{q}_n \rfloor + b_n} \rho_n \cdot \Big(1 + \frac{1}{\sqrt{\ell - n}}\Big)}_{s_{2n}} \Bigg]^{k - \lfloor \overline{q}_n \rfloor} \\ &= \frac{1}{b_n} \frac{s_{2n}^{b_n}}{1 - s_{2n}}, \end{aligned}$$

where we used the inequality of arithmetic and geometric means, the fact that the function inside the summation is decreasing, and that for app $\ell \geq \lfloor \overline{q}_n \rfloor + 1$ the terms in the summation are strictly bounded above by 1. For s_{2n}, if we let $\tilde{z}_n^2 = \lfloor \overline{q}_n \rfloor - n$, we have

$$\begin{aligned} s_{2n} &\leq \frac{\rho_n}{b_n} \cdot \int_{\lfloor \overline{q}_n \rfloor}^{\lfloor \overline{q}_n \rfloor + b_n} \Big(1 + \frac{1}{\sqrt{x-n}}\Big) dx \\ &= \frac{\rho_n}{b_n} \cdot \Big[2\sqrt{x-n} + x\Big]\Big|_{\lfloor \overline{q}_n \rfloor}^{\lfloor \overline{q}_n \rfloor + b_n} \\ &= \frac{\rho_n}{b_n} \cdot \Big[2\sqrt{\tilde{z}_n^2 + b_n} - 2\tilde{z}_n + b_n\Big], \end{aligned}$$

denoting this last term by $\tilde{s}_{2n}$ we have

$$\begin{aligned} b_n \cdot (1 - \tilde{s}_{2n}) &= b_n - \rho_n \cdot \Big[2\sqrt{\tilde{z}_n^2 + b_n} - 2\tilde{z}_n + b_n\Big] \\ &= b_n(1-\rho_n) - \rho_n \cdot \Big[\frac{b_n}{\tilde{z}_n} - \frac{b_n^2}{4\tilde{z}_n^3}\Big] + o(1) \\ &\to \frac{C^2\beta^3}{4}. \end{aligned}$$

Thus, since $\tilde{s}_{2n} \leq 1$ (for n large) we have

$$\frac{1}{b_n} \sum_{k=\lfloor \overline{q}_n \rfloor + b_n + 1}^{\infty} \frac{\pi_n(k)}{\pi_n(\lfloor \overline{q}_n \rfloor)} \leq \frac{1}{b_n} \frac{s_{2n}^{b_n}}{1 - s_{2n}} \leq \frac{1}{b_n} \frac{\tilde{s}_{2n}^{b_n}}{1 - \tilde{s}_{2n}} \to \frac{\exp\Big(-\frac{C^2\beta^3}{4}\Big)}{\frac{C^2\beta^3}{4}}.$$

Finally, since this limit is an upper bound we obtain the desired upper bound for the lim sup. □

Proof of Lemma 2.2. This result is a direct consequence of Lemmata B.1 and B.2 which were stated and proved right before the present proof. □

Proof of Proposition 2.3. Consider the following

$$\rho_n(y) = 1 - \frac{\beta_2^*}{n^{1/3}} - \frac{y}{n^{1/3}}, \quad y \in (-(\beta_2^* - \beta_1^*), (\beta_2^* - \beta_1^*)) = D.$$

Note that $n^{1/3}(1 - \rho_n(y)) = \beta_2^* + y > \beta_1^*$ and $\rho_n(y) \uparrow 1$, hence we can always find n_1 such that for all $n \geq n_1$ the leftmost equilibrium $\underline{q}$ is well defined. Note that for $\rho_n(y)$ we have

$$\begin{aligned} \log\Big(\frac{\pi_n(\lfloor \overline{q}_n \rfloor)}{\pi_n(\lfloor \underline{q} \rfloor)}\Big) &= (\lfloor \overline{q}_n \rfloor - \lfloor \underline{q} \rfloor) \log(\rho_n(y)) + \sum_{k=\lfloor \underline{q} \rfloor + 1}^{n-1} \log\Big[\frac{n}{k} \cdot \Big(1 + \frac{1}{\sqrt{n-k}}\Big)\Big] \\ &+ \sum_{k=1}^{\lfloor \overline{q}_n \rfloor - n} \log\Big[1 + \frac{1}{\sqrt{k}}\Big] + \log(2). \end{aligned} \tag{B.16}$$

Furthermore, observe that both $\underline{q}$ and $\overline{q}_n$ are continuous functions of y,

$$\underline{q}(y) = n - n(1-\rho_n(y)) \cdot r_{0,n}(\rho_n(y)) \quad \text{and} \quad \overline{q}_n(y) = n + \frac{\rho_n(y)^2}{(1-\rho_n(y))^2}.$$

Define,

$$f_n(y) \triangleq \log\Big(\frac{\pi_n(\lfloor \overline{q}_n(y) \rfloor)}{\pi_n(\lfloor \underline{q}(y) \rfloor)}\Big).$$

since we are using the floor function, $f_n(\cdot)$ might not be continuos. In the first step of this proof we show that the potential jumps of $f_n(\cdot)$ in D converge to zero (Step 1). Then we show that there exists a sequence γ_n^c such that $f_n(\gamma_n^c) \to c$ (Step 2) and $\gamma_n^c \to 0$ (Step 3).

Step 1. Fix $\epsilon > 0$. First, we prove that there exists $\tilde{n}$ such that for all $n \geq \tilde{n}$ we have that

$$\forall y \in D,\ \exists \delta > 0 \text{ such that } \forall \tilde{y}:\ |\tilde{y} - y| < \delta \Rightarrow |f_n(\tilde{y}) - f_n(y)| < \epsilon. \tag{B.17}$$

We choose $\tilde{n}$ such that for all $n \geq \tilde{n}$:

- $\sup_{z \in D} 2|\log(\rho_n(z))| \leq \epsilon/9$. This is possible because $\rho_n(z) \to 1$ uniformly in D.
- $$\sup_{z\in D}\Big|\log\Big[\frac{n}{\lfloor \underline{q}(z) \rfloor + 1}\Big]\Big| \leq \frac{\epsilon}{6}, \quad \text{and} \quad \sup_{z\in D}\Big|\log\Big(1 + \frac{1}{\sqrt{n - \lfloor \underline{q}(z) \rfloor - 1}}\Big)\Big| \leq \frac{\epsilon}{6}.$$
  This is possible because for any $z \in D$, $n/(\lfloor \underline{q}(z) \rfloor + 1) \to 1$.
- $$\sup_{z\in D}\Big|\log\Big[1 + \frac{1}{\sqrt{\lfloor \overline{q}_n(z) \rfloor - n}}\Big]\Big| \leq \frac{\epsilon}{3}.$$
  This is possible because for any $z \in D$, $(\lfloor \overline{q}_n(z) \rfloor - n) \uparrow \infty$.

Let $n \geq n_1$ and fix $y \in D$, we consider the first three terms in $f_n(\cdot)$, see Eq. (B.16). Let $Q_n(\tilde{y}) = \lfloor \overline{q}_n(\tilde{y}) \rfloor - \lfloor \underline{q}(\tilde{y}) \rfloor$ and $R_n(\tilde{y}) = \overline{q}_n(\tilde{y}) - \underline{q}(\tilde{y})$, and note that

$|Q_n(\tilde{y}) - R_n(\tilde{y})| \leq 2$ for any $\tilde{y}$. Also, $R_n(\tilde{y}) \log(\rho_n(\tilde{y}))$ is continuous; therefore, there exists δ_1 such that

$$|R_n(\tilde{y}) \log(\rho_n(\tilde{y})) - R_n(y) \log(\rho_n(y))| \leq \epsilon/9, \quad \forall \tilde{y} : \ |\tilde{y} - y| < \delta_1.$$

Using this, for the first term in Eq. (B.16), we have

$$\begin{aligned}
\Big| Q_n(\tilde{y}) \log(\rho_n(\tilde{y})) - Q_n(y) \log(\rho_n(y)) \Big| &= \Big| (Q_n(\tilde{y}) - R_n(\tilde{y})) \log(\rho_n(\tilde{y})) + R_n(\tilde{y}) \log(\rho_n(\tilde{y})) \\
&\quad - (Q_n(y) - R_n(y)) \log(\rho_n(y)) - R_n(y) \log(\rho_n(y)) \Big| \\
&\leq 2|\log(\rho_n(\tilde{y}))| + 2|\log(\rho_n(y))| \\
&\quad + |R_n(\tilde{y}) \log(\rho_n(\tilde{y})) - R_n(y) \log(\rho_n(y))| \\
&\leq \frac{\epsilon}{3},
\end{aligned}$$

for all $\tilde{y}$ such that $|\tilde{y} - y| < \delta_1$. For the second term in Eq. (B.16), observe that since $\underline{q}(\cdot)$ is continuous there always exists $\delta_2 > 0$ such that for all $\tilde{y}$ with $|\tilde{y} - y| < \delta_2$ we have $|\lfloor \underline{q}(\tilde{y}) \rfloor - \lfloor \underline{q}(y) \rfloor| \leq 1$. Therefore,

$$\begin{aligned}
&\Big| \sum_{k=\lfloor \underline{q}(\tilde{y}) \rfloor + 1}^{n-1} \log \Big[\frac{n}{k} \cdot \Big(1 + \frac{1}{\sqrt{n-k}} \Big) \Big] \\
&- \sum_{k=\lfloor \underline{q}(y) \rfloor + 1}^{n-1} \log \Big[\frac{n}{k} \cdot \Big(1 + \frac{1}{\sqrt{n-k}} \Big) \Big] \Big| \leq \Big| \log \Big[\frac{n}{\lfloor \underline{q}(\tilde{y}) \rfloor + 1} \cdot \Big(1 + \frac{1}{\sqrt{n - \lfloor \underline{q}(\tilde{y}) \rfloor - 1}} \Big) \Big] \Big| \\
&\leq \Big| \log \Big[\frac{n}{\lfloor \underline{q}(\tilde{y}) \rfloor + 1} \Big] \Big| \\
&+ \Big| \log \Big(1 + \frac{1}{\sqrt{n - \lfloor \underline{q}(\tilde{y}) \rfloor - 1}} \Big) \Big| \\
&\leq \frac{\epsilon}{3}.
\end{aligned}$$

Finally, for the third term in Eq. (B.16), since $\overline{q}_n(\cdot)$ is continuous there always exists $\delta_3 > 0$ such that for all $\tilde{y}$ with $|\tilde{y} - y| < \delta_3$ we have $|\lfloor \overline{q}_n(\tilde{y}) \rfloor - \lfloor \overline{q}_n(y) \rfloor| \leq 1$. Therefore,

$$\Big| \sum_{k=1}^{\lfloor \overline{q}_n(\tilde{y}) \rfloor - n} \log \Big[1 + \frac{1}{\sqrt{k}} \Big] - \sum_{k=1}^{\lfloor \overline{q}_n(y) \rfloor - n} \log \Big[1 + \frac{1}{\sqrt{k}} \Big] \Big| \leq \Big| \log \Big[1 + \frac{1}{\sqrt{\lfloor \overline{q}_n(y) \rfloor - n}} \Big] \Big| \leq \frac{\epsilon}{3}$$

Putting the three inequalities jus proved together, for $\delta \leq \min\{\delta_1, \delta_2, \delta_3\}$, delivers Eq. (B.17). Next define

$$\Delta_n \triangleq \sup_{y \in D} |f_n(y^+) - f_n(y^-)|,$$

then Eq. (B.17) ensures that $\Delta_n \to 0$.

Step 2. We construct γ_n^c and show that $f_n(\gamma_n^c) \to c$. Fix $y_1 \in (-(\beta_2^* - \beta_1^*), 0)$ and $y_2 \in (0, \beta_2^* - \beta_1^*)$, we next argue that there exists n_2 such that for all $n \geq n_2$ it holds that $f_n(y_1) > c > f_n(y_2)$. Indeed, consider first y_1 and note that $\beta_2^* + y_1 \in (\beta_1^*, \beta_2^*)$. For $g(\cdot)$ as in Proposition 2.2 part *iii)*, one has $g(\beta_2^* + y_1) > 0$. So, again by Proposition 2.2 part *iii)* we have that for any $\epsilon_1 \in (0, g(\beta_2^* + y_1))$ there exists $n_{1,2}$ such that for all $n \geq n_{1,2}$ we have

$$c < n^{1/3} \cdot (g(\beta_1) - \epsilon_1) < f_n(y_1).$$

A similar argument that leverages the fact that $g(\beta_2^* + y_2) < 0$ shows that there exists $n_{2,2}$ such that for all $n \geq n_{2,2}$ we have $f_n(y_2) < c$. We take $n_2 = \max\{n_{1,2}, n_{2,2}\}$ to conclude that for all $n \geq n_2$ it holds that $f_n(y_1) > c > f_n(y_2)$. To conclude consider $n \geq \max\{n_1, n_2\}$ then, by Step 1 we can always find $\gamma_n^c \in (y_1, y_2)$ such that

$$c - \frac{\Delta_n}{2} \leq f_n(\gamma_n^c) \leq c + \frac{\Delta_n}{2}$$

Taking limit at both sides and using that $\Delta_n \to 0$, we conclude that $f_n(\gamma_n^c) \to c$.

Step 3. To conclude the proof we need to argue that $\gamma_n^c \to 0$. Note from the argument above $\{\gamma_n^c\}$ is a bounded sequence. For the sake of contradiction fix $\epsilon > 0$ and suppose that

$$\limsup_{n \to \infty} \gamma_n^c > \epsilon.$$

This implies that there exists a subsequence $\{\gamma_{k(n)}^c\}$ that convergences to a point $\hat{\gamma}^c \geq \epsilon$. Let

$$\hat{\rho}_n = 1 - \frac{\beta_2^*}{n^{1/3}} - \frac{\gamma_n^c}{n^{1/3}},$$

then $k(n)^{1/3}(1-\hat{\rho}_{k(n)}) \to \beta_2^\star + \hat{\gamma}^c$. Because $g(\beta_2^\star + \hat{\gamma}^c) < 0$ from Proposition 2.2, for $\epsilon' > 0$ such that $g(\beta_2^\star + \hat{\gamma}^c) + \epsilon' < 0$, we can deduce that for all n large enough

$$f_{k(n)}(\gamma^c_{k(n)}) \leq n^{1/3}(g(\beta_2^\star + \hat{\gamma}^c) + \epsilon') \leq c - \epsilon'.$$

However, from Step 1 we know that $f_{k(n)}(\gamma^c_{k(n)}) \to c$. This, together with the previous inequality yields a contradiction. The case when $\liminf_{n\to\infty} \gamma_n^c < \epsilon$ can be treated similarly and is thus omitted. Therefore, for any $\epsilon > 0$

$$\epsilon \leq \liminf_{n\to\infty} \gamma_n^c \leq \limsup_{n\to\infty} \gamma_n^c \leq \epsilon,$$

since ϵ is arbitrary we have that $\gamma_n^c \to 0$, which concludes the proof.

□

Proof of Proposition 2.4. We prove both statement separately.

(i) We show that

$$\lim_{n\to\infty} \mathbf{P}[Q_n(\infty) < \lfloor \bar{q}_n \rfloor - C \cdot \sqrt{\log(n)} \cdot n^{1.5\alpha}] = 0,$$

the other case in analogous. To reduce notation let $b_n = C \cdot \sqrt{\log(n)} \cdot n^{1.5\alpha}$ for some $C > 0$ that we will choose later in the proof then

$$\begin{aligned}\mathbf{P}[Q_n(\infty) < \lfloor \bar{q}_n \rfloor - b_n] &\leq \mathbf{P}[Q_n(\infty) < n] + \mathbf{P}[n \leq Q_n(\infty) \leq \lfloor \bar{q}_n \rfloor - b_n] \\ &= \mathbf{P}[Q_n(\infty) < n] + \sum_{k=n}^{\lfloor \bar{q}_n \rfloor - b_n} \pi_n(k)\end{aligned}$$

by Theorem 2.2 part *(i)* the first term converges to zero. For the second term we have the following upper bound

$$\begin{aligned}
\sum_{k=n}^{\lfloor \bar{q}_n \rfloor - b_n} \pi_n(k) &\le \sum_{k=n}^{\lfloor \bar{q}_n \rfloor - b_n} \prod_{\ell=k+1}^{\lfloor \bar{q}_n \rfloor} \frac{1}{\rho_n \cdot \left(1 + \frac{1}{\sqrt{\ell - n}}\right)} \\
&\overset{(a)}{\le} \sum_{k=n}^{\lfloor \bar{q}_n \rfloor - b_n} \left[\frac{1}{\lfloor \bar{q}_n \rfloor - k} \cdot \sum_{\ell=k+1}^{\lfloor \bar{q}_n \rfloor - 1} \frac{1}{\rho_n \cdot \left(1 + \frac{1}{\sqrt{\ell - n}}\right)} \right]^{\lfloor \bar{q}_n \rfloor - k} \\
&\le \sum_{k=n}^{\lfloor \bar{q}_n \rfloor - b_n} \Bigg[\underbrace{\frac{1}{b_n} \cdot \sum_{\ell=\lfloor \bar{q}_n \rfloor - b_n + 1}^{\lfloor \bar{q}_n \rfloor - 1} \frac{1}{\rho_n \cdot \left(1 + \frac{1}{\sqrt{\ell - n}}\right)}}_{s_n} \Bigg]^{\lfloor \bar{q}_n \rfloor - k} \\
&\le \frac{s_n^{b_n}}{1 - s_n},
\end{aligned}$$

where in (a) we used the inequality of arithmetic and geometric means. We next show the last term above converges to zero.

