THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

Christophe-alain BRUNEEL-ZUPANC

Le 6 juillet 2021

Essays in Structural Econometrics

Ecole doctorale : **TSE - Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche :
UMR 5314 -TSE-R

Thèse dirigée par
Thierry MAGNAC

Jury

M. Robert MILLER, Rapporteur
M. Arnaud MAUREL, Rapporteur
M. Olivier DE GROOTE, Examinateur
M. Frank VERBOVEN, Examinateur
M. Thierry MAGNAC, Directeur de thèse
M. Pierre DUBOIS, Président

Essays in Structural Econometrics

Ph.D. Thesis

Christophe Bruneel-Zupanc[1]
Toulouse School of Economics

July 2021

[1]E-mail address: christophe.bruneel@tse-fr.eu.

Acknowledgements

First, I would like to thank my advisor Thierry Magnac for his invaluable guidance and supervision throughout my PhD. There is no word to describe how much I benefited from the numerous discussions we had together.

I am also very grateful to Olivier de Groote whose help has been crucial on the Job market. I am also indebted to many researchers in Toulouse who helped me improve my work along the years. Thanks to Christian Bontemps, Fabrice Collard, Pierre Dubois, Patrick Fève, Jean-Pierre Florens, Eric Gautier, Ana Gazmuri, Cristina Gualdani, Vishal Kamat, Jihyun Kim, Pascal Lavergne, Nour Meddahi, Shruti Sinha and François Poinas for all their questions and comments.

I would also like to thank Robert Miller for having me in Carnegie Mellon University. This visit has been decisive in determining my research interests.

Very importantly, I thank Christian Ben Lakhdar, Grégoire Cauchie, Fabrice Le Lec and especially Nicolas Vaillant, whom made me want and helped me do research in the first place.

I am also grateful to José Alfonso Muñoz-Alvarado, Guillaume Chapelle, Jean-Benoit Eymeoud, Gyung Mo Kim, Tuuli Vanhapelto, Etienne Wasmer, Philipp Wangner as well as many fellow PhD students in Toulouse for making this journey more joyful, discussing and working with me along the way.

In particular, I thank Alexandre Gaillard, not only for the work we are doing together, but for being an excellent friend without whom my years in Toulouse would not have been half as interesting.

Finally, I thank my family, especially my parents without whose support and education I would not have been able to study all these years.

Last but most important, I have no words to thank my wife Kaja Balcer who has been the best partner I could have hoped for through these years. I dedicate this thesis to her and our daughter Constance, whose recent birth brought us immeasurable joy!

Overview

The first chapter develops a general framework for models, static or dynamic, in which agents simultaneously make both discrete and continuous choices. I show that such models are non-parametrically identified. Based on the constructive identification arguments, I build a novel two-step estimation method in the lineage of Hotz and Miller (1993) but extended to discrete and continuous choice models. The method is especially attractive for complex dynamic models because it significantly reduces the computational burden associated with their estimation. To illustrate my new method, I estimate a dynamic micro-model of female labor supply and consumption. The method is also illustrated in the third chapter of the thesis.

In the second chapter, I build a dynamic search model to examine the decision problem of a homeowner who maximizes her expected profit from the sale of her property when market conditions are uncertain. Using a large dataset of real estate transactions in Pennsylvania between 2011 and 2014, I verify several stylized facts about the housing market. I develop a dynamic search model of the home-selling problem in which the homeowner learns about demand in a Bayesian way. I estimate the model and find that learning, especially the downward adjustment of the beliefs of sellers facing low demand, explains some of the key features of the housing data, such as the decreasing list price overtime and time on the market. By comparing with a perfect information benchmark, I derive an unexpected result: the value of information is not always positive. Indeed, an imperfectly informed seller facing low demand can obtain a better outcome than her perfectly informed counterpart thanks to a delusively stronger bargaining position.

In the third chapter, joint work with Thierry Magnac, we estimate a dynamic discrete and continuous choices model of households' decisions regarding their consumption, housing tenure and housing services over the life-cycle. We use non parametric identification arguments as in the first chapter to formulate an empirical strategy in two steps that (1) estimates discrete choice probabilities and continuous choices distribution summaries to be used in (2) Bellman and Euler equations that estimate the structural parameters. Specific modelling strategies are adopted because of unfrequent mobility due to housing transaction costs. Counterfactuals that can be evaluated are related to those transaction costs as well as of prudential policies such as downpayments.

Contents

Chapter 1

Discrete-Continuous Dynamic Choice Models: Identification and Conditional Choice Probability Estimation

Christophe Bruneel-Zupanc[1]

Abstract

This paper develops a general framework for models, static or dynamic, in which agents simultaneously make both discrete and continuous choices. I show that such models are non-parametrically identified. Based on the constructive identification arguments, I build a novel two-step estimation method in the lineage of Hotz and Miller (1993) but extended to discrete and continuous choice models. The method is especially attractive for complex dynamic models because it significantly reduces the computational burden associated with their estimation. To illustrate my new method, I estimate a dynamic model of female labor supply and consumption.

Keywords: Discrete and continuous choice, dynamic model, identification, structural estimation, female labor supply.

[1]E-mail address: christophe.bruneel@tse-fr.eu.

1.1 Introduction

Many economic problems involve joint discrete and continuous choices. For example, a firm can decide on a pricing scheme (per unit, flat rate) and the corresponding price level (Timmins, 2002). It can decide what to produce and the corresponding sales price (Crawford et al., 2019). Firms also decide whether to register their business and how many workers to hire (Ulyssea, 2018). Students select their majors and decide how much effort to exert into their study (Ahn et al., 2019). Consumers can decide what to buy and how much of it to consume (e.g. appliance choice and demand for energy, Dubin and McFadden, 1984). In housing, buyers decide their house size and housing tenure (Hanemann, 1984; Bajari et al., 2013). The buyer of a car selects a model and the mileage of the car (Bento et al., 2009). Individuals decide whether to retire or not and how much they plan to consume accordingly (Iskhakov et al., 2017). Similarly, labor force participation and consumption are joint choices for potential workers (Altuğ and Miller, 1998; Blundell et al., 2016a).

In these cases, a rational agent makes both decisions *simultaneously*. Here 'simultaneous' means that given the information she has, the agent jointly make both choices. As a result, the discrete choice is endogenous with respect to the continuous choice and vice versa. To take a labor market example, if she works, a woman consumes differently than if she does not work: she has two different conditional consumption choices. And at the same time, her decision of whether to work or not is dependent on these two conditional continuous choices. Unfortunately, the identification of models with simultaneous choices is difficult (Matzkin, 2007). Indeed, there is a core *observability problem* because we only observe the continuous choice made in the selected discrete alternative, and we do not know the counterfactual choices the individual would have made in the other alternatives. Ideally, we would like to recover counterfactual continuous choices using the choices of individuals with similar characteristics but who chose another alternative. However, doing so is not possible if individuals also differ on factors which are unobserved by the econometrician and affect both continuous and the discrete choices. In this case, two identical individuals as measured by their observed covariates might still differ on the unobserved dimension. There is a problem of *selection on unobservables*, which prevents the identification of counterfactual continuous choices. To further pursue the example, if an econometrician observes that working individuals consume more than unemployed individuals, she cannot identify whether this is because the consumption choice conditional on working is truly higher or because individuals with an unobserved higher taste for consumption select themselves more into working.

This paper develops a general framework of *simultaneous discrete-continuous choice* models suited for static or dynamic problems. I provide minimal necessary conditions under which non-parametric identification of the model can be obtained, using an *instrument* for the *unobserved selection*. Then, building upon the identification, I provide an estimation method for these models. The method is attractive because it yields significant computational gains over the estimation of dynamic models, in the lineage of Hotz and Miller (1993). I also show how to apply this method to a dynamic discrete-continuous choice model of female labor force participation and consumption.

The first contribution of this paper is that I provide a constructive proof for the non-parametric identification of a general class of structural models in which agents simultaneously make a discrete and a continuous choices. To do so, I require an instrument that must be *relevant* for the selection into the discrete choice and *excluded* from the conditional continuous choices. For example, the previous discrete choice can be a good instrument in the presence of switching costs. Indeed, it impacts the current discrete decision through the switching cost. Conditional on the current discrete decision, it is excluded from the current continuous choice. In this way, observable differences in the distribution of the choices with respect to the instrument can be attributed to unobserved differences in selection and not differences in continuous choices. I show that, paired with restrictions on the effect of unobserved heterogeneity on the continuous choice (monotonicity, rank invariance), the instrument allows us to achieve non-parametric identification of the optimal discrete and continuous choices. Once the optimal choices are identified, it can be further shown that the rest of the model is identified, for example, by exploiting first-order conditions (in the spirit of Blundell et al., 1997).