Recall that $\bar{q}_n = n + z_n^2$ where $z_n = \frac{\rho_n}{(1-\rho_n)}$. We have

$$\begin{aligned}
s_n &= \frac{1}{\rho_n \cdot b_n} \cdot \sum_{\ell=\lfloor \bar{q}_n \rfloor - b_n + 1}^{\lfloor \bar{q}_n \rfloor - 1} \frac{1}{\left(1 + \frac{1}{\sqrt{\ell - n}}\right)} \\
&\le \frac{1}{\rho_n \cdot b_n} \cdot \int_{\bar{q}_n - b_n}^{\bar{q}_n} \frac{1}{\left(1 + \frac{1}{\sqrt{x - n}}\right)} dx \\
&= \frac{1}{\rho_n \cdot b_n} \cdot \Big[-2\sqrt{x-n} + 2\log(\sqrt{x-n} + 1) + x - n \Big] \Big|_{\bar{q}_n - b_n}^{\bar{q}_n} \\
&= \frac{1}{\rho_n \cdot b_n} \cdot \Big[-2z_n + 2\log(z_n + 1) + 2\sqrt{z_n^2 - b_n} - 2\log(\sqrt{z_n^2 - b_n} + 1) + b_n \Big],
\end{aligned}$$

denote this last term by $\tilde{s}_n$. Then

$$\begin{aligned}
\tilde{s}_n &= \frac{1}{\rho_n \cdot b_n} \cdot \Big[-2z_n + 2\Big(\frac{1}{z_n} + O(n^{-2\alpha})\Big) + 2z_n \Big(1 - \frac{b_n}{2z_n^2} - \frac{b_n^2}{8z_n^4} + O\Big(\frac{b_n^3}{z_n^6}\Big)\Big) \\
&\quad - 2\Big(\sqrt{1 - \frac{b_n}{z_n^2}} + \frac{1}{z_n} - 1 + O\Big(\frac{b_n^2}{z_n^4}\Big)\Big) + b_n \Big] \\
&= \frac{1}{\rho_n \cdot b_n} \cdot \Big[2z_n \Big(- \frac{b_n}{2z_n^2} - \frac{b_n^2}{8z_n^4} \Big) - 2\Big(- \frac{b_n}{2z_n^2} \Big) + b_n \Big] + O(n^{-2\alpha} \log(n)) \\
&= \Big[1 + \frac{(1-\rho_n)^3}{\rho_n^3} - \frac{b_n (1-\rho_n)^3}{4\rho_n^4} \Big] + O(n^{-2\alpha} \log(n)).
\end{aligned}$$

Hence, $\tilde{s}_n \to 1$ and

$$b_n \cdot (1 - \tilde{s}_n) = \underbrace{\frac{(1-\rho_n)^3}{\rho_n^3}}_{O(n^{-3\alpha})} \cdot \underbrace{\frac{b_n^2}{4\rho_n}}_{O(n^{3\alpha}\log(n))} + O(n^{-\alpha/2}\log(n)^{3/2}) = O(\log(n)). \tag{B.18}$$

From this we can deduce that $b_n \cdot (1 - \tilde{s}_n) \to +\infty$ (which implies that $\tilde{s}_n \leq 1$) and

$$b_n \cdot (1-\tilde{s}_n)^2 = O(\log(n)) \cdot (1-\tilde{s}_n) = O(\log(n)) \cdot O(n^{-3\alpha/2}\sqrt{\log(n)}) \longrightarrow 0 \quad \text{as } n \to \infty.$$

Putting all this together yields, for n large enough,

$$\begin{aligned}
\sum_{k=n}^{\lfloor \bar{q}_n \rfloor - b_n} \pi_n(k) &\leq \frac{s_n^{b_n}}{1 - s_n} \\
&\leq \frac{\tilde{s}_n^{b_n}}{1 - \tilde{s}_n} \\
&= \frac{\exp\Big(- b_n \cdot (1 - \tilde{s}_n) + O(b_n(1-\tilde{s}_n)^2) \Big)}{1 - \tilde{s}_n} \\
&\overset{Eq.(\text{B.18})}{=} \frac{\exp\Big(- \frac{(1-\rho_n)^3 b_n^2}{4\rho_n^4} + O(n^{-\alpha/2}\log(n)^{3/2}) \Big)}{1 - \tilde{s}_n} \\
&= \frac{n^{-\frac{n^{3\alpha}(1-\rho_n)^3 C^2}{4\rho_n^4}} \exp\Big(O(n^{-\alpha/2}\log(n)^{3/2}) \Big)}{1 - \tilde{s}_n},
\end{aligned}$$

observe that the exponential term above converges to 1. Also, the denominator is $O(n^{-3\alpha/2}\sqrt{\log(n)})$ while $\frac{n^{3\alpha}(1-\rho_n)^3 C^2}{4\rho_n^4} \to \beta^3 C^2/4$. So if we choose C such that $\beta^3 C^2/4 > 3\alpha/2$ then we have that

$$\lim_{n\to\infty} \sum_{k=n}^{\lfloor \bar{q}_n \rfloor - b_n} \pi_n(k) = 0,$$

as desired.

(ii) We show that

$$\lim_{n\to\infty} \mathbf{P}[Q_n(\infty) < \lfloor \underline{q} \rfloor - C \cdot \sqrt{\log(n)} \cdot \sqrt{n}] = 0,$$

the other case is analogous. To reduce notation let $b_n = C \cdot \sqrt{\log(n) \cdot n}$ for some $C > 0$ that we identify later then

$$\begin{aligned}
\mathbf{P}[Q_n(\infty) < \lfloor \underline{q} \rfloor - b_n] &= \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \pi_n(k) \\
&= \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \prod_{\ell=k+1}^{\lfloor \underline{q} \rfloor} \frac{1}{\rho_n} \cdot \frac{\ell}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-\ell}}\right)} \\
&\overset{(a)}{\leq} \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \prod_{\ell=k+1}^{\lfloor \underline{q} \rfloor - 1} \frac{1}{\rho_n} \cdot \frac{\ell}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-\ell}}\right)} \\
&\overset{(b)}{\leq} \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \left\{ \frac{1}{\lfloor \underline{q} \rfloor - k - 1} \cdot \sum_{\ell=k+1}^{\lfloor \underline{q} \rfloor - 1} \frac{1}{\rho_n} \cdot \frac{\ell}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-\ell}}\right)} \right\}^{\lfloor \underline{q} \rfloor - k - 1} \\
&\overset{(c)}{\leq} \sum_{k=0}^{\lfloor \underline{q} \rfloor - b_n - 1} \Bigg\{ \underbrace{\frac{1}{b_n} \cdot \sum_{\ell=\lfloor \underline{q} \rfloor - b_n}^{\lfloor \underline{q} \rfloor - 1} \frac{1}{\rho_n} \cdot \frac{\ell}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-\ell}}\right)}}_{s_{1n}} \Bigg\}^{\lfloor \underline{q} \rfloor - k - 1} \\
&\leq \frac{s_{1n}^{b_n}}{(1 - s_{1n})},
\end{aligned}$$

where in (a) we use that

$$\frac{1}{\rho_n} \cdot \frac{\lfloor \underline{q} \rfloor}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n - \lfloor \underline{q} \rfloor}}\right)} \leq 1,$$

in (b) the inequality of arithmetic and geometric means, and in (c) the fact that the term we are summing in the second summation is decreasing in ℓ is decreasing for $\ell \leq \underline{q} + b_n$. In order to simplify notation let $\tilde{z}_n^2 = n - \lfloor \underline{q} \rfloor$. Let us analyze s_{1n},

$$
\begin{aligned}
s_{1n} &= \frac{1}{b_n} \cdot \sum_{\ell=\lfloor \underline{q} \rfloor - b_n}^{\lfloor \underline{q} \rfloor - 1} \frac{1}{\rho_n} \cdot \frac{\ell}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-\ell}}\right)} \\
&\leq \frac{1}{b_n} \cdot \int_{\lfloor \underline{q} \rfloor - b_n}^{\lfloor \underline{q} \rfloor} \frac{1}{\rho_n} \cdot \frac{x}{n} \cdot \frac{1}{\left(1 + \frac{1}{\sqrt{n-x}}\right)} dx \\
&= \frac{1}{b_n n \rho_n} \cdot \Bigg[\frac{1}{6} \Big(-3n^2 + n(8\sqrt{n-x} + 6) + 4x\sqrt{n-x} - 12\sqrt{n-x} + 3x^2 - 6x \Big) \\
&\quad - 2(n-1)\log(\sqrt{n-x} + 1) \Bigg] \Bigg|_{\lfloor \underline{q} \rfloor - b_n}^{\lfloor \underline{q} \rfloor} \\
&= \frac{1}{b_n n \rho_n} \cdot \Bigg[\frac{12n - 4\tilde{z}_n^2 - 12}{6} \Big(\tilde{z}_n - \sqrt{\tilde{z}_n^2 + b_n} \Big) + \frac{4}{6} b_n \sqrt{\tilde{z}_n^2 + b_n} + (n - \tilde{z}_n^2) b_n \\
&\quad - \frac{b_n^2}{2} - 2(n-1)\log\left(\frac{\tilde{z}_n + 1}{\sqrt{\tilde{z}_n^2 + b_n} + 1} \right) \Bigg],
\end{aligned}
$$

If we denote this last expression $\tilde{s}_{1n}$ then for $(1-\tilde{s}_{1n})$ we have that

$$\begin{aligned}
(1-\tilde{s}_{1n}) &= 1 - \frac{1}{b_n n \rho_n}\Bigg[\frac{12n - 4\tilde{z}_n^2 - 12}{6}\tilde{z}_n\Big(1-\sqrt{1+\frac{b_n}{\tilde{z}_n^2}}\Big) + (n-\tilde{z}_n^2)b_n - \frac{b_n^2}{2}\Bigg] \\
&+ o\Big(\sqrt{\frac{1}{n\log(n)}}\Big) \\
&= 1 - \frac{1}{b_n n \rho_n}\Bigg[\frac{12n - 4\tilde{z}_n^2 - 12}{6}\Big(-\frac{b_n}{2\tilde{z}_n} + \frac{b_n^2}{8\tilde{z}_n^3}\Big) + (n-\tilde{z}_n^2)b_n - \frac{b_n^2}{2}\Bigg] \\
&+ o\Big(\sqrt{\frac{1}{n\log(n)}}\Big) \\
&= 1 - \frac{1}{b_n n \rho_n}\Bigg[2n\Big(-\frac{b_n}{2\tilde{z}_n} + \frac{b_n^2}{8\tilde{z}_n^3}\Big) + (n-\tilde{z}_n^2)b_n - \frac{b_n^2}{2}\Bigg] \\
&+ o\Big(\sqrt{\frac{1}{n\log(n)}}\Big) \\
&\overset{\text{Eq.(B.11)}}{=} 1 - \Big(1-\frac{\tilde{z}_n^2}{n}\Big)\frac{1}{\rho_n} + \Big(\frac{(1-\rho_n)}{\rho_n} - \frac{z_n^2}{n\rho_n}\Big)\frac{1}{\rho_n}\cdot\frac{z_n}{\tilde{z}_n} - \frac{b_n}{4\rho_n\tilde{z}_n^3} + \frac{b_n}{2\rho_n n} \\
&+ o\Big(\sqrt{\frac{1}{n\log(n)}}\Big) \\
&= -\underbrace{\frac{b_n}{4\rho_n\tilde{z}_n^3}}_{O(\sqrt{\frac{\log(n)}{n^{2-3\alpha}}})} + \underbrace{\frac{b_n}{2\rho_n n}}_{O(\sqrt{\frac{\log(n)}{n}})} + o\Big(\sqrt{\frac{1}{n\log(n)}}\Big),
\end{aligned}$$

hence, $\tilde{s}_n \to 1$ and

$$b_n \cdot (1-\tilde{s}_n) = b_n \cdot \Big(-\frac{b_n}{4\rho_n\tilde{z}_n^3} + \frac{b_n}{2\rho_n n} + o\Big(\sqrt{\frac{1}{n\log(n)}}\Big)\Big) = O(\log(n)). \tag{B.19}$$

From this we can deduce that $b_n \cdot (1-\tilde{s}_n) \to +\infty$ (which implies that $\tilde{s}_n \leq 1$) and

$$b_n \cdot (1-\tilde{s}_n)^2 = O(\log(n)) \cdot (1-\tilde{s}_n) = O(\log(n)) \cdot O(\sqrt{\frac{\log(n)}{n}}) \longrightarrow 0 \quad \text{as } n \to \infty.$$

Putting all this together yields, for n large enough

$$\begin{aligned}
\sum_{k=0}^{\lfloor q \rfloor - b_n - 1} \pi_n(k) &\leq \frac{s_n^{b_n}}{1-s_n} \\
&\leq \frac{\tilde{s}_n^{b_n}}{1-\tilde{s}_n} \\
&= \frac{\exp\Big(- b_n \cdot (1-\tilde{s}_n) + O(b_n(1-\tilde{s}_n)^2)\Big)}{1-\tilde{s}_n} \\
&\overset{Eq.(B.19)}{=} \frac{\exp\Big(- \Big(- \frac{b_n^2}{4\rho_n \tilde{z}_n^3} + \frac{b_n^2}{2\rho_n n}\Big) + O(\sqrt{\frac{\log(n)^3}{n}})\Big)}{1-\tilde{s}_n} \\
&= \frac{n^{-\Big(-\frac{C^2 n}{4\rho_n \tilde{z}_n^3} + \frac{C^2 n}{2\rho_n n}\Big)} \exp\Big(O(\sqrt{\frac{\log(n)^3}{n}})\Big)}{1-\tilde{s}_n},
\end{aligned}$$

observe that the exponential term above converges to 1. Also, the denominator is $O(\sqrt{\frac{\log(n)}{n}})$ while

$$-\frac{C^2 n}{4\rho_n \tilde{z}_n^3} + \frac{C^2 n}{2\rho_n n} \to \frac{C^2}{2}\Big(1 - \mathbf{1}_{\{\alpha=1/3\}} \cdot \frac{1}{2(\beta \cdot r(\beta))^{3/2}}\Big),$$

where $r(\beta) = \lim_{n\to\infty} r_{0,n}(\rho_n)$, and the term in brackets in the expression above is strictly positive when $\alpha = 1/3$ and $\beta > \beta_1^*$. So if we choose C such that

$$\frac{C^2}{2}\Big(1 - \mathbf{1}_{\{\alpha=1/3\}} \cdot \frac{1}{2(\beta \cdot r(\beta))^{3/2}}\Big) > \frac{1}{2}$$

then we have that

$$\lim_{n\to\infty} \sum_{k=0}^{\lfloor q \rfloor - b_n - 1} \pi_n(k) = 0,$$

as desired. $\square$

Appendix C

The Scope of Sequential Screening With Ex-Post Participation Constraints

C.1 Proofs for Section 3.5

Proof of Lemma 3.1. The proof of this result is standard and thus omitted. □

Proof of Lemma 3.2. The fact that the optimal solution is a threshold allocation is explained in the main text. Thus, we only need to provide a proof for $\widehat{\theta}$ being in the interval $[\widehat{\theta}_1, \widehat{\theta}_K]$. Note that for all θ below $\widehat{\theta}_1$, $\mu_k(\theta)$ is negative for all $k \in \{1, \ldots, K\}$. Therefore, $\bar{\mu}(\theta)$ is negative for all θ below $\widehat{\theta}_1$. Similarly, for all θ above $\widehat{\theta}_K$, $\bar{\mu}(\theta)$ is positive. Since the allocation is of the threshold type, it is optimal to set $x(\theta)$ equal to 0 for θ below $\widehat{\theta}_1$ and to set $x(\theta)$ equal to 1 for θ above $\widehat{\theta}_K$. This necessarily implies that $\widehat{\theta}$ is in $[\widehat{\theta}_1, \widehat{\theta}_K]$. □

Proof of Theorem 3.1. We first show the sufficiency of our condition and then its necessity. We denote by Ω the space of non-decreasing allocations, that is,

$$\Omega \triangleq \{x : [0,1] \to [0,1] : x(\cdot) \text{ is non-decreasing}\}.$$

Sufficiency. We assume condition (APR) holds, we want to verify the static contract is optimal. In order to do so we dualize the IC constraints. The Lagrangian is

$$\begin{aligned}\mathcal{L}(\boldsymbol{u}, \mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{w}) &= u_L(w_L - \lambda_{HL} - \alpha_L) + u_H(\lambda_{HL} - \alpha_H + w_H) \\ &+ \int_0^{\theta_{max}} x_L(z) \cdot \Big[\alpha_L \mu_L(z) f_L(z) - \lambda_{HL} \overline{F}_H(z)\Big] dz \\ &+ \int_0^{\theta_{max}} x_H(z) \cdot \Big[\alpha_H \mu_H(z) f_H(z) + \lambda_{HL} \overline{F}_H(z)\Big] dz,\end{aligned}$$

where w_L, w_H correspond to the multipliers for the ex-post IR constraints, and $\boldsymbol{\lambda} \in \{\lambda_{HL}, \lambda_{LH}\}$ to the multipliers for IC constraints. In the Lagrangian above we have chosen the multipliers as follows

$$w_L = \alpha_L - \alpha_H r^{HH}(\widehat{\theta}), \;\; w_H = \alpha_H + \alpha_H r^{HH}(\widehat{\theta}), \;\; \lambda_{HL} = \alpha_L r^{LH}(\widehat{\theta}), \;\; \lambda_{LH} = 0, \quad \text{(C.1)}$$

these multipliers are non-negative because $r^{HH}(\widehat{\theta}) \leq 0$, $r^{LH}(\widehat{\theta}) \geq 0$ and

$$w_H = \alpha_H + \alpha_H r^{HH}(\widehat{\theta}) \geq 0 \Leftrightarrow r^{HH}(\widehat{\theta}) \geq -1 \Leftrightarrow [\widehat{\theta} - \frac{\overline{F}_H}{f_H}(\widehat{\theta})] \geq -\frac{\overline{F}_H}{f_H}(\widehat{\theta}) \Leftrightarrow \widehat{\theta} \geq 0.$$

Hence, maximizing the Lagrangian over non-decreasing allocation x_L and x_H yields an upper bound for the relaxed problem. Note that this choice of multipliers eliminates the u_L and u_H terms in the Lagrangian. We next show that under (APR) the solution to the Lagrangian relaxation is the static solution. We first claim that

$$\max_{x_L \in \Omega} \int_0^{\theta_{max}} x_L(z) \cdot \Big[\alpha_L \mu_L(z) f_L(z) - \lambda_{HL} \overline{F}_H(z)\Big] dz = \int_{\widehat{\theta}}^{\theta_{max}} \Big[\alpha_L \mu_L(z) f_L(z) - \lambda_{HL} \overline{F}_H(z)\Big] dz. \quad \text{(C.2)}$$

To prove this first note that the optimal solution x_L on the left hand side of (C.2) must be of the threshold type, that is, $x_L(\theta) = \mathbf{1}_{\{\theta \geq \theta^\star\}}$, because $x_L(\cdot)$ is non-decreasing (see, e.g., [52] or [57]). Hence (C.2) is equivalent to

$$\int_{\theta^\star}^{\theta_{max}} \Big[\alpha_L \mu_L(z) f_L(z) - \lambda_{HL} \overline{F}_H(z)\Big] dz \leq \int_{\widehat{\theta}}^{\theta_{max}} \Big[\alpha_L \mu_L(z) f_L(z) - \lambda_{HL} \overline{F}_H(z)\Big] dz, \quad \forall \theta^\star \in [0, 1].$$

Replacing the value of λ_{HL}, this equation can be cast over values $\theta_1^\star \leq \widehat{\theta}$ and $\theta_2^\star \geq \widehat{\theta}$ as

$$\frac{\int_{\theta_1^\star}^{\widehat{\theta}} \alpha_L \mu_L(z) f_L(z) dz}{\int_{\theta_1^\star}^{\widehat{\theta}} \overline{F}_H(z) dz} \leq \alpha_L r^{LH}(\widehat{\theta}) \leq \frac{\int_{\widehat{\theta}}^{\theta_2^\star} \alpha_L \mu_L(z) f_L(z) dz}{\int_{\widehat{\theta}}^{\theta_2^\star} \overline{F}_H(z) dz}, \quad \forall \theta_1^\star \leq \widehat{\theta} \leq \theta_2^\star \quad \text{(C.3)}$$

Condition (APR) ensures the equation above always hold. Indeed, condition (APR) implies that for any $\theta_1^\star \leq \widehat{\theta}$ and $\epsilon > 0$

$$\frac{\int_{\theta_1^\star}^{\widehat{\theta}} \alpha_L \mu_L(z) f_L(z) dz}{\int_{\theta_1^\star}^{\widehat{\theta}} \overline{F}_H(z) dz} \leq \frac{\int_{\widehat{\theta}}^{\widehat{\theta}+\epsilon} \alpha_L \mu_L(z) f_L(z) dz}{\int_{\widehat{\theta}}^{\widehat{\theta}+\epsilon} \overline{F}_H(z) dz}.$$

Taking $\epsilon \downarrow 0$ yields the left hand side inequality in (C.3). The right hand side inequality in (C.3) can be verified using an analogous argument. This shows (C.2), that is, the static contract maximizes the part of the Lagrangian that corresponds to interim type L. We now prove the same for type H. Note first that the optimality of the static contract implies

$$\lambda = \alpha_L r^{LH}(\widehat{\theta}) = -\alpha_H r^{HH}(\widehat{\theta}).$$

Then

$$\begin{aligned}
&\max_{x_H \in \Omega} \int_0^{\theta_{max}} x_H(z) \cdot \Big[\alpha_H \mu_H(z) f_H(z) + \lambda_{HL} \overline{F}_H(z)\Big] dz \\
&= \max_{x_H \in \Omega} \int_0^{\theta_{max}} x_H(z) \cdot \alpha_H \cdot \Big[\mu_H(z) f_H(z) - r^{HH}(\theta_s) \overline{F}_H(z)\Big] dz \\
&\overset{(a)}{=} \max_{x_H \in \Omega} \int_0^{\theta_{max}} x_H(z) \cdot \alpha_H \cdot \Big[r^{HH}(z) - r^{HH}(\theta_s)\Big] \overline{F}_H(z) dz \\
&\overset{(b)}{=} \int_{\widehat{\theta}}^{\theta_{max}} \alpha_H \cdot \Big[r^{HH}(z) - r^{HH}(\theta_s)\Big] \overline{F}_H(z) dz
\end{aligned}$$

where in (a) we have used the definition of $r^{HH}(\cdot)$ and in (b) our assumption that $r^{HH}(\cdot)$ is increasing. Thus, we have proved that for this choice of Lagrange multipliers the static contract maximizes the Lagrangian. Since the value of the Lagrangian coincides with the primal objective at the static solution, and this solution is always primal feasible. We conclude that the static contract is optimal.

Necessity. We differ this proof to the proof of Proposition 3.1. In it we show that whenever condition (APR) is not satisfied, there is a contract different from the static one with a strictly larger revenue. □

Proof of Proposition 3.1. Assume (APR) does not hold, then by Lemma C.1 (which we state and prove after the current proof) there exist $\theta_1 < \widehat{\theta} < \theta_2$ such that

$$\frac{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) r^{LH}(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) dz} > \frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) r^{LH}(z) dz}{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) dz}, \tag{C.4}$$

Consider a solution in which we set $u_L = u_H = 0$, and

$$x_L(\theta) = \begin{cases} 0 & \text{if } \theta < \theta_1 \\ x & \text{if } \theta_1 \leq \theta \leq \theta_2 \\ 1 & \text{if } \theta_2 < \theta, \end{cases} \quad x_H(\theta) = \begin{cases} 0 & \text{if } \theta < \widehat{\theta} \\ 1 & \text{if } \widehat{\theta} \leq \theta, \end{cases}$$

where $x = \int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) dz / \int_{\theta_1}^{\theta_2} \overline{F}_H(z) dz$. We next show that this solution is feasible and that yields an strict revenue improvement over the static contract.

Feasibility. The ex-post participation constraints are clearly satisfied. Also, since $\theta_1 < \widehat{\theta} < \theta_2$ we have $x_L \in (0,1)$, and both $x_L(\cdot)$ and $x_H(\cdot)$ are non-decreasing allocations. We verify the IC constraints

$$u_L + \int_0^{\theta_{max}} x_L(\theta) \overline{F}_L(\theta) d\theta \geq u_H + \int_0^{\theta_{max}} x_H(\theta) \overline{F}_L(\theta) d\theta,$$
$$u_H + \int_0^{\theta_{max}} x_H(\theta) \overline{F}_H(\theta) d\theta \geq u_L + \int_0^{\theta_{max}} x_L(\theta) \overline{F}_H(\theta) d\theta.$$

By replacing the allocations and ex-post utilities we obtain that the IC constraints are equivalent to

$$\frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z) dz} \geq \frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_L(z) dz}{\int_{\theta_1}^{\theta_2} \overline{F}_L(z) dz}. \tag{C.5}$$

To see why this is true, rewrite equation (C.4) as

$$\frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) dz} > \frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) r^{LH}(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) r^{LH}(z) dz}, \tag{C.6}$$

note that we are using here that by Lemma C.1 the denominator on the right hand side is strictly positive. Also, note that

$$\begin{aligned}
\frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) r^{LH}(z) dz}{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_L(z) dz} &= \frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_L(z) r^{LL}(z) dz}{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_L(z) dz} \\
&\geq r^{LL}(\widehat{\theta}) \frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_L(z) dz}{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_L(z) dz} \\
&= r^{LL}(\widehat{\theta}) \frac{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_L(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_L(z) dz} \\
&\geq \frac{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_L(z) r^{LL}(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_L(z) dz} \\
&= \frac{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) r^{LH}(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_L(z) dz},
\end{aligned}$$

where the inequalities come from the fact that $r^{LL}(\cdot)$ is an increasing function and $r^{LL}(\widehat{\theta}) \geq 0$. This gives

$$\frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) r^{LH}(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) r^{LH}(z) dz} \geq \frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_L(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_L(z) dz},$$

note that we are using here that by Lemma C.1 the denominator on the left hand side is strictly positive. This inequality together with (C.6) yields (C.5) and, therefore, the proposed solution is feasible.

Revenue improvement. We need to prove that

$$\begin{aligned}
\int_{\widehat{\theta}}^{\theta_{max}} [\alpha_L f_L(z) \mu_L(z) + \alpha_H f_H(z) \mu_H(z)] dz < \chi \cdot \int_{\theta_1}^{\theta_2} \alpha_L f_L(z) \mu_L(z) dz + \int_{\theta_2}^{\theta_{max}} \alpha_L f_L(z) \mu_L(z) dz \\
+ \int_{\widehat{\theta}}^{\theta_{max}} \alpha_H f_H(z) \mu_H(z) dz,
\end{aligned}$$

this is equivalent to

$$\int_{\widehat{\theta}}^{\theta_2} \alpha_L f_L(z) \mu_L(z) dz < \frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z) dz} \cdot \int_{\theta_1}^{\theta_2} \alpha_L f_L(z) \mu_L(z) dz$$

which is the same as

$$\frac{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) r^{LH}(z) dz}{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) dz} < \frac{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) r^{LH}(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) dz}$$

which is exactly the property satisfied by θ_1, θ_2 in (C.4). $\square$

Lemma C.1 *Suppose*

$$\max_{0\leq\theta\leq\widehat{\theta}} R^{LH}(\theta,\widehat{\theta}) > \min_{\widehat{\theta}\leq\theta\leq\theta_{max}} R^{LH}(\widehat{\theta},\theta).$$

Then, there exist $\theta_a, \theta_b \in [0, \theta_{max}]$ *with* $\theta_a < \widehat{\theta} < \theta_b$ *such that* $R^{LH}(\theta_a, \widehat{\theta}) > R^{LH}(\widehat{\theta}, \theta_b)$. *Moreover,* $0 < \int_{\theta_a}^{\widehat{\theta}} \overline{F}_H(z) r^{LH}(z)dz = \int_{\theta_a}^{\widehat{\theta}} \overline{F}_L(z) r^{LL}(z)dz$, *and* $0 < \int_{\widehat{\theta}}^{\theta_b} \overline{F}_H(z) r^{LH}(z)dz = \int_{\widehat{\theta}}^{\theta_b} \overline{F}_L(z) r^{LL}(z)dz$.

Proof of Lemma C.1. Note that both $r^{LH}(\cdot, \widehat{\theta})$ and $r^{LH}(\widehat{\theta}, \cdot)$ are continuous functions. Thus the maximum and the minimum in the statement are achieved by some $\tilde{\theta}_a \in [0, \widehat{\theta}]$ and $\tilde{\theta}_b \in [\widehat{\theta}, \overline{\theta}]$, respectively. Therefore, by assumption, we have that

$$R^{LH}(\tilde{\theta}_a, \widehat{\theta}) > R^{LH}(\widehat{\theta}, \tilde{\theta}_b).$$

Using the continuity of both function we can find $\theta_a < \widehat{\theta}$ and $\theta_b > \widehat{\theta}$ such that the inequality above is satisfied.

To finalize, we argue why $0 < \int_{\theta_a}^{\widehat{\theta}} \overline{F}_2(z) r^{LH}(z)dz$. Note that since $\theta_b > \widehat{\theta} \geq \widehat{\theta}_a$ (see Lemma 3.2) we have $R^{LH}(\widehat{\theta}, \theta_b) > 0$. Therefore, $R^{LH}(\theta_a, \widehat{\theta}) > 0$ which imply the desired inequalities. □

Proof of Lemma 3.3. From Lemma 3.2 we have that $\widehat{\theta}_L \leq \widehat{\theta} \leq \widehat{\theta_H}$. For exponential distributions, $\widehat{\theta}_L = 1/\lambda_L$ and $\widehat{\theta}_H = 1/\lambda_H$. Therefore, $\widehat{\theta} \in [1/\lambda_L, 1/\lambda_L]$. Moreover, $\widehat{\theta}$ must satisfy (3.8), if not we could increase it or decrease and obtain an strict revenue improvement.

We provide a proof for the rest of the properties for general distributions satisfying (DHR). Note first that $\widehat{\theta}$ can be seen as a function of α_L and α_H but since α_H equals $1 - \alpha_L$, we can effectively consider $\widehat{\theta}$ just a function of α_L. Then, when α_L equals 0 is as we only had type H buyers and, therefore, the optimal threshold is $\widehat{\theta}_H$. While when α_L equals 1 is as we only had type L buyers so the optimal threshold is $\widehat{\theta}_L$. Hence, $\widehat{\theta}(0)$ equals $\widehat{\theta}_H$ and $\widehat{\theta}(1)$ equals $\widehat{\theta}_L$.

Now we prove that $\widehat{\theta}(\alpha_L)$ is non-increasing. Consider $\alpha_L^a < \alpha_L^b$ and suppose that

$\widehat{\theta}(\alpha_L^a) < \widehat{\theta}(\alpha_L^b)$. Define

$$\ell(\theta, \alpha_L) \triangleq \int_{\theta}^{\theta_{max}} \alpha_L f_L(z)\mu_L(z) + (1-\alpha_L) f_H(z)\mu_H(z) dz,$$

note that this is a linear function of α_L and, for fixed α_L, it is maximized at $\widehat{\theta}(\alpha_L)$. Hence,

$$\begin{aligned} \ell(\widehat{\theta}(\alpha_L^a), \alpha_L^b) &\leq \ell(\widehat{\theta}(\alpha_L^b), \alpha_L^b) \\ &= \ell(\widehat{\theta}(\alpha_L^b), \alpha_L^b - \alpha_L^a) + \ell(\widehat{\theta}(\alpha_L^b), \alpha_L^a) \\ &\leq \ell(\widehat{\theta}(\alpha_L^b), \alpha_L^b - \alpha_L^a) + \ell(\widehat{\theta}(\alpha_L^a), \alpha_L^a) \end{aligned}$$

therefore

$$\int_{\widehat{\theta}(\alpha_L^a)}^{\widehat{\theta}(\alpha_L^b)} \alpha_L^b f_L(z)\mu_L(z) + (1-\alpha_L^b) f_H(z)\mu_H(z) dz \leq \int_{\widehat{\theta}(\alpha_L^a)}^{\widehat{\theta}(\alpha_L^b)} \alpha_L^a f_L(z)\mu_L(z) + (1-\alpha_L^a) f_H(z)\mu_H(z) dz. \tag{C.7}$$

Recall that $\widehat{\theta}$ is in $[\widehat{\theta}_L, \widehat{\theta}_H]$ and, therefore, $\widehat{\theta}_L \leq \widehat{\theta}(\alpha_L^a) < \widehat{\theta}(\alpha_L^b) \leq \widehat{\theta}_H$. This in turn implies that

$$\mu_L(z) > 0 \quad \text{and} \quad \mu_H(z) < 0, \quad \forall z \in (\widehat{\theta}(\alpha_L^a), \widehat{\theta}(\alpha_L^b)),$$

so for z in $(\widehat{\theta}(\alpha_L^a), \widehat{\theta}(\alpha_L^b))$ we have

$$\alpha_L^a f_L(z)\mu_L(z) + (1-\alpha_L^a) f_H(z)\mu_H(z) < \alpha_L^b f_L(z)\mu_L(z) + (1-\alpha_L^b) f_H(z)\mu_H(z),$$

which contradicts (C.7). □

Proof of Proposition 3.2. We make use of Theorem 3.1. Condition (APR) for the exponential distribution is

$$\max_{\theta \leq \widehat{\theta}} \left\{ \frac{\widehat{\theta} e^{-\lambda_L \widehat{\theta}} - \theta e^{-\lambda_L \theta}}{e^{-\lambda_H \widehat{\theta}} - e^{-\lambda_H \theta}} \right\} \leq \min_{\widehat{\theta} \leq \theta} \left\{ \frac{\theta e^{-\lambda_L \theta} - \widehat{\theta} e^{-\lambda_L \widehat{\theta}}}{e^{-\lambda_H \theta} - e^{-\lambda_H \widehat{\theta}}} \right\}. \tag{C.8}$$

Before we begin the proof we need some definitions and observations. Define the following functions

$$\underline{g}(\theta) \triangleq \frac{\widehat{\theta} e^{-\lambda_L \widehat{\theta}} - \theta e^{-\lambda_L \theta}}{e^{-\lambda_H \widehat{\theta}} - e^{-\lambda_H \theta}} \quad \text{and} \quad \overline{g}(\theta) \triangleq \frac{\theta e^{-\lambda_L \theta} - \widehat{\theta} e^{-\lambda_L \widehat{\theta}}}{e^{-\lambda_H \theta} - e^{-\lambda_H \widehat{\theta}}}.$$

Note the following

$$\lim_{\theta \to \widehat{\theta}+} \overline{g}(\theta) = \lim_{\theta \to \widehat{\theta}-} \underline{g}(\theta) = \frac{(\lambda_L \widehat{\theta} - 1)}{\lambda_H} \cdot e^{-\widehat{\theta}(\lambda_L - \lambda_H)}, \tag{C.9}$$

and

$$\lim_{\theta \to \infty} \overline{g}(\theta) = \widehat{\theta} \cdot e^{-\widehat{\theta}(\lambda_L - \lambda_H)}. \tag{C.10}$$

Finally note that

$$\frac{(\lambda_L \widehat{\theta} - 1)}{\lambda_H} \cdot e^{-\widehat{\theta}(\lambda_L - \lambda_H)} \leq \widehat{\theta} \cdot e^{-\widehat{\theta}(\lambda_L - \lambda_H)} \iff \widehat{\theta} \leq \frac{1}{\lambda_L - \lambda_H}. \tag{C.11}$$

Now, suppose condition (APR) holds and

$$\widehat{\theta} > \frac{1}{\lambda_L - \lambda_H} \tag{C.12}$$

From equations (C.9),(C.10) and (C.11) we see that

$$\overline{g}(\widehat{\theta}) = \underline{g}(\widehat{\theta}) > \lim_{\theta \to \infty} \underline{g}(\theta),$$

which implies

$$\max_{\theta \leq \widehat{\theta}} \left\{ \frac{\widehat{\theta} e^{-\lambda_L \widehat{\theta}} - \theta e^{-\lambda_L \theta}}{e^{-\lambda_H \widehat{\theta}} - e^{-\lambda_H \theta}} \right\} > \min_{\widehat{\theta} \leq \theta} \left\{ \frac{\theta e^{-\lambda_L \theta} - \widehat{\theta} e^{-\lambda_L \widehat{\theta}}}{e^{-\lambda_H \theta} - e^{-\lambda_H \widehat{\theta}}} \right\} \tag{C.13}$$

contradicting the fact that condition (APR) holds.

For the other direction, assume equation (3.9) holds. We first prove that for $\theta \leq \widehat{\theta}$ we have $\underline{g}(\theta) \leq \underline{g}(\widehat{\theta})$, indeed

$$\begin{aligned}
\underline{g}(\theta) \leq \underline{g}(\widehat{\theta}) &\iff \frac{\widehat{\theta} e^{-\lambda_L \widehat{\theta}} - \theta e^{-\lambda_L \theta}}{e^{-\lambda_H \widehat{\theta}} - e^{-\lambda_H \theta}} \leq \frac{(\lambda_L \widehat{\theta} - 1)}{\lambda_H} \cdot e^{-\widehat{\theta}(\lambda_L - \lambda_H)} \\
&\iff \lambda_H \cdot (\widehat{\theta} e^{-\lambda_L \widehat{\theta}} - \theta e^{-\lambda_L \theta}) \geq (e^{-\lambda_H \widehat{\theta}} - e^{-\lambda_H \theta}) \cdot (\lambda_L \widehat{\theta} - 1) \cdot e^{-\widehat{\theta}(\lambda_L - \lambda_H)} \\
&\iff \lambda_H \widehat{\theta} \cdot (1 - \frac{\theta}{\widehat{\theta}} e^{-\lambda_L (\theta - \widehat{\theta})}) - (1 - e^{-\lambda_H (\theta - \widehat{\theta})}) \cdot (\lambda_L \widehat{\theta} - 1) \geq 0,
\end{aligned}$$

so we just need to see that this last inequality holds for $\theta \leq \widehat{\theta}$. For doing so define

$$H(\theta) \triangleq \lambda_H \widehat{\theta} \cdot (1 - \frac{\theta}{\widehat{\theta}} e^{-\lambda_L (\theta - \widehat{\theta})}) - (1 - e^{-\lambda_H (\theta - \widehat{\theta})}) \cdot (\lambda_L \widehat{\theta} - 1),$$

and note that $H(\widehat{\theta}) = 0$ and

$$H(0) = \lambda_H\widehat{\theta} + (e^{\lambda_H\widehat{\theta}} - 1) \cdot (\lambda_L\widehat{\theta} - 1) \geq \lambda_H\widehat{\theta} + \lambda_H\widehat{\theta}(\lambda_L\widehat{\theta} - 1) = \lambda_H\widehat{\theta} \cdot \lambda_L\widehat{\theta} > 0,$$

where the inequality comes from convexity of the exponential function and the fact that $\widehat{\theta} \geq 1/\lambda_L$. Furthermore the derivative of H is given by

$$\frac{dH}{d\theta} = \lambda_H(\lambda_L\theta - 1)e^{-\lambda_L(\theta-\widehat{\theta})} - \lambda_H(\lambda_L\widehat{\theta} - 1)e^{-\lambda_H(\theta-\widehat{\theta})},$$

and it can be easily verified that for $\theta \leq \widehat{\theta}$ we have $dH/d\theta \leq 0$. This together to the facts that $H(0) > 0$ and $H(\widehat{\theta}) = 0$ imply that $\underline{g}(\theta) \leq \underline{g}(\widehat{\theta})$ for all $\theta \leq \widehat{\theta}$. Which in turn implies

$$\max_{\theta \leq \widehat{\theta}} \left\{ \frac{\widehat{\theta}e^{-\lambda_L\widehat{\theta}} - \theta e^{-\lambda_L\theta}}{e^{-\lambda_H\widehat{\theta}} - e^{-\lambda_H\theta}} \right\} = \frac{(\lambda_L\widehat{\theta} - 1)}{\lambda_H} \cdot e^{-\widehat{\theta}(\lambda_L - \lambda_H)}.$$

Now we prove that for $\theta \geq \widehat{\theta}$ we have $\overline{g}(\theta) \geq \overline{g}(\widehat{\theta})$. Note that if we prove this we are done because this and what we have just proven imply condition (APR). As before we do

$$\begin{aligned}
\overline{g}(\theta) \geq \overline{g}(\widehat{\theta}) &\iff \frac{\theta e^{-\lambda_L\theta} - \widehat{\theta}e^{-\lambda_L\widehat{\theta}}}{e^{-\lambda_H\theta} - e^{-\lambda_H\widehat{\theta}}} \geq \frac{(\lambda_L\widehat{\theta} - 1)}{\lambda_H} \cdot e^{-\widehat{\theta}(\lambda_L - \lambda_H)} \\
&\iff \lambda_H(\widehat{\theta}e^{-\lambda_L\widehat{\theta}} - \theta e^{-\lambda_L\theta}) \geq (\lambda_L\widehat{\theta} - 1) \cdot (e^{-\lambda_H\widehat{\theta}} - e^{-\lambda_H\theta}) \cdot e^{-\widehat{\theta}(\lambda_L - \lambda_H)} \\
&\iff \lambda_H(\widehat{\theta} - \theta e^{-\lambda_L(\theta-\widehat{\theta})}) - (\lambda_L\widehat{\theta} - 1) \cdot (1 - e^{-\lambda_H(\theta-\widehat{\theta})}) \geq 0,
\end{aligned}$$

note that the LHS of this last inequality is again the function $H(\cdot)$ but this time defined for $\theta \geq \widehat{\theta}$. We have $H(\widehat{\theta}) = 0$. It is easy to prove that for $\widehat{\theta} \leq \theta \leq \tilde{\theta}$ the function $H(\theta)$ is increasing, and then for $\theta > \tilde{\theta}$ is decreasing, where $\tilde{\theta} > \widehat{\theta}$ and $dH(\tilde{\theta})/d\theta = 0$. Also,

$$\lim_{\theta \to \infty} H(\theta) = \lambda_H\widehat{\theta} - (\lambda_L\widehat{\theta} - 1) \geq 0,$$

hence for $\theta \geq \widehat{\theta}$ we have $H(\theta) \geq 0$ and, therefore, $\overline{g}(\theta) \geq \overline{g}(\widehat{\theta})$ for all $\theta \geq \widehat{\theta}$, as desired.

□

Proof of Corollary 3.1. Recall that for any $\lambda_L > \lambda_H$ from Lemma 3.3 we have

$$\frac{1}{\lambda_L} \leq \widehat{\theta}(\alpha_L) \leq \frac{1}{\lambda_H},$$

and

$$\lambda_L \leq 2\lambda_H \Longleftrightarrow \frac{1}{\lambda_H} \leq \frac{1}{\lambda_L - \lambda_H},$$

therefore, for any $\alpha_L \in [0,1]$ equation (3.9) is satisfied. Then by Proposition 3.2 we conclude that the static contract is optimal for any $\alpha_L \in [0,1]$. $\square$

Proof of Corollary 3.2. First we show $\widehat{\theta}(\cdot)$ is continuous from the right at zero. Let $\{\alpha_L^n\} \in [0,1]$ be any sequence such that

$$\lim_{n\to\infty} \alpha_L^n = 0,$$

and suppose $\widehat{\theta}(\alpha_L^n)$ does not converge to $\widehat{\theta}(0) = 1/\lambda_H$. That is,

$$\exists \epsilon > 0, \forall n_0, \exists n \geq n_0, \quad |\frac{1}{\lambda_H} - \widehat{\theta}(\alpha_L^n)| > \epsilon,$$

since $\widehat{\theta}(\alpha_L^n) \leq \frac{1}{\lambda_H}$ we have

$$|\frac{1}{\lambda_H} - \widehat{\theta}(\alpha_L^n)| > \epsilon \Longleftrightarrow \frac{1}{\lambda_H} - \widehat{\theta}(\alpha_L^n) > \epsilon.$$

This in turn means that we can create a subsequence $\{\alpha_L^{\ell_n}\} \subset \{\alpha_L^n\}$ such that

$$\forall n, \quad \frac{1}{\lambda_H} - \epsilon > \widehat{\theta}(\alpha_L^{\ell_n}). \tag{C.14}$$

But since $\widehat{\theta}(\alpha_L^{\ell_n})$ is a maximizer of $\Pi^{\text{static}}(\cdot)$ we must have

$$\alpha_L^{\ell_n}\widehat{\theta}(\alpha_L^{\ell_n})e^{-\lambda_L\widehat{\theta}(\alpha_L^{\ell_n})}+(1-\alpha_L^{\ell_n})\widehat{\theta}(\alpha_L^{\ell_n})e^{-\lambda_H\widehat{\theta}(\alpha_L^{\ell_n})} \geq \alpha_L^{\ell_n}\frac{1}{\lambda_H}e^{-\lambda_L\frac{1}{\lambda_H}}+(1-\alpha_L^{\ell_n})\frac{1}{\lambda_H}e^{-\lambda_H\frac{1}{\lambda_H}},$$

because $\lambda_L > \lambda_H$ we can bound the LHS above to obtain

$$\widehat{\theta}(\alpha_L^{\ell_n})e^{-\lambda_H\widehat{\theta}(\alpha_L^{\ell_n})} \geq \alpha_L^{\ell_n}\frac{1}{\lambda_H}e^{-\lambda_L\frac{1}{\lambda_H}} + (1-\alpha_L^{\ell_n})\frac{1}{\lambda_H}e^{-\lambda_H\frac{1}{\lambda_H}}. \tag{C.15}$$

Note that the function $\theta e^{-\lambda_H\theta}$ has a unique maximum at $\theta = 1/\lambda_H$ and since $\widehat{\theta}(\alpha_L^{\ell_n})$ satisfies equation (C.14), we can always find $\delta(\epsilon) > 0$ such that

$$\Big(\frac{1}{\lambda_H} + \delta(\epsilon)\Big)e^{-\lambda_H(\frac{1}{\lambda_H}+\delta(\epsilon))} > \widehat{\theta}(\alpha_L^{\ell_n})e^{-\lambda_H\widehat{\theta}(\alpha_L^{\ell_n})}, \quad \forall n,$$

plugging this in equation (C.15) yields

$$\Big(\frac{1}{\lambda_H}+\delta(\epsilon)\Big)e^{-\lambda_H(\frac{1}{\lambda_H}+\delta(\epsilon))} > \alpha_L^{\ell_n}\frac{1}{\lambda_H}e^{-\lambda_L\frac{1}{\lambda_H}} + (1-\alpha_L^{\ell_n})\frac{1}{\lambda_H}e^{-\lambda_H\frac{1}{\lambda_H}}, \quad \forall n,$$

so taking the limit over n gives a contradiction. In conclusion we have proved that $\widehat{\theta}(\cdot)$ is continuous from the right at zero. Now, to finalize the proof recall that we are assuming $\lambda_L > 2\lambda_H$ or equivalently $\frac{1}{\lambda_H} > \frac{1}{\lambda_L-\lambda_H}$. However, since $\widehat{\theta}(0) = 1/\lambda_H$ and $\widehat{\theta}(\cdot)$ is continuous from the right we can always find $\bar{\alpha}_L \in (0,1]$ such that

$$\frac{1}{\lambda_H} \geq \widehat{\theta}(\bar{\alpha}_L) \geq \frac{1}{\lambda_L-\lambda_H},$$

so thanks to Proposition 3.2, the sequential contract is optimal when we set $\alpha_L > \bar{\alpha}_L$. Note that the same arguments is valid for $1/\lambda_L$. That is, we can show that $\widehat{\theta}(\alpha_L)$ is continuos from the left at 1 and then using the fact that

$$\frac{1}{\lambda_L-\lambda_H} > \frac{1}{\lambda_L},$$

we can find $\bar{\alpha}_H \in [\bar{\alpha}_L, 1)$ such that

$$\frac{1}{\lambda_L-\lambda_H} > \widehat{\theta}(\bar{\alpha}_H) \geq \frac{1}{\lambda_L},$$

hence in $[\bar{\alpha}_H, 1]$ the static contract is optimal. All of this implies that since $\widehat{\theta}(\cdot)$ is a non-increasing function we can always find $\bar{\alpha} \in (0,1)$ with the desired property. □

Proof of Corollary 3.3. Fix λ_H and α_L. Suppose the result is not true, that is,

$$\forall \bar{\lambda}_L \geq 2\lambda_H, \exists \lambda_L \geq \bar{\lambda}_L, \quad \widehat{\theta}(\lambda_L) \leq \frac{1}{\lambda_L-\lambda_H}.$$

From this we can construct a sequence $\lambda_L^n \geq 2\lambda_H$ such that

$$\lim_{n\to\infty}\lambda_L^n = \infty \quad \text{and} \quad \widehat{\theta}(\lambda_L^n) \leq \frac{1}{\lambda_L^n-\lambda_H}, \quad \forall n \in \mathbb{N},$$

therefore $\widehat{\theta}(\lambda_L^n)$ converges to 0, and we have

$$\Pi^{\text{static}}(\widehat{\theta}(\lambda_L^n)) = \widehat{\theta}(\lambda_L^n)e^{-\lambda_H\widehat{\theta}(\lambda_L^n)}\Big(\alpha_L e^{-(\lambda_L^n-\lambda_H)\widehat{\theta}(\lambda_L^n)} + \alpha_H\Big) \leq \widehat{\theta}(\lambda_L^n)e^{-\lambda_H\widehat{\theta}(\lambda_L^n)} \overset{n\to\infty}{\to} 0.$$

However, since $\widehat{\theta}(\lambda_L^n)$ maximizes $\Pi^{\text{static}}(\cdot)$ it must be the case that $\Pi^{\text{static}}(1/\lambda_H) \leq \Pi^{\text{static}}(\widehat{\theta}(\lambda_L^n))$, that is,

$$\alpha_L \frac{1}{\lambda_H} e^{-\lambda_L^n \frac{1}{\lambda_H}} + \alpha_H \frac{1}{\lambda_H} e^{-\lambda_H \frac{1}{\lambda_H}} \leq \Pi^{\text{static}}(\widehat{\theta}(\lambda_L^n)).$$

Taking limit over n at both sides of the previous equation yields

$$\alpha_H \frac{1}{\lambda_H} e^{-\lambda_H \frac{1}{\lambda_H}} \leq 0,$$

a contradiction. □

C.2 Proofs for Section 3.6

Proof of Theorem 3.2. For ease of exposition we restate the problem's formulation,

$$\begin{aligned}
(\mathcal{P}_R) \quad \max_{0 \leq \mathbf{x} \leq 1} \quad & - \sum_{k \in \{L,H\}} \alpha_k u_k + \sum_{\in \{L,H\}} \alpha_k \int_0^{\theta_{max}} x_k(z) \mu_k(z) f_k(\theta) d\theta \\
\text{s.t} \quad & x_k(\theta) \quad \text{non-decreasing}, \quad \forall k \in \{L,H\} \\
& u_k \geq 0, \forall k \in \{L,H\} \\
& u_H + \int_0^{\theta_{max}} x_H(z) \overline{F}_H(z) dz \geq u_L + \int_0^{\theta_{max}} x_L(z) \overline{F}_H(z) dz.
\end{aligned}$$

We separate this proof into two parts. In part 1 we show that the optimal solution has the structure in the statement of the theorem. Note that it is enough to provide a proof for the structure of the allocation, the transfers can be readily derived from Lemma 3.1. In part 2 we derive the properties about the thresholds, x_L and u_H and u_L.

Part 1. For any optimal solution to $(\mathcal{P}_R)$ two possible situations may arise:

1. The allocation has at least one interval in which is continuously strictly increasing.

2. The allocation does not have an interval in which is continuously strictly increasing, but is a piecewise constant non-decreasing function.

For each interim type, we prove that if we are in case (1), we can modify the allocation in that interval to be constant and obtain at least a weak improvement in the objective. This implies that for any optimal allocation, we can construct another optimal allocation that is a piecewise constant non-decreasing function. Therefore, we can always assume we are in case (2). In this case, we show that for interim type L there is only one intermediate step, and for interim type H there is no intermediate step.

We split the proof in interim type L and H. Let $x_L^\star(\theta)$ and $x_H^\star(\theta)$ denote the optimal allocations. We begin with interim type L.

- **interim type L case (1):** Suppose there is an interval (θ_1, θ_2) in which $x_L^\star(\theta)$ is continuously strictly increasing. Before we start with the main argument, note that if $\widehat{\theta}_L > \theta_1$ then we can set $x_L^\star(\theta)$ to be equal to $x_L^\star(\theta_1)$ for all θ in $(\theta_1, \widehat{\theta}_L)$. This strictly increases the objective function while maintaining feasibility. So we can assume $\widehat{\theta}_L \leq \theta_1$, which in turn implies that $\mu_L(\cdot)$ is non-negative in the interval (θ_1, θ_2).

  Now we give the main argument. Note that by Theorem 1 in [44, p. 217], $x_L^\star(\theta)$ must maximize the Lagrangian:[1]

$$\begin{aligned}\mathcal{L}(\boldsymbol{u}, \mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{w}) = {} & u_L(w_L - \lambda - \alpha_L) + u_H(\lambda - \alpha_H + w_H) \\ & + \int_0^{\theta_{max}} x_L(z) \cdot \Big[\alpha_L \mu_L(z) f_L(z) - \lambda \overline{F}_H(z)\Big] dz \\ & + \int_0^{\theta_{max}} x_H(z) \cdot \Big[\alpha_H \mu_H(z) f_H(z) + \lambda \overline{F}_H(z)\Big] dz,\end{aligned}$$

  with $\lambda, w_L, w_H \geq 0$. Define $L_L(\cdot)$ by

$$L_L(\theta) \triangleq \alpha_L \mu_L(\theta) f_L(\theta) - \lambda \overline{F}_H(\theta),$$

[1] To use this theorem we need to verify that there is a feasible solution that strictly satisfies all inequalities. We can take $u_L = u_H > 0$, $x_L(\theta) = \mathbf{1}_{\{\theta \geq \theta_L\}}$ and $x_H(\theta) = \mathbf{1}_{\{\theta \geq \theta_H\}}$ with $\theta_H < \theta_L$.

then it must be the case that $L_L(\theta) = 0$ for all $\theta \in (\theta_1, \theta_2)$. Suppose this is not true, then we could have $\widehat{\theta} \in (\theta_1, \theta_2)$ such that $L_L(\widehat{\theta}) > 0$, since $L_L(\cdot)$ is a continuous function this must also be true for all $\theta \in (\widehat{\theta} - \epsilon, \widehat{\theta} + \epsilon)$ for $\epsilon > 0$ small enough. But then we can obtain a strict improvement by setting $x_1(\theta) = x_L^\star(\widehat{\theta} + \epsilon)$ for all $\theta \in (\widehat{\theta} - \epsilon, \widehat{\theta} + \epsilon)$. A similar argument holds when $L_L(\widehat{\theta}) < 0$. Therefore, we have just proved that $L_L(\theta) = 0$ for all $\theta \in (\theta_1, \theta_2)$. In other words,

$$\alpha_L \frac{\mu_L(\theta) f_L(\theta)}{\overline{F}_H(\theta)} = \lambda \geq 0, \quad \forall \theta \in (\theta_1, \theta_2), \tag{C.16}$$

Also, by the second mean value theorem for integrals there exists $\widehat{\theta} \in (\theta_1, \theta_2)$ such that

$$x_L^\star(\widehat{\theta}) = \frac{\int_{\theta_1}^{\theta_2} x_L^\star(z) \overline{F}_2(z) dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z) dz}. \tag{C.17}$$

Going back to $(\mathcal{P}_R)$, we have that the part of objective associated to $x_L^\star$ in (θ_1, θ_2) is

$$\int_{\theta_1}^{\theta_2} \alpha_L x_L^\star(z) \mu_L(z) f_L(z) dz = \lambda \cdot \int_{\theta_1}^{\theta_2} x_L^\star(z) \overline{F}_H(z) dz, \tag{C.18}$$

where in the equality we have used (C.16). Now, consider modifying $x_L^\star$ to be $\tilde{x}_L^\star$ equal to $x_L^\star(\widehat{\theta})$ in (θ_1, θ_2). Then from (C.16), (C.17) and (C.18) we get

$$\begin{aligned}
\int_{\theta_1}^{\theta_2} x_L^\star(z) \alpha_L \mu_L(z) f_L(z) dz &= \lambda \cdot x_L^\star(\hat{\theta}) \cdot \int_{\theta_1}^{\theta_2} \overline{F}_H(z) dz \\
&= x_L^\star(\hat{\theta}) \cdot \int_{\theta_1}^{\theta_2} \alpha_L \mu_L(z) f_L(z) dz \\
&= \int_{\theta_1}^{\theta_2} \tilde{x}_L^\star(z) \alpha_L \mu_L(z) f_L(z) dz,
\end{aligned}$$

therefore, the modified $\tilde{x}_L^\star$ has the same objective value than the old one. Also, note that we have preserved feasibility because

$$
\begin{aligned}
u_L + \int_0^{\theta_{max}} \tilde{x}_L^\star(z)\overline{F}_H(z)dz &= u_L + \int_{\theta_1}^{\theta_2} \tilde{x}_L^\star(z)\overline{F}_H(z)dz + \int_{(\theta_1,\theta_2)^c} \tilde{x}_L^\star(z)\overline{F}_H(z)dz \\
&= u_L + x_L^\star(\hat{\theta}) \cdot \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{(\theta_1,\theta_2)^c} x_L^\star(z)\overline{F}_H(z)dz \\
&\overset{(a)}{=} u_L + \int_{\theta_1}^{\theta_2} x_L^\star(z)\overline{F}_H(z)dz + \int_{(\theta_1,\theta_2)^c} x_L^\star(z)\overline{F}_H(z)dz \\
&= u_L + \int_0^{\theta_{max}} x_L^\star(z)\overline{F}_H(z)dz,
\end{aligned}
$$

where in (a) we used equation (C.17).

- **interim type L case (2):** Suppose for $x_L^\star(\cdot)$ there exists $\theta_1 < \theta_2 < \theta_3$ and $0 < x_1 < x_2 < 1$ such that $x_L^\star(\theta) = x_1$ in (θ_1, θ_2) and $x_L^\star(\theta) = x_2$ in (θ_2, θ_3). Since type's L allocation is piecewise constant we must have $x_L^\star(\theta_1^-) < x_1$ and $x_2 < x_L^\star(\theta_3^+)$.

  Then, the part of objective associated to interim type L in these intervals is

$$
\alpha_L \cdot x_1 \cdot \int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z)dz + \alpha_L \cdot x_2 \cdot \int_{\theta_2}^{\theta_3} \mu_L(z) f_L(z)dz. \tag{C.19}
$$

  If $\mu_L(\widehat{\theta}) \leq 0$ for some $\widehat{\theta} \in (\theta_1, \theta_3)$ then because of (DHR), $\mu_L(\theta) \leq 0$ for all $\theta \leq \widehat{\theta}$ and, therefore, we can always find a better solution by setting $x_L^\star(\theta) = 0$ for all $\theta \leq \widehat{\theta}$ (note that this does not affect feasibility in $(\mathcal{P}_R)$). So assume $\mu_L(\theta) > 0$ for all $\theta \in (\theta_1, \theta_3)$, then it must be the case that

$$
u_H + \int_0^{\theta_{max}} x_H(z)\overline{F}_H(z)dz = u_L + \int_0^{\theta_{max}} x_L(z)\overline{F}_H(z)dz, \tag{C.20}
$$

  otherwise we could increase x_1 and obtain an strict improvement in the objective. There are two cases:

  a) $\frac{\int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z)dz}{\int_{\theta_1}^{\theta_2} \overline{F}_2(z)dz} \geq \frac{\int_{\theta_2}^{\theta_3} \mu_L(z) f_L(z)dz}{\int_{\theta_2}^{\theta_3} \overline{F}_2(z)dz}$: In this case consider decreasing x_2 by $\epsilon_2 > 0$ and increasing x_1 by $\epsilon_1 > 0$, in such a way that equation (C.20) remains with equality, that is,

$$
\epsilon_1 \cdot \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz - \epsilon_2 \cdot \int_{\theta_2}^{\theta_3} \overline{F}_H(z)dz = 0. \tag{C.21}
$$

The change in equation (C.19) is

$$\alpha_L \cdot \frac{\epsilon_2 \cdot \int_{\theta_2}^{\theta_3} \overline{F}_H(z)dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz} \cdot \int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z)dz - \alpha_L \cdot \epsilon_2 \cdot \int_{\theta_2}^{\theta_3} \mu_L(z) f_L(z)dz, \quad \text{(C.22)}$$

which under our current assumption is non-negative. So we can weakly improve our objective, indeed we can do it so until $x_1 + \epsilon_1$ and $x_2 - \epsilon_2$ are equal,

$$x_1 + \epsilon_1 = x_2 - \epsilon_2 \Leftrightarrow x_1 + \epsilon_2 \cdot \frac{\int_{\theta_2}^{\theta_3} \overline{F}_H(z)dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz} = x_2 - \epsilon_2 \Leftrightarrow \epsilon_2 = \frac{(x_2 - x_1)}{1 + \frac{\int_{\theta_2}^{\theta_3} \overline{F}_H(z)dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz}},$$

since $x_2 > x_1$ we have $\epsilon_2 > 0$ and, therefore, we have shown that it is possible to increase x_1 and to decrease x_2 in such a way the objective is weakly improved and the solution is constant in (θ_1, θ_3).

b) $\frac{\int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z)dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz} < \frac{\int_{\theta_2}^{\theta_3} \mu_L(z) f_L(z)d\theta}{\int_{\theta_2}^{\theta_3} \overline{F}_H(z)dz}$: In this case consider increasing x_2 by $\epsilon_2 > 0$ and decreasing x_1 by $\epsilon_1 > 0$ in such a way that equation (C.20) remains with equality. By doing this the change in the objective is strictly positive, and we do it until either $x_1 = x_L^\star(\theta_1^-)$ or $x_2 = x_L^\star(\theta_3^+)$.

This proves the result for interim type L and case (2).

In conclusion, putting together what we have proved for type L in cases (1) and (2), we can always consider $x_L^\star$ to be a step function with at most one intermediate step as in the statement of the proposition.

Now we proceed with interim type 2.

- **interim type H case (1):** Suppose there is an arbitrary interval (θ_1, θ_2) in which $x_H^\star(\theta)$ is continuously strictly increasing. Before we start with the main argument, note that if $\widehat{\theta}_H < \theta_2$ then we can set $x_H^\star(\theta)$ to be equal to $x_H^\star(\theta_2)$ for all θ in $(\widehat{\theta}_H, \theta_2)$. This strictly increases the objective function and maintains feasibility. So we can assume $\widehat{\theta}_H \geq \theta_2$, which in turn implies that $\mu_H(\cdot)$ is non-positive in the interval (θ_1, θ_2).

Now we give the main argument. Note that by Theorem 1 in [44, p. 217], $x_H^\star(\theta)$ must maximize the Lagrangian

$$\begin{aligned}\mathcal{L}(\boldsymbol{u}, \mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{w}) &= u_L(w_L - \lambda - \alpha_L) + u_H(\lambda - \alpha_H + w_H) \\ &+ \int_0^{\theta_{max}} x_L(z) \cdot \Big[\alpha_L \mu_L(z) f_L(z) - \lambda \overline{F}_H(z)\Big] dz \\ &+ \int_0^{\theta_{max}} x_H(z) \cdot \Big[\alpha_H \mu_H(z) f_H(z) + \lambda \overline{F}_H(z)\Big] dz,\end{aligned}$$

with $\lambda, w_L, w_H \geq 0$. Define $L_H(\cdot)$ by

$$L_H(\theta) \triangleq \alpha_H \mu_H(\theta) f_H(\theta) + \lambda \overline{F}_H(\theta),$$

then it must be the case that $L_H(\theta) = 0$ for all $\theta \in (\theta_1, \theta_2)$. Suppose this is not true, then we could have $\widehat{\theta} \in (\theta_1, \theta_2)$ such that $L_H(\widehat{\theta}) > 0$, since $L_H(\cdot)$ is a continuous function this must also be true for all $\theta \in (\widehat{\theta} - \epsilon, \widehat{\theta} + \epsilon)$ for $\epsilon > 0$ small enough. But then we can obtain an strict improvement by setting $x_2(\theta) = x_H^\star(\widehat{\theta} + \epsilon)$ for all $\theta \in (\widehat{\theta} - \epsilon, \widehat{\theta} + \epsilon)$. A similar argument holds when $L_H(\widehat{\theta}) < 0$. Therefore, we have just proved that $L_H(\theta) = 0$ for all $\theta \in (\theta_1, \theta_2)$. In other words,

$$\alpha_H \frac{\mu_H(\theta) f_H(\theta)}{\overline{F}_H(\theta)} = -\lambda, \quad \forall \theta \in (\theta_1, \theta_2). \tag{C.23}$$

Also note that by the second mean value theorem for integrals, there exists $\widehat{\theta} \in (\theta_1, \theta_2)$ such that

$$x_H^\star(\widehat{\theta}) = \frac{\int_{\theta_1}^{\theta_2} x_H^\star(z) \overline{F}_H dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z) dz}. \tag{C.24}$$

Going back to $(\mathcal{P}_R)$, we have that the part of objective associated to $x_H^\star$ in (θ_1, θ_2) is

$$\int_{\theta_1}^{\theta_2} \alpha_H x_H^\star(z) \mu_H(z) f_H(z) dz = -\lambda \cdot \int_{\theta_1}^{\theta_2} x_H^\star(z) \overline{F}_H(z) dz, \tag{C.25}$$

where in the equality we have used (C.23). Now, consider modifying $x_H^\star$ to be $\tilde{x}_H^\star$ equal to $x_H^\star(\widehat{\theta})$ in (θ_1, θ_2). Then from (C.23), (C.24) and (C.25) we get

$$\begin{aligned}\int_{\theta_1}^{\theta_2} x_H^\star(z)\alpha_H\mu_H(z)f_H(z)dz &= -\lambda \cdot x_H^\star(\hat{\theta}) \cdot \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz \\ &= x_H^\star(\hat{\theta}) \cdot \int_{\theta_1}^{\theta_2} \alpha_H\mu_H(z)f_H(z)dz \\ &= \int_{\theta_1}^{\theta_2} \tilde{x}_H^\star(z)\alpha_H\mu_H(z)f_H(z)dz,\end{aligned}$$

therefore, the modified $\tilde{x}_H^\star$ has the same objective value than the old one. Also, note that we have preserved feasibility because

$$\begin{aligned}u_H + \int_0^{\theta_{max}} \tilde{x}_H^\star(z)\overline{F}_H(z)dz &= u_H + \int_{\theta_1}^{\theta_2} \tilde{x}_H^\star(z)\overline{F}_H(z)dz + \int_{(\theta_1,\theta_2)^c} \tilde{x}_H^\star(z)\overline{F}_H(z)dz \\ &= u_H + x_H^\star(\hat{\theta}) \cdot \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{(\theta_1,\theta_2)^c} x_H^\star(z)\overline{F}_H(z)dz \\ &\overset{(a)}{=} u_H + \int_{\theta_1}^{\theta_2} x_H^\star(z)\overline{F}_H(z)dz + \int_{(\theta_1,\theta_2)^c} x_H^\star(z)\overline{F}_H(z)dz \\ &= u_H + \int_0^{\theta_{max}} x_H^\star(z)\overline{F}_H(z)dz,\end{aligned}$$

where in (a) we used equation (C.24).

- **interim type H case (2):** Suppose $x_H^\star(\cdot)$ is an optimal solution to $(\mathcal{P}_R)$ for which there exists $\theta_1 < \theta_2$ and $0 < x < 1$ such that $x_H^\star(\theta) = x$ in (θ_1, θ_2). Similar to the proof of type L assume $x_H^\star(\theta_1^-) < x < x_H^\star(\theta_2^+)$.

  Then the part of the objective for the interim type 2 in this interval is

$$\alpha_H \cdot x \cdot \int_{\theta_1}^{\theta_2} \mu_H(z)f_H(z)dz. \tag{C.26}$$

  If $\mu_H(\widehat{\theta}) \geq 0$ for some $\widehat{\theta} \in (\theta_1, \theta_2)$ then because of (DHR), $\mu_H(\theta) \geq 0$ for all $\theta \geq \widehat{\theta}$ and, therefore, we can always find a better solution by setting $x_H^\star(\theta) = 1$ for all $\theta \geq \widehat{\theta}$ (note that this does not affect feasibility in $(\mathcal{P}_R)$). So assume $\mu_H(\theta) < 0$ for all $\theta \in (\theta_1, \theta_2)$, then it must be the case that

$$u_H + \int_0^{\theta_{max}} x_H(z)\overline{F}_H(z)dz = u_L + \int_0^{\theta_{max}} x_L(z)\overline{F}_H(z)dz, \tag{C.27}$$

otherwise we could decrease x and obtain an strict improvement in the objective. Now, consider splitting the interval in half, that is, take $\widehat{\theta} = (\theta_1 + \theta_2)/2$ and note that because of (DHR) we always have

$$\frac{\int_{\theta_1}^{\widehat{\theta}} \mu_H(z) f_H(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) dz} \leq \frac{\int_{\widehat{\theta}}^{\theta_2} \mu_H(z) f_H(z) dz}{\int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) dz}. \tag{C.28}$$

We can modify $x_H^\star(\theta)$ in (θ_1, θ_2) as follows and obtain an, at least weakly, objective improvement. For $\theta \in (\theta_1, \widehat{\theta})$ set $x_H^\star(\theta) = x - \epsilon_1$ and for $\theta \in (\widehat{\theta}, \theta_2)$ set $x_H^\star(\theta) = x + \epsilon_2$ with $\epsilon_1, \epsilon_2 > 0$, and such that equation (C.27) remains with equality. That is,

$$-\epsilon_1 \cdot \int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) dz + \epsilon_2 \cdot \int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) dz = 0.$$

With this modification the change in the objective is

$$-\alpha_H \cdot \frac{\epsilon_2 \cdot \int_{\widehat{\theta}}^{\theta_2} \overline{F}_H(z) dz}{\int_{\theta_1}^{\widehat{\theta}} \overline{F}_H(z) dz} \cdot \int_{\theta_1}^{\widehat{\theta}} \mu_H(z) f_H(z) dz + \alpha_H \cdot \epsilon_2 \cdot \int_{\widehat{\theta}}^{\theta_2} \mu_H(z) f_H(z) dz,$$

which by equation (C.28) is non-negative. Then we can keep increasing ϵ_2 until either $x - \epsilon_1 = x_H^\star(\theta_1^-)$ or $x + \epsilon_2 = x_H^\star(\theta_2^+)$. This proofs we can, at least weakly, improve the objective. It also proves that we can modify the solution in such a way that for one of the two halves of the intervals the step reaches the boundary bound given by either $x_H^\star(\theta_1^-)$ or $x_H^\star(\theta_2^+)$. For the half that did not reach the boundary, we can do the same procedure described above and then repeat this procedure until we completely get rid of the intermediate step between $(x_H^\star(\theta_1^-), x_H^\star(\theta_2^+))$. Note that this process can be potentially infinite, in which case a more rigorous argument is required.

Suppose the process described above goes for infinitely many steps. In this case, an allocation sequence $\{x_H^n(\theta)\}_{n \in \mathbb{N}}$ defined in $[\theta_1, \theta_2]$ is generated. To prove that the argument works, we need to show that there exists $\theta_\infty \in [\theta_1, \theta_2]$ such that

$$\lim_{n \to \infty} \int_{\theta_1}^{\theta_2} x_H^n(z) \mu_H(z) f_H(z) dz = x_H^\star(\theta_1) \int_{\theta_1}^{\theta_\infty} \mu_H(z) f_H(z) dz + x_H^\star(\theta_2) \int_{\theta_\infty}^{\theta_2} \mu_H(z) f_H(z) dz. \tag{C.29}$$

To prove this, let $\{\underline{\theta}_n, \overline{\theta}_n, \widehat{\theta}_n\}_{n\in\mathbb{N}}$ be the sequence generated in the infinite process where:

- $\underline{\theta}_n$ and $\overline{\theta}_n$ correspond to the lower and upper bound of the interval. For example, at the beginning $\underline{\theta}_1 = \theta_1$ and $\overline{\theta}_1 = \theta_2$. At the next iteration we will have either $\underline{\theta}_2 = \theta_1$ and $\overline{\theta}_2 = \widehat{\theta}$ or $\underline{\theta}_2 = \widehat{\theta}$ and $\overline{\theta}_2 = \theta_2$. Note that for all $n \in \mathbb{N}$: $\underline{\theta}_n, \overline{\theta}_n \in [\theta_1, \theta_2]$.
- $\widehat{\theta}_n$ is defined to be the half of the interval. So $\widehat{\theta}_1 = \widehat{\theta}$, and $\widehat{\theta}_2 = (\underline{\theta}_2 + \overline{\theta}_2)/2$.

From these definitions we have that $\underline{\theta}_n$ and $\overline{\theta}_n$ are bounded monotonic sequences (the first non-decreasing and the second non-increasing), thus both converge to a limit. Also,

$$\widehat{\theta}_n = \frac{\underline{\theta}_n + \overline{\theta}_n}{2},$$

then all three quantities, $\underline{\theta}_n, \overline{\theta}_n$ and $\widehat{\theta}_n$, converge to the same limit which we denote by $\theta_\infty \in [\theta_1, \theta_2]$ (if the limit was not the same we could continue iterating the process).From this we can conclude that the following limit holds almost surely

$$\lim_{n\to\infty} x_H^n(\theta) = \begin{cases} x_H^\star(\theta_1^-) & \text{if } \theta < \theta_\infty \\ x_H^\star(\theta_2^+) & \text{if } \theta \geq \theta_\infty, \end{cases} \qquad \forall \theta \in [\theta_1, \theta_2].$$

Finally, we can use the almost surely version of the dominated convergence theorem to obtain (C.29). This completes the proof for interim type 2 and case (2).

In conclusion, putting together what we have proved for type H in cases (1) and (2), we can always consider $x_H^\star$ to be a threshold allocation as in the statement of the proposition.

Part 2. From what we have just proved we can write down $(\mathcal{P}_R)$ as follows

$$\begin{aligned} \max \quad & -\sum_{k\in\{L,H\}} \alpha_k u_k + \alpha_1 \chi \int_{\theta_1}^{\theta_2} \mu_1(z) f_1(z) dz + \alpha_1 \int_{\theta_2}^{\theta_{max}} \mu_1(z) f_1(z) dz \\ & + \alpha_2 \int_{\theta_H}^{\theta_{max}} \mu_H(z) f_H(z) dz \\ \text{s.t} \quad & \chi \in [0,1], \quad \theta_1 \le \theta_2 \\ & u_k \ge 0, \;\; k \in \{L,H\} \\ & u_H + \int_{\theta_H}^{\theta_{max}} \overline{F}_H(z) dz \ge u_L + \chi \int_{\theta_1}^{\theta_2} \overline{F}_H(z) dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z) dz. \end{aligned}$$

- $u_L = 0$: From the formulation above it is clear that is always optimal to set $u_L = 0$.

- $\widehat{\theta}_L \le \theta_1$: Suppose the opposite, that is, $\widehat{\theta}_L > \theta_1$. This implies that between θ_1 and $\widehat{\theta}_1$, $\mu_L(\cdot)$ is negative. Then, we can increase θ_1 while keeping feasibility and, at the same time, increasing the objective function. Note this argument is also valid when $\theta_1 = \theta_2$. Also, note that we can obtain a strict improvement only when $x > 0$; however, when $x = 0$ we can only obtain a weak improvement. In either case, we can always consider $\widehat{\theta}_L \le \theta_1$.

- $\theta_H \le \widehat{\theta}_H$: Suppose the opposite, $\theta_H > \widehat{\theta}_H$. Since $\mu_H(\theta) > 0$ for all $\theta \ge \widehat{\theta}_H$, we can can decrease θ_H and obtain an objective improvement while maintaining feasibility.

- $u_H = 0$: Suppose $u_H > 0$, then we must have

$$u_H + \int_{\theta_H}^{\theta_{max}} \overline{F}_H(z) dz = x \int_{\theta_1}^{\theta_2} \overline{F}_H(z) dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z) dz, \tag{C.30}$$

otherwise, we could decrease u_H and, by doing so, improve the objective. Since $u_H > 0$, equation (C.30) yields

$$0 < u_H = x \int_{\theta_1}^{\theta_2} \overline{F}_H(z) dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z) dz - \int_{\theta_H}^{\theta_{max}} \overline{F}_H(z) dz, \tag{C.31}$$

then it must be true that $\theta_1 < \theta_H$; otherwise, from equation (C.31) we would have ($\theta_1 \leq \theta_2$)

$$\int_{\theta_H}^{\theta_1} \overline{F}_H(z)dz + \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z)dz < x \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z)dz,$$

which implies

$$\int_{\theta_H}^{\theta_1} \overline{F}_H(z)dz < 0,$$

a contradiction. Thus, $\theta_1 < \theta_H$.

Now consider, a new contract for type H which consists on decreasing the cut-off θ_H by $\epsilon > 0$ sufficiently small, but at the same time maintaining the equality in equation (C.30). Specifically, let $\theta_H(\epsilon) = \theta_H - \epsilon > 0$ (which we can do because as we just saw $\theta_H > \theta_1 \geq 0$) and let $u_H(\epsilon)$ be

$$u_H(\epsilon) = x \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z)dz - \int_{\theta_H(\epsilon)}^{\theta_{max}} \overline{F}_H(z)dz,$$

note that by taking ϵ small we still have $u_H(\epsilon) > 0$. We claim that this new contract, characterized by $\theta_1, \theta_2, x, \theta_H(\epsilon)$ and $u_H(\epsilon)$, yields a larger objective that the old contract, characterized by $\theta_1, \theta_2, x, \theta_H$ and u_H. The old contract objective's is

$$-\alpha_H u_H + \alpha_L x \int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z)dz + \alpha_L \int_{\theta_2}^{\theta_{max}} \mu_L(z) f_L(z)dz + \alpha_H \int_{\theta_H}^{\theta_{max}} \mu_H(z) f_H(z)dz,$$

and using equation (C.30) it becomes

$$x \int_{\theta_1}^{\theta_2} (\alpha_L \mu_L(z) f_L(z) - \alpha_H \overline{F}_H(z))dz + \int_{\theta_2}^{\theta_{max}} (\alpha_L \mu_L(z) f_L(z) - \alpha_H \overline{F}_H(z))dz$$
$$+\alpha_H \int_{\theta_H}^{\theta_{max}} z f_H(z)dz.$$

We obtain a similar expression for the new contract's objective. Specifically, the first two terms in the expression above are the same and the third term differs in θ_H. Hence, the new contract yields an improvement over the old one if and only if

$$\int_{\theta_H}^{\theta_{max}} z f_H(z)dz < \int_{\theta_H(\epsilon)}^{\theta_{max}} z f_H(z)dz.$$

Since $\theta_H(\epsilon) < \theta_H$ this last inequality is true. Thus, if $u_H > 0$ we can always construct a new contract yielding a larger objective value and, therefore, at any optimal contract we must have $u_H = 0$.

- $\theta_H \leq \theta_2$: Since at any optimal solution $u_H = 0$, the IC constraint is
$$\int_{\theta_H}^{\theta_{max}} \overline{F}_H(z)dz \geq x \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z)dz.$$
Hence, if $\theta_H > \theta_2$ from the expression above we would have
$$\int_{\theta_H}^{\theta_{max}} \overline{F}_H(z)dz \geq x \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{\theta_2}^{\theta_H} \overline{F}_H(z)dz + \int_{\theta_H}^{\theta_{max}} \overline{F}_H(z)dz,$$
which implies $\theta_H = \theta_2$, a contradiction.

- $\theta_1 \leq \theta_H$: First we show that $\theta_1 \leq \widehat{\theta}_H$. Suppose the opposite, that is, $\theta_1 > \widehat{\theta}_H$. Then, since $\widehat{\theta}_H \geq \theta_H$ we must have $\theta_1 > \theta_H$ and, therefore,
$$\begin{aligned}\int_{\theta_H}^{\theta_{max}} \overline{F}_H(z)dz &= \int_{\theta_H}^{\theta_1} \overline{F}_H(z)dz + \int_{\theta_1}^{\theta_{max}} \overline{F}_H(z)dz \\ &> \int_{\theta_1}^{\theta_{max}} \overline{F}_H(z)dz \\ &= \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z)dz \\ &\geq \chi \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z)dz.\end{aligned}$$
That is, the IC constraint is not binding. Therefore, since $\theta_1 > \widehat{\theta}_H \geq \widehat{\theta}_L$ we can slightly decrease θ_1 and, in this way, obtain an objective improvement whenever $x > 0$. When $x = 0$, because $\theta_2 \geq \theta_1$, we can decrease θ_2 and obtain an objective improvement as well. Hence, at any optimal solution we must have $\theta_1 \leq \widehat{\theta}_H$.

  In order to complete the proof, suppose $\theta_1 > \theta_H$ then, as before, we have
$$\int_{\theta_H}^{\theta_{max}} \overline{F}_H(z)dz > x \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z)dz.$$
Using that $\theta_1 \leq \widehat{\theta}_H$ implies $\theta_H < \widehat{\theta}_H$, we can slightly increase θ_H (maintaining feasibility) and thus obtain an objective improvement. In conclusion, at any optimal solution we must have $\theta_1 \leq \theta_H$.

- $x = \int_{\theta_H}^{\theta_2} \overline{F}_H(z)dz / \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz$: since $\widehat{\theta}_L \leq \theta$, the part of the objective that involves x is always non-negative and, therefore, it is optimal to make x as large as possible. The IC constraints gives an upper bound for x which is precisely $\int_{\theta_H}^{\theta_2} \overline{F}_H(z)dz / \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz$, thus the result.

□

Proof of Theorem 3.3. We divide the proof into two part. In part 1 we show that the solution to the relaxed problem and the original problem coincide. In part 2 we prove that the three conditions that we state after the theorem are sufficient to characterize the optimality of the static contract.

Part 1. It is enough to show that the solution of $(\mathcal{P}_R)$ is feasible in $(\mathcal{P})$. From Theorem 3.2 we know that we can formulate $(\mathcal{P}_R)$ as

$$
\begin{aligned}
(\mathcal{P}_R^d) \quad \max \quad & \alpha_L \chi \int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z)dz + \alpha_L \int_{\theta_2}^{\theta_{max}} \mu_L(z) f_L(z)dz \\
& + \alpha_H \int_{\theta_H}^{\theta_{max}} \mu_H(z) f_H(z)dz \\
\text{s.t} \quad & \chi = \frac{\int_{\theta_H}^{\theta_2} \overline{F}_H(z)dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz} \\
& \hat{\theta}_L \leq \theta_1 \leq \theta_H \leq \theta_2, \theta_H \leq \hat{\theta}_H \\
& \int_{\theta_H}^{\theta_{max}} \overline{F}_H(z)dz \geq \chi \int_{\theta_1}^{\theta_2} \overline{F}_H(z)dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_H(z)dz.
\end{aligned}
$$

Let $\theta_1, \theta_H, \theta_2$ and x be the optimal solution to $(\mathcal{P}_R)$. If this solution corresponds to the optimal static contract or yields the same objective than it, we are done because this contract is always feasible in $(\mathcal{P})$. If this solution is different from the optimal static contract and yields a strictly larger objective than it, it must be the case that

$$
\begin{aligned}
\int_{\theta_H}^{\theta_{max}} \bar{\mu}(z)dz < \alpha_L x & \int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z)dz \\
& + \alpha_L \int_{\theta_2}^{\theta_{max}} \mu_L(z) f_L(z)dz + \alpha_H \int_{\theta_H}^{\theta_{max}} \mu_H(z) f_H(z)dz. \qquad \text{(C.32)}
\end{aligned}
$$

This is true because the contract $(u_1, u_2, x_1, x_2) = (0, 0, \mathbf{1}_{\{\theta \geq \theta_H\}}, \mathbf{1}_{\{\theta \geq \theta_H\}})$ is a feasible static contract and, therefore, its associated revenue is bounded by that of the optimal static contract. From the formulation of $(\mathcal{P}_R)$ we know that $\widehat{\theta}_L \leq \theta_1 \leq \theta_H \leq \theta_2$, this and equation (C.32) deliver

$$0 \leq \int_{\theta_H}^{\theta_2} \mu_L(z) f_L(z) dz < x \int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z) dz.$$

Hence, $\theta_1 < \theta_2$, $\theta_H < \theta_2$ (otherwise $x = 0$) and

$$\frac{\int_{\theta_H}^{\theta_2} \mu_L(z) f_L(z) dz}{\int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z) dz} < x. \tag{C.33}$$

Also, since $x \leq 1$ we must have $\theta_1 < \theta_H$. Note that since $\widehat{\theta}_L \leq \theta_1 < \theta_2$ the denominator above is strictly positive.

Now we argue that the contract optimizing $(\mathcal{P}_R)$ characterized by $\theta_1, \theta_H, \theta_2$ and x is feasible for $(\mathcal{P})$. Since the high to low IC constraint is satisfied, we only need to verify the low to high IC constraint. That is, we need to verify the following inequality

$$x \int_{\theta_1}^{\theta_2} \overline{F}_L(z) dz + \int_{\theta_2}^{\theta_{max}} \overline{F}_L(z) dz \geq \int_{\theta_H}^{\theta_{max}} \overline{F}_L(z) dz, \tag{C.34}$$

or, equivalently, $x \geq \int_{\theta_H}^{\theta_2} \overline{F}_L(z) dz / \int_{\theta_1}^{\theta_2} \overline{F}_L(z) dz$. In order to see why (C.34) holds, observe that from Lemma C.2 (which we state and prove after the present proof) we have

$$\frac{\int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z) dz}{\int_{\theta_1}^{\theta_2} \overline{F}_L(z) dz} \leq \frac{\int_{\theta_H}^{\theta_2} \mu_L(z) f_L(z) dz}{\int_{\theta_H}^{\theta_2} \overline{F}_L(z) dz} \Leftrightarrow \frac{\int_{\theta_1}^{\theta_H} \mu_L(z) f_L(z) dz}{\int_{\theta_1}^{\theta_H} \overline{F}_L(z) dz} \leq \frac{\int_{\theta_H}^{\theta_2} \mu_L(z) f_L(z) dz}{\int_{\theta_H}^{\theta_2} \overline{F}_L(z) dz}. \tag{C.35}$$

The right hand side in (C.35) always holds thanks to (DHR), indeed,

$$\begin{aligned} \frac{\int_{\theta_1}^{\theta_H} \mu_L(z) f_L(z) dz}{\int_{\theta_1}^{\theta_H} \overline{F}_L(z) dz} = \frac{\int_{\theta_1}^{\theta_H} \overline{F}_L r^{LL}(z) dz}{\int_{\theta_1}^{\theta_H} \overline{F}_L(z) dz} \leq r^{LL}(\theta_H) \leq \frac{\int_{\theta_H}^{\theta_2} \overline{F}_L r^{LL}(z) dz}{\int_{\theta_H}^{\theta_2} \overline{F}_L(z) dz} \\ = \frac{\int_{\theta_H}^{\theta_2} \mu_L(z) f_L(z) dz}{\int_{\theta_H}^{\theta_2} \overline{F}_L(z) dz}. \end{aligned}$$

Thus the left hand side in (C.35) holds. Equivalently,

$$\frac{\int_{\theta_H}^{\theta_2} \overline{F}_L(z) dz}{\int_{\theta_1}^{\theta_2} \overline{F}_L(z) dz} \leq \frac{\int_{\theta_H}^{\theta_2} \mu_L(z) f_L(z) dz}{\int_{\theta_1}^{\theta_2} \mu_L(z) f_L(z) dz}$$

Using this, together with equation (C.33), delivers equation (C.34). This concludes the proof for Part 1.

Part 2. In this part we prove the following optimality conditions for the thresholds $\theta_1 \leq \theta_H \leq \theta_2$:

1. $R^{LH}(\theta_1, \theta_2) \leq \min_{\theta_2 \leq \theta} R^{LH}(\theta_2, \theta)$;

2. $\max_{\theta \leq \theta_2} R^{LH}(\theta, \theta_2) \leq R^{LH}(\theta_1, \theta_2)$;

3. $\alpha_L \cdot R^{LH}(\theta_1, \theta_2) + \alpha_H r^{HH}(\theta_H) = 0$.

It is enough to prove that under the conditions the optimal contract characterized by $(\theta_1, \theta_H, \theta_2)$ is optimal for $(\mathcal{P}_R)$. To prove this we use a Lagrangian relaxation (we do not relax the monotonicity constraints) and show that this relaxation is optimized by the contract characterized by $(\theta_1, \theta_H, \theta_2)$.

First, we establish some properties that can be derived from conditions (1) to (3). Condition (3) implies that $\theta_2 \geq \widehat{\theta}_L$; otherwise, $\theta_1, \theta_2 < \widehat{\theta}_L$ which would imply that $R^{LH}(\theta_1, \theta_2) < 0$. In turn, condition (3) would give $R^{HH}(\theta_H) > 0$ which would imply that $\widehat{\theta}_H < \theta_H$. Since $\theta_H \leq \theta_2$ we would have $\widehat{\theta}_H < \theta_H \leq \theta_2 < \widehat{\theta}_L$, that is, $\widehat{\theta}_H < \widehat{\theta}_L$ which is not possible. Moreover, condition (2) together with the fact that $\theta_2 \geq \widehat{\theta}_L$ imply that $\theta_1 \geq \widehat{\theta}_L$. This yields $R^{LH}(\theta_1, \theta_2) \geq 0$, and thus we can use condition (3) again to deduce that $\theta_H \leq \widehat{\theta}_H$. In summary, $\widehat{\theta}_L \leq \theta_1$ and $\theta_H \leq \widehat{\theta}_H$.

Now we provide the main argument. If $\theta_1 = \theta_2$, then we also have $\theta_1 = \theta_2 = \theta_H$. Condition (3) implies that the contract characterize by $(\theta_1, \theta_H, \theta_2)$ is the static contract. Conditions (1) and (2) together yield (APR) and, therefore, from Theorem 3.1 we deduce that the static contract is optimal. Next suppose that $\theta_1 < \theta_2$, and define

$$\Omega \triangleq \{x : [0,1] \to [0,1] : x(\cdot) \text{ is non-decreasing}\}.$$

We use $\mathbf{x}^\star$ to denote the solution characterize by $(\theta_1, \theta_H, \theta_2)$. The Lagrangian for $(\mathcal{P}_R)$ is

$$\begin{aligned}\mathcal{L}(\boldsymbol{u}, \mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{w}) = {} & u_L(w_L - \lambda - \alpha_L) + u_H(\lambda - \alpha_H + w_H) \\ & + \int_0^{\theta_{max}} x_L(z) \cdot \Big[\alpha_L \mu_L(z) f_L(z) - \lambda \overline{F}_H(z)(z)\Big] dz \\ & + \int_0^{\theta_{max}} x_H(z) \cdot \Big[\alpha_H \mu_H(z) f_H(z) + \lambda \overline{F}_H(z)\Big] dz,\end{aligned}$$

consider the following multipliers

$$\lambda = \alpha_L \cdot R^{LH}(\theta_1, \theta_2), \quad w_L = \lambda + \alpha_L, w_H = -\lambda + \alpha_H,$$

note that λ and w_L are non-negative, and for w_H we have

$$w_H \geq 0 \Leftrightarrow \alpha_H + \alpha_H r^{HH}(\theta_H) \geq 0 \Leftrightarrow r^{HH}(\theta_H) \geq -1$$

if and only if

$$[\theta_H - h^{HH}(\theta_H)] \geq -h^{HH}(\theta_H) \Leftrightarrow \theta_H \geq 0,$$

where in the first if and only if we used condition (3) in our hypothesis. Thus when we optimize the Lagrangian we obtain:

$$\begin{aligned}\max_{(\boldsymbol{u},\mathbf{x})\in\Omega} \mathcal{L}(\mathbf{x}, \boldsymbol{u}, \boldsymbol{\lambda}, \boldsymbol{w}) = {} & \max_{0\leq\theta\leq\theta_{max}} \int_\theta^{\theta_{max}} \Big[\alpha_1 \mu_1(z) f_1(z) - \lambda \overline{F}_2(z)\Big] dz \\ & + \max_{0\leq\theta\leq\theta_{max}} \int_\theta^{\theta_{max}} \Big[\alpha_H \mu_H(z) f_H(z) + \lambda \overline{F}_H(z)\Big] dz\ , \end{aligned} \tag{C.36}$$

where we can reduce attention to threshold strategies because $x_L(\cdot), x_H(\cdot)$ are non-decreasing (see, e.g., [52] or [57]). If we are able to show that $\mathcal{L}(\mathbf{x}, \boldsymbol{u}, \boldsymbol{\lambda}, \boldsymbol{w})$ evaluated at our candidate solution is an upper bound for the RHS above we are done. Let's begin with the second term, take any $0 \leq \theta \leq \theta_{max}$ then

$$\begin{aligned}\int_\theta^{\theta_{max}} \Big[\alpha_H \mu_H(z) f_H(z) + \lambda \overline{F}_H(z)\Big] dz &= \int_\theta^{\theta_{max}} \Big[\alpha_H \mu_H(z) f_H(z) - \alpha_H r^{HH}(\theta_H) \overline{F}_H(z)\Big] dz \\ &= \int_\theta^{\theta_{max}} \alpha_H \overline{F}_H(z) \Big[r^{HH}(z) - r^{HH}(\theta_H)\Big] dz \\ &\leq \int_{\theta_H}^{\theta_{max}} \alpha_H \overline{F}_H(z) \Big[r^{HH}(z) - r^{HH}(\theta_H)\Big] dz \\ &= \int_0^{\theta_{max}} x_H^\star(z) \Big[\alpha_H \mu_H(z) f_H(z) + \lambda \overline{F}_H(z)\Big] dz,\end{aligned}$$

where in the first equality we used condition (3) and the inequality comes from the fact that $r^{HH}(\cdot)$ is non-decreasing. Now we look into the first term in equation (C.36), consider first $\theta \geq \theta_2$

$$\begin{aligned}\int_{\theta}^{\theta_{max}} \Big[\alpha_L \mu_L(z) f_L(z) - \lambda \overline{F}_H(z)\Big] dz &= \int_{\theta_L^2}^{\theta_{max}} \Big[\alpha_L \mu_L(z) f_L(z) - \lambda \overline{F}_H(z)\Big] dz \\ &\quad - \int_{\theta_L^2}^{\theta} \Big[\alpha_L \mu_L(z) f_L(z) - \lambda \overline{F}_H(z)\Big] dz \\ &\leq \int_{\theta_L^2}^{\theta_{max}} \Big[\alpha_L \mu_L(z) f_L(z) - \lambda \overline{F}_H(z)\Big] dz,\end{aligned}$$

where we have used that

$$-\int_{\theta_2}^{\theta} \Big[\alpha_L \mu_L(z) f_L(z) - \lambda \overline{F}_2(z)\Big] dz \leq 0$$

if and only if

$$\alpha_L \cdot \frac{\int_{\theta_1}^{\theta_2} \overline{F}_H(z) r^{LH}(z) dz}{\int_{\theta_1}^{\theta_2} \overline{F}_H(z) dz} = \lambda \leq \alpha_L \cdot \frac{\int_{\theta_2}^{\theta} \overline{F}_2(z) r^{LH}(z) dz}{\int_{\theta_2}^{\theta} \overline{F}_H(z) dz},$$

which thanks to condition (1) in our hypothesis is true. A similar argument holds for $\theta \leq \theta_2$, but using condition (2). Since $\mathcal{L}(\mathbf{x}^\star, 0, \boldsymbol{\lambda}, \boldsymbol{w})$ equals

$$\begin{aligned}x \int_{\theta_1}^{\theta_2} \Big[\alpha_1 \mu_1(z) f_1(z) - \lambda \overline{F}_H(z)\Big] dz &+ \int_{\theta_2}^{\theta_{max}} \Big[\alpha_1 \mu_1(z) f_1(z) - \lambda \overline{F}_H(z)\Big] dz \\ &+ \int_{\theta_H}^{\theta_{max}} \Big[\alpha_H \mu_H(z) f_H(z) + \lambda \overline{F}_H(z)\Big] dz,\end{aligned}$$

which by the definition of λ simplifies to

$$\int_{\theta_2}^{\theta_{max}} \Big[\alpha_1 \mu_1(z) f_1(z) - \lambda \overline{F}_H(z)\Big] dz + \int_{\theta_H}^{\theta_{max}} \Big[\alpha_H \mu_H(z) f_H(z) + \lambda \overline{F}_H(z)\Big] dz,$$

we conclude that $\max_{(\boldsymbol{u},\mathbf{x})\in\Omega} \mathcal{L}(\boldsymbol{u}, \mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{w}) \leq \mathcal{L}(0, \mathbf{x}^\star, \boldsymbol{\lambda}, \boldsymbol{w})$, as required. □

Lemma C.2 *Let $\theta_i \in [0,1]$ for $i = 1, 2, 3$ be such that $\theta_1 < \theta_2 < \theta_3$. Also, consider functions $f, g : [\theta_1, \theta_3] \to I$, with $\int_{\theta_1}^{\theta_2} g(z)dz, \int_{\theta_2}^{\theta_3} g(z)dz > 0$. Then,*

$$\frac{\int_{\theta_1}^{\theta_3} f(z)dz}{\int_{\theta_1}^{\theta_3} g(z)dz} \leq \frac{\int_{\theta_2}^{\theta_3} f(z)dz}{\int_{\theta_2}^{\theta_3} g(z)dz} \quad \textit{if and only if} \quad \frac{\int_{\theta_1}^{\theta_2} f(z)dz}{\int_{\theta_1}^{\theta_2} g(z)dz} \leq \frac{\int_{\theta_2}^{\theta_3} f(z)dz}{\int_{\theta_2}^{\theta_3} g(z)dz}.$$

Proof of Lemma C.2.

$$\frac{\int_{\theta_1}^{\theta_3} f(z)dz}{\int_{\theta_1}^{\theta_3} g(z)dz} \le \frac{\int_{\theta_2}^{\theta_3} f(z)dz}{\int_{\theta_2}^{\theta_3} g(z)dz} \Leftrightarrow \Big(\int_{\theta_2}^{\theta_3} g(z)dz\Big)\Big(\int_{\theta_1}^{\theta_3} f(\theta)dz\Big) \le \Big(\int_{\theta_1}^{\theta_3} g(z)dz\Big)\Big(\int_{\theta_2}^{\theta_3} f(z)dz\Big)$$

$$\Leftrightarrow \Big(\int_{\theta_2}^{\theta_3} g(z)dz\Big)\Big(\int_{\theta_1}^{\theta_2} f(z)dz\Big) \le \Big(\int_{\theta_1}^{\theta_2} g(z)dz\Big)\Big(\int_{\theta_2}^{\theta_3} f(z)dz\Big)$$

$$\Leftrightarrow \frac{\int_{\theta_1}^{\theta_2} f(z)dz}{\int_{\theta_1}^{\theta_2} g(z)dz} \le \frac{\int_{\theta_2}^{\theta_3} f(z)dz}{\int_{\theta_2}^{\theta_3} g(z)dz}$$

□

Proof of Proposition 3.3. We use the sufficient condition in Theorem 3.3. First note that since the support of the exponential distribution is unbounded from above, we can take $\theta_2 = \infty$ which eliminates condition (1). Conditions (2) and (3) can be cast as

$$\theta_1 e^{-\theta_1(\lambda_L-\lambda_H)} \ge \theta e^{-\theta(\lambda_L-\lambda_H)} \quad \forall\theta \ge 0 \quad \text{and} \quad \alpha_L\cdot\lambda_H\theta_1 e^{-\theta_1(\lambda_L-\lambda_H)} = -\alpha_H\cdot(\lambda_H\theta_H-1), \tag{C.37}$$

By optimizing the first term in (C.37) we obtain

$$\theta_1 = \frac{1}{\lambda_L-\lambda_H},$$

and then solving for θ_H yields

$$\theta_H = \frac{1}{\lambda_H} - \frac{\alpha_L}{\alpha_H}\frac{e^{-1}}{\lambda_L-\lambda_H}.$$

What we need to check (and it is not obvious at a first glance) is that $\theta_1 \le \theta_H$. First, we show

$$\alpha_L(\theta_1 - \frac{1}{\lambda_L})\lambda_L e^{-\lambda_L\theta_1} + \alpha_H(\theta_1 - \frac{1}{\lambda_H})\lambda_H e^{-\lambda_H\theta_1} < 0. \tag{C.38}$$

To prove this inequality notice that since $\widehat{\theta}$ is the optimal static cutoff we have

$$\alpha_L\widehat{\theta}e^{-\lambda_L\widehat{\theta}} + \alpha_H\widehat{\theta}e^{-\lambda_H\widehat{\theta}} \ge \alpha_L\theta_1 e^{-\lambda_L\theta_1} + \alpha_H\theta_1 e^{-\lambda_H\theta_1}, \tag{C.39}$$

then

$$\begin{aligned}
&\alpha_L(\theta_L^1 - \frac{1}{\lambda_L})\lambda_L e^{-\lambda_L \theta_L^1} \\
&+\alpha_H(\theta_L^1 - \frac{1}{\lambda_H})\lambda_H e^{-\lambda_H \theta_L^1} = \alpha_L \theta_L^1(\lambda_L - \lambda_H)e^{-\lambda_L\theta_L^1} + \alpha_L\theta_L^1\lambda_H e^{-\lambda_L\theta_L^1} \\
&\qquad + \alpha_H\theta_L^1\lambda_H e^{-\lambda_H\theta_L^1} - \alpha_L e^{-\lambda_L\theta_L^1} - \alpha_H e^{-\lambda_H\theta_L^1} \\
&\qquad = \alpha_L e^{-\lambda_L\theta_L^1} + \lambda_H(\alpha_L\theta_L^1 e^{-\lambda_L\theta_L^1} + \alpha_H\theta_L^1 e^{-\lambda_H\theta_L^1}) \\
&\qquad - \alpha_L e^{-\lambda_L\theta_L^1} - \alpha_H e^{-\lambda_H\theta_L^1} \\
&\qquad \overset{(a)}{\leq} \lambda_H(\alpha_L\widehat{\theta}e^{-\lambda_L\widehat{\theta}} + \alpha_H\widehat{\theta}e^{-\lambda_H\widehat{\theta}}) - \alpha_H e^{-\lambda_H\theta_L^1} \\
&\qquad \overset{(b)}{<} \lambda_H(\alpha_L\widehat{\theta}e^{-\lambda_L\widehat{\theta}} + \alpha_H\widehat{\theta}e^{-\lambda_H\widehat{\theta}}) - \alpha_H e^{-\lambda_H\widehat{\theta}} \\
&\qquad = \lambda_H\alpha_L\widehat{\theta}e^{-\lambda_L\widehat{\theta}} + \lambda_H\alpha_H e^{-\lambda_H\widehat{\theta}}(\widehat{\theta} - \frac{1}{\lambda_H}) \\
&\qquad \overset{(c)}{=} \lambda_H\alpha_L\widehat{\theta}e^{-\lambda_L\widehat{\theta}} - \lambda_L\alpha_L e^{-\lambda_L\widehat{\theta}}(\widehat{\theta} - \frac{1}{\lambda_L}) \\
&\qquad = \alpha_L e^{-\lambda_L\widehat{\theta}}\Big(- \widehat{\theta}(\lambda_L - \lambda_H) + 1\Big) \\
&\qquad \overset{(d)}{<} 0,
\end{aligned}$$

where (a) comes from equation (C.39), (b) is true because the function $-e^{-\lambda_H\theta}$ increasing and $\theta_1 < \widehat{\theta}$, (c) comes from equation (3.8). And (d) comes from $\theta_1 < \widehat{\theta}$. With this we have proven (C.38) and thus

$$\begin{aligned}
\lambda_L\alpha_H \cdot (\theta_H - \frac{1}{\lambda_H}) &\overset{(a)}{=} -\lambda_L\alpha_L \cdot \theta_L^1 e^{-\theta_L^1(\lambda_L - \lambda_H)} \\
&= -\lambda_L\alpha_L \cdot \Big(\theta_L^1 - \frac{1}{\lambda_L}\Big)e^{-\theta_L^1(\lambda_L-\lambda_H)} - \lambda_L\alpha_L \cdot \frac{1}{\lambda_L}e^{-\theta_L^1(\lambda_L-\lambda_H)} \\
&\overset{(b)}{>} \alpha_H(\theta_L^1 - \frac{1}{\lambda_H})\lambda_H - \alpha_L \cdot e^{-\theta_L^1(\lambda_L-\lambda_H)} \\
&\overset{(c)}{=} \alpha_H(\theta_L^1 - \frac{1}{\lambda_H})\lambda_H + \frac{\alpha_H}{\theta_L^1} \cdot (\theta_H - \frac{1}{\lambda_H}),
\end{aligned}$$

where in (a) and (c) we used the definition of θ_H, and in (b) we used equation (C.38). From this we have that

$$(\theta_H - \frac{1}{\lambda_H}) \cdot \Big(\lambda_L\alpha_H - \frac{\alpha_H}{\theta_1}\Big) > \alpha_H(\theta_1 - \frac{1}{\lambda_H})\lambda_H,$$

but replacing θ_1 with $1/(\lambda_L - \lambda_H)$ in this last expression we get $\theta_H > \theta_1$.

Finally, x is given by

$$x = \frac{\int_{\theta_H}^{\theta_3} \overline{F}_H(z)dz}{\int_{\theta_1}^{\theta_3} \overline{F}_H(z)dz} = \frac{e^{-\lambda_H\theta_H}}{e^{-\lambda_H\theta_1}} = \exp\Big(- \lambda_H\Big[\frac{1}{\lambda_H} - \frac{\alpha_L}{\alpha_H}\frac{e^{-1}}{\lambda_L - \lambda_H} - \frac{1}{\lambda_L - \lambda_H}\Big]\Big).$$

□

C.3 Proofs for Section 3.7

Proof of Theorem 3.4. In Lemma C.3 (which we state and prove after this proof) we show that $\mathcal{A}$ is non-empty. Next we prove the necessary and sufficient condition.

We prove both directions separately. First we show that if there exists $\boldsymbol{\lambda} \in \mathcal{A}$ satisfying the properties then the static contract is optimal. Then we show that if the static contract is optimal then we can always solve for $\boldsymbol{\lambda}$ satisfying the properties.

Define

$$\Omega \triangleq \{x : [0,1] \longrightarrow [0,1] : x(\cdot) \text{ is non-decreasing}\}, \quad \text{and} \quad \Omega^K \triangleq \underbrace{\Omega \times \cdots \times \Omega}_{K \text{ times}}.$$

For the first part we use a Lagrangian relaxation approach. That is, we dualize the IC constraints for a specific set of multipliers. This gives an upper bound to the seller's problem. Then we show that for our choice of multipliers the relaxation is maximized at the static allocation. The Lagrangian is

$$\begin{aligned}\mathcal{L}(x, u, \boldsymbol{\lambda}, \mathbf{w}) = &\sum_{k=1}^{K} u_k\Big(- \alpha_k + w_k + \sum_{j:j\neq k} \lambda_{kj} - \sum_{j:j\neq k} \lambda_{jk}\Big)\\ &+ \sum_{k=1}^{K} \int_0^{\theta_{max}} x_k(z)\Big(\alpha_k\mu_k(z)f_k(z) + \overline{F}_k(z) \cdot \sum_{j:j\neq k} \lambda_{kj} - \sum_{j:j\neq k} \lambda_{jk}\overline{F}_j(z)\Big)dz,\end{aligned}$$

where $\boldsymbol{\lambda}$ correspond to the multipliers associated with the ICs, and $\mathbf{w}$ to the multipliers associated with the ex-post IR constraints. Let us define $\boldsymbol{\lambda}$ to be equal to the

$(\lambda_{ij})_{i,j\in\{1,\cdots,K\}^2}$ we are assuming to exist, that is $\boldsymbol{\lambda} \in \mathcal{A}$, and let

$$w_k = \alpha_k + \sum_{j:j\neq k} \lambda_{jk} - \sum_{j:j\neq k} \lambda_{kj}, \forall k \in \{1, \ldots, K\}. \tag{C.40}$$

Note that by our choice of $\boldsymbol{\lambda}$ ($\boldsymbol{\lambda} \in \mathcal{A}$), w_k is non-negative for all k. With this choice of $\mathbf{w}$ the first summation in the Lagrangian becomes zero. Now, we need to show that for this choice of multipliers the Lagrangian is maximized at the static contract. In order to show this observe that

$$\max_{x\in\Omega^K, u\geq 0} \mathcal{L}(x, u, \boldsymbol{\lambda}, \mathbf{w}) = \sum_{k=1}^{K} \max_{x_k\in\Omega} \int_0^{\theta_{max}} x_k(z)\Big(\alpha_k\mu_k(z)f_k(z) + \overline{F}_k(z) \cdot \sum_{j:j\neq k} \lambda_{kj} - \sum_{j:j\neq k} \lambda_{jk}\overline{F}_j(z)\Big)dz, \tag{C.41}$$

thus we just need to verify that the RHS of (C.41) is bounded above by

$$\sum_{k=1}^{K} \int_{\widehat{\theta}}^{\theta_{max}} \Big(\alpha_k\mu_k(z)f_k(z) + \overline{F}_k(z) \cdot \sum_{j:j\neq k} \lambda_{kj} - \sum_{j:j\neq k} \lambda_{jk}\overline{F}_j(z)\Big)dz. \tag{C.42}$$

Note that the RHS of (C.41), for each k, is maximized at some threshold contract $\theta_k \in [0, 1]$. So to prove that (C.42) is an upper bound of (C.41) is enough to show that for all k and for any $\theta_k \in [0, 1]$

$$\int_{\theta_k}^{\theta_{max}} \Big(\alpha_k\mu_k(z)f_k(z) + \overline{F}_k(z) \cdot \sum_{j:j\neq k} \lambda_{kj} - \sum_{j:j\neq k} \lambda_{jk}\overline{F}_j(z)\Big)dz \leq \int_{\widehat{\theta}}^{\theta_{max}} \Big(\alpha_k\mu_k(z)f_k(z) + \overline{F}_k(z) \cdot \sum_{j:j\neq k} \lambda_{kj} - \sum_{j:j\neq k} \lambda_{jk}\overline{F}_j(z)\Big)dz. \tag{C.43}$$

Consider $\theta_k \geq \widehat{\theta}$ in (C.43), then (C.43) becomes

$$0 \leq \int_{\widehat{\theta}}^{\theta_k} \Big(\alpha_k\mu_k(z)f_k(z) + \overline{F}_k(z) \cdot \sum_{j:j\neq k} \lambda_{kj} - \sum_{j:j\neq k} \lambda_{jk}\overline{F}_j(z)\Big)dz,$$

this is equivalent to

$$-\Big(\sum_{j:j\neq k} \lambda_{kj}\Big) \cdot \int_{\widehat{\theta}}^{\theta_k} \overline{F}_k(z)dz \leq \int_{\widehat{\theta}}^{\theta_k} \Big(\alpha_k\mu_k(z)f_k(z) - \sum_{j:j\neq k} \lambda_{jk}\overline{F}_j(z)\Big)dz, \quad \forall\theta_k \geq \widehat{\theta},$$

which can be rewritten as

$$-\Big(\sum_{j:j\neq k}\lambda_{kj}\Big)\leq \min_{\widehat{\theta}\leq\theta}\left\{\alpha_k\frac{\int_{\widehat{\theta}}^{\theta}\mu_k(z)f_k(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_k(z)dz}-\sum_{j:j\neq k}\lambda_{jk}\cdot\frac{\int_{\widehat{\theta}}^{\theta}\overline{F}_j(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_k(z)dz}\right\}. \tag{C.44}$$

Similarly, if $\theta_k\leq\widehat{\theta}$ then (C.43) is equivalent to

$$0\geq\int_{\theta_k}^{\widehat{\theta}}\Big(\alpha_k\mu_k(z)f_k(z)+\overline{F}_k(z)\cdot\sum_{j:j\neq k}\lambda_{kj}-\sum_{j:j\neq k}\lambda_{jk}\overline{F}_j(z)\Big)dz,\quad\forall\theta_k\leq\widehat{\theta},$$

which is equivalent to

$$\max_{\theta\leq\widehat{\theta}}\left\{\alpha_k\frac{\int_{\theta}^{\widehat{\theta}}\mu_k(z)f_k(z)dz}{\int_{\theta}^{\widehat{\theta}}\overline{F}_k(z)dz}-\sum_{j:j\neq k}\lambda_{jk}\cdot\frac{\int_{\theta}^{\widehat{\theta}}\overline{F}_j(z)dz}{\int_{\theta}^{\widehat{\theta}}\overline{F}_k(z)dz}\right\}\leq-\Big(\sum_{j:j\neq k}\lambda_{kj}\Big). \tag{C.45}$$

In summary, proving that (C.43) holds is equivalent to showing that both (C.44) and (C.45) hold. To see why this is true, note that

$$\begin{aligned}\lim_{\theta\to\widehat{\theta}+}\alpha_k\frac{\int_{\widehat{\theta}}^{\theta}\mu_k(z)f_k(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_k(z)dz}-\sum_{j:j\neq k}\lambda_{jk}\cdot\frac{\int_{\widehat{\theta}}^{\theta}\overline{F}_j(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_k(z)dz}&=\frac{\alpha_k\cdot\mu_k(\widehat{\theta})\cdot f_k(\widehat{\theta})-\sum_{j:j\neq k}\lambda_{jk}\cdot\overline{F}_j(\widehat{\theta})}{\overline{F}_k(\widehat{\theta})}\\&=-\Big(\sum_{j:j\neq k}\lambda_{kj}\Big),\end{aligned} \tag{C.46}$$

where the last equality comes from the choice of the multipliers. Since the limit is taken for values above $\widehat{\theta}$, this implies that

$$\min_{\widehat{\theta}\leq\theta}\left\{\alpha_k\frac{\int_{\widehat{\theta}}^{\theta}\mu_k(z)f_k(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_k(z)dz}-\sum_{j:j\neq k}\lambda_{jk}\cdot\frac{\int_{\widehat{\theta}}^{\theta}\overline{F}_j(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_k(z)dz}\right\}$$

is bounded above by

$$\lim_{\theta\to\widehat{\theta}+}\alpha_k\frac{\int_{\widehat{\theta}}^{\theta}\mu_k(z)f_k(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_k(z)dz}-\sum_{j:j\neq k}\lambda_{jk}\cdot\frac{\int_{\widehat{\theta}}^{\theta}\overline{F}_j(z)dz}{\int_{\widehat{\theta}}^{\theta}\overline{F}_k(z)dz}=-\Big(\sum_{j:j\neq k}\lambda_{kj}\Big).$$

A similar argument(taken the limit for values below $\widehat{\theta}$ this time) can be used to show that

$$-\Big(\sum_{j:j\neq k}\lambda_{kj}\Big)\leq\max_{\theta\leq\widehat{\theta}}\left\{\alpha_k\frac{\int_{\theta}^{\widehat{\theta}}\mu_k(z)f_k(z)dz}{\int_{\theta}^{\widehat{\theta}}\overline{F}_k(z)dz}-\sum_{j:j\neq k}\lambda_{jk}\cdot\frac{\int_{\theta}^{\widehat{\theta}}\overline{F}_j(z)dz}{\int_{\theta}^{\widehat{\theta}}\overline{F}_k(z)dz}\right\}.$$

Since we are assuming that the minimum is an upper bound to the maximum above, we can conclude that both (C.44) and (C.45) hold (with equality). This concludes the proof for the first direction.

For the second direction we need to show that if the static contract is optimal then we can find $\boldsymbol{\lambda}$ satisfying condition (APR^M). Theorem 1 in Luenberger (1969, p. 217) gives then the existence of Lagrange multipliers such that the static contract maximizes the Lagrangian(here we use the interior point condition in the assumptions). In other words, $\exists \boldsymbol{\lambda}, \boldsymbol{w} \geq 0$ such that

$$\mathcal{L}(\mathbf{x}^{\mathbf{s}}, \mathbf{0}, \boldsymbol{\lambda}, \boldsymbol{w}) \geq \mathcal{L}(\mathbf{x}, \boldsymbol{u}, \boldsymbol{\lambda}, \boldsymbol{w}), \quad \forall \boldsymbol{u}, \mathbf{x} \in \mathbb{R}_+^K \times \Omega^K. \tag{C.47}$$

Note that (C.47) holds for any $\boldsymbol{u}, \mathbf{x} \in \mathbb{R}_+^K \times \Omega^K$. Thus we can first consider $\mathbf{x}$ equal to $\mathbf{x}^{\mathbf{s}}$ in (C.47), this yields

$$0 \geq \sum_{k=1}^{K} u_k \Big(-\alpha_k + w_k + \sum_{j:j \neq k} \lambda_{kj} - \sum_{j:j \neq k} \lambda_{jk} \Big), \quad \forall \boldsymbol{u} \in \mathbb{R}_+^K.$$

Which implies that

$$-\alpha_k + w_k + \sum_{j:j \neq k} \lambda_{kj} - \sum_{j:j \neq k} \lambda_{jk} = 0, \quad \forall k,$$

and since $w_k \geq 0$ we can conclude that

$$\alpha_k \geq \sum_{j:j \neq k} \lambda_{kj} - \sum_{j:j \neq k} \lambda_{jk}, \quad \forall k,$$

as required. Now, fix k and consider a solution $\mathbf{x} \in \Omega^K$ such that $x_j \equiv x^s$ for all $j \neq k$ and x_k is $\mathbf{1}_{\{\theta \geq \theta_k\}}$ for some $\theta_k \in [0, 1]$. Then equation (C.47) delivers equation (C.43). And we already saw that (C.43) is equivalent to both equations (C.44) and (C.45). Putting these two equations together yields

$$\max_{\theta \leq \widehat{\theta}} \left\{ \alpha_k \frac{\int_\theta^{\widehat{\theta}} \mu_k(z) f_k(z) dz}{\int_\theta^{\widehat{\theta}} \overline{F}_k(z) dz} - \sum_{j:j \neq k} \lambda_{jk} \cdot \frac{\int_\theta^{\widehat{\theta}} \overline{F}_j(z) dz}{\int_\theta^{\widehat{\theta}} \overline{F}_k(z) dz} \right\} \leq -\Big(\sum_{j:j \neq k} \lambda_{kj} \Big)$$

which is bounded above by

$$\min_{\widehat{\theta} \leq \theta} \left\{ \alpha_k \frac{\int_{\widehat{\theta}}^{\theta} \mu_k(z) f_k(z) dz}{\int_{\widehat{\theta}}^{\theta} \overline{F}_k(z) dz} - \sum_{j:j \neq k} \lambda_{jk} \cdot \frac{\int_{\widehat{\theta}}^{\theta} \overline{F}_j(z) dz}{\int_{\widehat{\theta}}^{\theta} \overline{F}_k(z) dz} \right\},$$

that is, condition (APR^M) holds for any k. We only need to check that $\boldsymbol{\lambda} \in \mathcal{A}$. Observe that both the maximum and the minimum are bounded from below and above (respectively) by

$$\frac{\alpha_k \cdot \mu_k(\widehat{\theta}) \cdot f_k(\widehat{\theta}) - \sum_{j:j\neq k} \lambda_{jk} \cdot \overline{F}_j(\widehat{\theta})}{\overline{F}_k(\widehat{\theta})}. \tag{C.48}$$

To see this we can take the limit as before. For the maximum we take the limit of θ approaching to $\widehat{\theta}$ from below. This limit converges to the expression in (C.48) and is bounded above by the maximum. The same argument applies to the minimum but this time taking the limit from above $\widehat{\theta}$. In turn implies that

$$\frac{\alpha_k \cdot \mu_k(\widehat{\theta}) \cdot f_k(\widehat{\theta}) - \sum_{j:j\neq k} \lambda_{jk} \cdot \overline{F}_j(\widehat{\theta})}{\overline{F}_k(\widehat{\theta})} = -\Big(\sum_{j:j\neq k} \lambda_{kj}\Big),$$

and we can conclude that $\boldsymbol{\lambda} \in \mathcal{A}$. $\square$

Lemma C.3 *The set $\mathcal{A}$ is non-empty.*

Proof of Lemma C.3. We want to show that $\mathcal{A} \neq \emptyset$, which amount to proving that the linear system

$$\sum_{j=1,j\neq k}^{K} \lambda_{jk} \cdot \overline{F}_j(\widehat{\theta}) = \alpha_k \cdot \mu_k(\widehat{\theta}) \cdot f_k(\widehat{\theta}) + \overline{F}_k(\widehat{\theta}) \cdot \sum_{j=1,j\neq k}^{K} \lambda_{kj}, \quad \forall k \in \{1, \ldots, K\},$$

$$\alpha_k = w_k + \sum_{j=1,j\neq k}^{K} \lambda_{kj} - \sum_{j=1,j\neq k}^{K} \lambda_{jk}, \quad \forall k \in \{1, \ldots, K\},$$

with $(\lambda, \mathbf{w}) \geq 0$ has a solution. We begin by writing down the system with matrices and then we apply Farkas' lemma.

First, the vector λ is given by

$$(\underbrace{\lambda_{12}, \lambda_{13}, \cdots, \lambda_{1K}}_{Type1}, \underbrace{\lambda_{21}, \lambda_{23}, \cdots, \lambda_{2K}}_{Type2}, \cdots, \underbrace{\lambda_{K1}, \lambda_{K2}, \cdots, \lambda_{KK-1}}_{TypeK}),$$

note that the terms λ_{kk} for any $k \in \{1, \ldots, K\}$ do not form part of the vector. Now, consider matrix A with $K(K-1)+K$ columns and $2K$ rows given by

$$A = \begin{bmatrix} \mathbf{F}^1 & \mathbf{F}^2 & \cdots & \mathbf{F}^K & 0_{K\times K} \\ B^1 & B^2 & \cdots & B^K & I_{K\times K} \end{bmatrix},$$

where $0_{K\times K}$ is the zero matrix of dimension $K\times K$ and $I_{K\times K}$ is the identity matrix of dimension $K\times K$. Also, $\mathbf{F}^k$ is a matrix of dimension $K\times(K-1)$ defined by

$$\mathbf{F}^k_{ij}=\begin{cases}-\overline{F}_k(\widehat{\theta}) & \text{if } i=k\\ \overline{F}_k(\widehat{\theta}) & \text{if } i<k, j=i\\ \overline{F}_k(\widehat{\theta}) & \text{if } i>k, j=i-1\\ 0 & \text{if } o.w\end{cases},$$

and B^k is a matrix of dimension $K\times(K-1)$ defined by

$$B^k_{ij}=\begin{cases}1 & \text{if } i=k\\ -1 & \text{if } i<k, j=i\\ -1 & \text{if } i>k, j=i-1\\ 0 & \text{if } o.w\end{cases}.$$

Finally, let b be a vector defined by

$$b=(\alpha_L\mu_1(\widehat{\theta})f_1(\widehat{\theta}),\alpha_2\mu_2(\widehat{\theta})f_2(\widehat{\theta}),\cdots,\alpha_K\mu_K(\widehat{\theta})f_K(\widehat{\theta}),\alpha_L,\cdots,\alpha_K).$$

Then, the linear system can be rewritten as

$$A\cdot\begin{bmatrix}\lambda\\ \mathbf{w}\end{bmatrix}=b,\quad \lambda,\mathbf{w}\geq 0.$$

Now we use Farkas' lemma, if this system does not have a solution then it must be the case that the following system has a solution

$$A^{\intercal}\cdot\begin{bmatrix}y^F\\ y^B\end{bmatrix}\geq 0,\quad b^{\intercal}\cdot\begin{bmatrix}y^F\\ y^B\end{bmatrix}<0. \tag{C.49}$$

Explicitly, we have (y^F,y^B) solve

$$\overline{F}_k(\widehat{\theta})\cdot(y^F_j-y^F_k)-(y^B_j-y^B_k)\geq 0,\quad \forall k,\forall j\neq k$$
$$y^B_k\geq 0,\quad \forall k$$
$$\sum_{k=1}^{K}\alpha_k\mu_k(\widehat{\theta})f_k(\widehat{\theta})\cdot y^F_k+\sum_{k=1}^{K}\alpha_k\cdot y^B_k<0.$$

Let y_m^F be equal to $\min_k\{y_k^F\}$ (m is the index that achieves the minimum) then

$$\begin{aligned}
\sum_{k=1}^K \alpha_k\mu_k(\widehat{\theta})f_k(\widehat{\theta})\cdot y_k^F + \sum_{k=1}^K \alpha_k\cdot y_k^B &\overset{(a)}{=} \sum_{k=1}^K \alpha_k\mu_k(\widehat{\theta})f_k(\widehat{\theta})\cdot(y_k^F - y_m^F) + \sum_{k=1}^K \alpha_k\cdot y_k^B \\
&= \sum_{k=1}^K \alpha_k\Big(\widehat{\theta} - \frac{\overline{F}_k(\widehat{\theta})}{f_k(\widehat{\theta})}\Big)f_k(\widehat{\theta})\cdot(y_k^F - y_m^F) + \sum_{k=1}^K \alpha_k\cdot y_k^B \\
&= \sum_{k=1}^K \alpha_k\Big(\widehat{\theta}f_k(\widehat{\theta}) - \overline{F}_k(\widehat{\theta})\Big)\cdot(y_k^F - y_m^F) + \sum_{k=1}^K \alpha_k\cdot y_k^B \\
&\overset{(b)}{\geq} -\sum_{k=1}^K \alpha_k\overline{F}_k(\widehat{\theta})\cdot(y_k^F - y_m^F) + \sum_{k=1}^K \alpha_k\cdot y_k^B \\
&= \sum_{k=1}^K \alpha_k\overline{F}_k(\widehat{\theta})\cdot(y_m^F - y_k^F) + \sum_{k=1}^K \alpha_k\cdot y_k^B \\
&\overset{(c)}{\geq} \sum_{k=1}^K \alpha_k\cdot(y_m^B - y_k^B) + \sum_{k=1}^K \alpha_k\cdot y_k^B \\
&= \sum_{k=1}^K \alpha_k\cdot y_m^B \\
&\overset{(d)}{=} y_m^B \geq 0,
\end{aligned}$$

a contradiction. Where in (a) we use the fact that $\sum_{k=1}^K \alpha_k\mu_k(\widehat{\theta})f_k(\widehat{\theta}) = 0$, in (b) we use the definition of y_m^F, in (c) we use the first set of equations in (C.49) and in (d) we use the fact that $\sum_{k=1}^K \alpha_k = 1$ and $y_m^B \geq 0$. $\square$

Proof of Proposition 3.4. We make use of Lemma C.4 which we state and prove after the present proof. In that lemma we need to define the function

$$L_k(z|\boldsymbol{\lambda}) \triangleq \alpha_k\mu_k(z) + \frac{\bar{F}_k(z)}{f_k(z)}\cdot\sum_{\ell:\ell\neq k}\lambda_{k\ell} - \sum_{\ell:\ell\neq k}\lambda_{\ell k}\frac{\bar{F}_\ell(z)}{f_k(z)},$$

for any $\boldsymbol{\lambda} \geq 0$. For exponential distributions $L_k(z|\boldsymbol{\lambda})$ becomes:

$$L_k(z|\boldsymbol{\lambda}) = \underbrace{\alpha_k\cdot z + \frac{1}{\lambda_k}\cdot\Big(\sum_{\ell:\ell\neq k}\lambda_{k\ell} - \alpha_k\Big)}_{\text{linear}} - \underbrace{\sum_{\ell:\ell>k}\lambda_{\ell k}\frac{e^{-z(\lambda_l-\lambda_k)}}{\lambda_k}}_{\text{increasing and convex}} - \underbrace{\sum_{\ell:\ell<k}\lambda_{\ell k}\frac{e^{-z(\lambda_l-\lambda_k)}}{\lambda_k}}_{\text{decreasing and convex}}.$$

Hence, $L_k(\cdot|\boldsymbol{\lambda})$ is concave, which means that it crosses zero at most two times. Using Lemma C.4 we conclude that in the exponential case allocations have at most one step in which randomization occurs. $\square$

Lemma C.4 *For any dual-feasible variable $\boldsymbol{\lambda}$ associated to the IC constraints define*

$$L_k(z|\boldsymbol{\lambda}) \triangleq \alpha_k \mu_k(z) + \frac{\bar{F}_k(z)}{f_k(z)} \cdot \sum_{\ell:\ell\neq k} \lambda_{k\ell} - \sum_{\ell:\ell\neq k} \lambda_{\ell k} \frac{\bar{F}_\ell(z)}{f_k(z)}. \tag{F}$$

If $L_k(z|\boldsymbol{\lambda})$ crosses zero at most p times then the optimal allocation x_k has at most $\lfloor p/2 \rfloor$ intervals where randomization occurs.

Proof of Lemma C.4. We divide the proof into two parts. In the first part we construct a new dual problem and state the complementary slackness conditions. This part of the proof follows the general theory of linear programming in infinite dimensional space developed by [6]. In the second part we exploit the complementary slackness conditions to show that the optimal allocation x_k has at most $\lfloor p/2 \rfloor$ intervals where randomization occurs.

Part 1. Define the cone of non-negative non-decreasing functions

$$\mathcal{K} \triangleq \{x : [0, \theta_{max}] \to \mathbb{R} | x \text{ is non-negative and non-decreasing function}\}. \tag{Primal Cone}$$

The general formulation of the seller's problem is

$$
\begin{aligned}
(\mathcal{P}) \quad \max \quad & -\sum_{k=1}^{K} \alpha_k u_k + \sum_{k=1}^{K} \alpha_k \int_0^{\theta_{max}} x_k(z)\mu_k(z) f_k(z) dz \\
\text{s.t} \quad & x_k(\cdot) \in \mathcal{K}, \quad \forall k \in \{1, \ldots, K\} \\
& x_k(\theta) \leq 1, \quad \forall \theta \in [0, \theta_{max}] \quad , \forall k \in \{1, \ldots, K\} \\
& u_k \geq 0, \quad \forall k \in \{1, \ldots, K\} \\
& u_k + \int_0^{\theta_{max}} x_k(z) \bar{F}_k(z) dz \geq u_{k'} + \int_0^{\theta_{max}} x_{k'}(z) \bar{F}_k(z) dz, \quad \forall k, k' \in \{1, \ldots, K\}.
\end{aligned}
$$

Note that the dual cone of $\mathcal{K}$ is

$$\mathcal{K}^* = \{\beta : \int_\theta^{\theta_{max}} \beta(z) dz \geq 0, \quad \forall \theta \in [0, \theta_{max}]\}. \tag{Dual Cone}$$

The Lagrangian is

$$\begin{aligned}\mathcal{L}(x,u,\boldsymbol{\lambda},\beta,\mathbf{w}) = &\sum_{k=1}^{K} u_k \cdot \Big(-\alpha_k + w_k + \sum_{\ell:\ell\neq k}\lambda_{k\ell} - \sum_{\ell:\ell\neq k}\lambda_{\ell k}\Big) \\ &+ \sum_{k=1}^{K}\int_0^{\theta_{max}} x_k(z)\Big(\alpha_k\mu_k(z)f_k(z) + \bar{F}_k(z)\cdot\sum_{\ell:\ell\neq k}\lambda_{k\ell} \\ &- \sum_{\ell:\ell\neq k}\lambda_{\ell k}\bar{F}_\ell(z) + \beta_k(z) - \eta_k(z)\Big)dz \\ &+ \sum_{k=1}^{K}\int_0^{\theta_{max}} \eta_k(z)dz,\end{aligned}$$

where β_k are the dual variables associated with the monotonicity constraints, η_k are dual variables associated with the constraints $x_k(\theta) \leq 1$. While $\boldsymbol{\lambda}, \boldsymbol{w}$ correspond to the dual variables associated with the IC an non-negativity constraints respectively. This yields the following Dual program (D):

$$\begin{aligned}(D) \quad \min \quad &\sum_{k=1}^{K}\int_0^{\theta_{max}} \eta_k(z)dz \\ \text{s.t} &-\alpha_k + w_k + \sum_{\ell:\ell\neq k}\lambda_{k\ell} - \sum_{\ell:\ell\neq k}\lambda_{\ell k} = 0, \quad \forall k \\ &\alpha_k\mu_k(z)f_k(z) + \bar{F}_k(z)\cdot\sum_{\ell:\ell\neq k}\lambda_{k\ell} - \sum_{\ell:\ell\neq k}\lambda_{\ell k}\bar{F}_\ell(z) \\ &= \eta_k(z) - \beta_k(z), \quad \forall k, \quad \forall z\in[0,\theta_{max}] \\ &\lambda, \mathbf{w}, \eta_k(\cdot) \geq 0, \quad \beta_k \in \mathcal{K}^*, \quad \forall k.\end{aligned}$$

And we must have complementary slackness:

- For the monotonicity constraints (the cone constraints) this means that if $x_k(\cdot)$ changes at some θ then $\int_\theta^{\theta_{max}} \beta_k(z)dz = 0$. Also $x(0)\cdot\int_0^{\theta_{max}}\beta(z)dz = 0$. All of this for all k.

- For the upper bound constraints: $(1 - x_k(\theta))\cdot\eta_k(\theta) = 0$ for all $\theta \in [0, \theta_{max}]$ and for all k.

Part 2. Consider an optimal primal-dual pair. Let x_k be the primal solution for interim type k, and β_k, η_k and $\lambda, \mathbf{w}$ the corresponding dual solutions. Observe that from dual feasibility we must have

$$f_k(z) \cdot L_k(z|\boldsymbol{\lambda}) = \eta_k(z) - \beta_k(z), \quad \forall z \in [0, \theta_{max}]. \tag{C.50}$$

Let us denote by $\hat{z}_1 < \cdots < \hat{z}_p$ the points where $L_k(\cdot|\boldsymbol{\lambda})$ crosses zero, and we let $\hat{z}_0 = 0$ and $\hat{z}_{p+1} = \theta_{max}$. Note that $L_k(\theta_{max}|\boldsymbol{\lambda}) = \alpha \cdot \theta_{max} > 0$, and by the feasibility of $\boldsymbol{\lambda}$ we have $L_k(0|\boldsymbol{\lambda}) = -w_k/f_k(0) \leq 0$.

Let $z_1^\star \triangleq \inf\{z \in [0, \theta_{max}] : x_k(z) = 1\}$ (if $x_k(z)$ never equals 1 we take $z_1^\star = \theta_{max}$). We can assume that $z_1^\star > 0$, otherwise $x_k(z)$ would be equal to 1 everywhere in $[0, \theta_{max}]$ and the result would follow. In turn, there has to be a change on x_k around $z_1^\star$ and, therefore, complementary slackness implies that $\int_{z_1^\star}^{\theta_{max}} \beta_k(z)dz = 0$. Moreover, since $x_k(z) < 1$ for all $z < z_1^\star$ complementary slackness implies that $\eta_k(z) = 0$ for all $z < z_1^\star$. Therefore, Eq. (C.50) becomes

$$f_k(z) \cdot L_k(z|\boldsymbol{\lambda}) = -\beta_k(z), \quad \forall z \in [0, z_1^\star). \tag{C.51}$$

Let q be the largest index in $\{0, 1, \ldots, p\}$ such that $\hat{z}_q \leq z_1^\star$. Note that $z_1^\star \in [\hat{z}_q, \hat{z}_{q+1}]$. We show the following claim:

Claim 1. $L_k(\cdot|\boldsymbol{\lambda})$ is positive in $(\hat{z}_q, \hat{z}_{q+1})$ and $z_1^\star = \hat{z}_q$.

Proof of Claim 1. First suppose that $L_k(\cdot|\boldsymbol{\lambda})$ is positive in $(\hat{z}_q, \hat{z}_{q+1})$ we show that $z_1^\star = \hat{z}_q$. If not then for any $z \in (\hat{z}_q, z_1^\star)$ we have $L_k(z|\boldsymbol{\lambda}) > 0$ which thanks to Eq. (C.51) yields $\beta_k(z) < 0$ for any $z \in (\hat{z}_q, z_1^\star)$ and, therefore,

$$\int_z^{\theta_{max}} \beta_k(z)dz = \int_z^{z_1^\star} \beta_k(z)dz + \underbrace{\int_{z_1^\star}^{\theta_{max}} \beta_k(z)dz}_{=0} = \int_z^{z_1^\star} \beta_k(z)dz < 0, \tag{C.52}$$

but this contradicts the fact that $\beta_k \in \mathcal{K}^*$. That is, $z_1^\star \leq \hat{z}_q$ but since $\hat{z}_q \leq z_1^\star$ we conclude that $\hat{z}_q = z_1^\star$. To complete the argument suppose $L_k(\cdot|\boldsymbol{\lambda})$ is negative in $(\hat{z}_q, \hat{z}_{q+1})$ then, in particular, $L_k(\cdot|\boldsymbol{\lambda})$ is negative in $(z_1^\star, \hat{z}_{q+1})$ and from Eq. (C.50) we deduce that $\beta_k(z') > 0$ for all $z_1'^{\star}, \hat{z}_{q+1})$. Hence, for any $z_1'^{\star}, \hat{z}_{q+1})$

$$0 = \int_{z_1^\star}^{\theta_{max}} \beta_k(z)dz = \underbrace{\int_{z_1^\star}^{z'} \beta_k(z)dz}_{>0} + \underbrace{\int_{z'}^{\theta_{max}} \beta_k(z)dz}_{\geq 0} > 0, \qquad \text{(C.53)}$$

a contradiction. In the second bracket we use the fact that $\beta_k \in \mathcal{K}^*$. This concludes the proof of Claim 1.

This shows that $x_k(\cdot)$ equals 1 in $(\hat{z}_q, \theta_{max}]$ and that it changes value at $\hat{z}_q$. Now, from Claim 1 we now that $L_k(\cdot|\boldsymbol{\lambda})$ is negative in $(\hat{z}_{q-1}, \hat{z}_q)$ and, therefore, from Eq. (C.51) we deduce that $\beta_k(\cdot)$ is positive in $(\hat{z}_{q-1}, \hat{z}_q)$. This together with $\int_{z_1^\star}^{\theta_{max}} \beta_k(z)dz = 0$ imply that $x_k(\cdot)$ is constant in $(\hat{z}_{q-1}, \hat{z}_q)$ (by means of complementary slackness any change would yield a contradiction). Let's denote the value of $x_k(\cdot)$ in $(\hat{z}_{q-1}, \hat{z}_q)$ by χ_q. Note that of $\chi_q = 0$ we are done. Similarly to what we did before we define $z_2^\star \triangleq \inf\{z \in [0, \hat{z}_{q-1}] : x_k(z) = \chi_q\}$. Note that $z_2^\star < \hat{z}_{q-1}$. If $z_2^\star = 0$ then we $x_k(\cdot)$ equals χ_q for all values below z_q and, therefore, there is nothing more to prove. So assume $z_2^\star > 0$. If $z_2^\star = \hat{z}_{q-1}$ then $x_k(\cdot)$ changes value at $\hat{z}_{q-1}$ and, therefore, by complementary slackness $\int_{\hat{z}_{q-1}}^{\theta_{max}} \beta_k(z)dz = 0$. However, $L_k(\cdot|\boldsymbol{\lambda})$ is positive in $(\hat{z}_{q-2}, \hat{z}_{q-1})$ which by Eq. (C.51) implies that β_k is negative in $(\hat{z}_{q-2}, \hat{z}_{q-1})$ but this would contradict the dual feasibility of β_k. Hence, we can assume that $z_2^\star < \hat{z}_{q-1}$.

Let q_2 be the largest index in $\{0, 1, \ldots, q-1\}$ such that $\hat{z}_{q_2} \leq z_2^\star$. Note that $z_2^\star \in [\hat{z}_{q_2}, \hat{z}_{q_2+1}]$. As before we can show that $L_k(\cdot|\boldsymbol{\lambda})$ is positive in $(\hat{z}_{q_2}, \hat{z}_{q_2+1})$ and $z_2^\star = \hat{z}_{q_2}$. Note that this implies that the value χ_q of $x_k(\cdot)$ extends for at least two intervals, namely, $(\hat{z}_{q-2}, \hat{z}_{q-1})$ and $(\hat{z}_{q-1}, \hat{z}_q)$.

The previous argument can be applied iteratively over all intervals defined by $\hat{z}_1 < \cdots < \hat{z}_p$. Since in each step of the argument we cover two interval we deduce that there can be at most $\lfloor p/2 \rfloor$ different value of $\chi_{q'}$ where q' is defined in every step as we did before. Moreover, if $L_k(0|\boldsymbol{\lambda}) < 0$ then in the interval $(0, \hat{z}_1)$ the dual variable $\beta_k(\cdot)$ is positive. Because $\int_{\hat{z}_1}^{\theta_{max}} \beta_k(z)dz = 0$ (this follows from the steps of the argument) and $x(0) \cdot \int_0^{\theta_{max}} \beta(z)dz = 0$ we must have $x(0) = 0$ and so in the last interval x_k equals 0. $\square$