The second contribution of the paper is in terms of estimation. I build a two-step estimation method, similar to Hotz and Miller (1993) but for discrete *and continuous* choices. In the first step, one estimates the policies, which I name after Hotz-Miller's CCPs: Conditional Continuous

Choices (*CCCs*), and Conditional Choice Probabilities (*CCPs*). This step is built on the identification arguments. The policies are estimated directly from the data without solving the structural model. To this end, I propose a novel method that estimates the entire monotone continuous choice functions directly instead of proceeding pointwise. In the second step, one uses the estimated CCCs and CCPs to estimate the structure of the model. For example, I exploit the fact that within my general framework, the payoffs are related to optimal choices through the first-order conditions. My estimation method is attractive because it yields sizeable *computational gains*. Typical dynamic discrete *or* continuous choice models are difficult to estimate because they involve solving the theoretical model (either by backward recursion or fixed point algorithms). Dynamic discrete-continuous choices model are even more difficult to estimate because the mixed choices can introduce kinks and non-concavities in the value function (Iskhakov et al., 2017). Given that I can recover the CCCs and CCPs in the first stage, I can exploit them to estimate the rest of the model without having to compute the value function or solve the model. It yields computational gains comparable to the computational gains generated by Hotz and Miller (1993) in the dynamic discrete choice literature. The gains are so important that they not only reduce the time required to estimate the models, but also make it possible to estimate models that have thus far been deemed computationally intractable. In this respect, my method may facilitate the use of simultaneous discrete-continuous choice models.

Finally, I illustrate my method by building and estimating a dynamic life-cycle model of women's consumption and labor force participation, in the spirit of Blundell et al. (1997, 2016a). This application has been implemented under a parametric framework for practical reasons. First, doing so avoids 'curse of dimensionality' concerns, and second, it makes my empirical findings comparable to the existing literature. I add to existing models a more flexible distribution of unobserved heterogeneity. Thanks to my method, I flexibly estimate the complete *distribution* of consumption choices and working probabilities at any given set of observed covariates (assets, earnings, family status, education, etc.). Hence, I can recover distributions such as that of the marginal propensity to consume when earnings or benefits increase for any individual. I use these estimated policies to estimate the parameters of the structural model. For example, I find a constant relative risk aversion of 1.63, close to the value of 1.56 in Blundell et al. (1994) and the value of 1.53 in Alan et al. (2009). All things considered, the method developed in this paper allows for more complete models in terms of unobserved heterogeneity, with a faster estimation and I still find estimates consistent with the existing literature. Therefore the method is very attractive in practice.

Related Literature:

There is a vast empirical literature that uses *dynamic discrete choice* models. For example, such works study labor market transition and career choice (Keane and Wolpin, 1997), fertility choice (Eckstein and Wolpin, 1989) and education choice (Arcidiacono, 2004). Starting from the bus replacement problem of Rust (1987), developments have been made in the estimation and identification of these models. They include non-exhaustively: Hotz and Miller (1993); Hotz et al. (1994); Rust (1994); Magnac and Thesmar (2002); Aguirregabiria and Mira (2002, 2007); Kasahara and Shimotsu (2009); Arcidiacono and Miller (2011); Hu and Shum (2012); Arcidiacono and Miller (2019, 2020); Abbring and Daljord (2020). For a survey, see Aguirregabiria and Mira (2010) or Arcidiacono and Ellickson (2011).

Similarly, the literature on *dynamic continuous choice* models is also voluminous, especially concerning consumption/saving choices (Carroll, 2006) or investment choices (Hong and Shum, 2010). There are also methods such as Bajari et al. (2007) that can be applied to either dynamic discrete choice models or dynamic continuous choice models (but not both).

However, in many cases, economic problems involve several joint decisions, not only one discrete choice or only one continuous choice. For example, labor force participation is very much related to saving decisions. By focusing only on one of these two dimensions and ignoring the other endogenous choice, one might be missing something important. Unfortunately, empirical applications of the dynamic discrete-continuous choice framework are less popular, as there was no generally identified setup prior to this work. For example, Blundell et al. (1997) provide identification of such models once the optimal choices are identified but do not directly address the identification of the choices. The existing literature employs several tricks to overcome the unobserved selection problem. The most common is to have implicit or explicit assumptions about the unobserved selection process. For example, Dubin and McFadden (1984), Hanemann (1984) or Bento et al. (2009) have specific assumptions about their error disturbances (independence, measurement errors, known joint distribution), which generate a specific selection process. Blevins (2014) studies the non-parametric identification of dynamic discrete-continuous choice models but assumes a very specific timing in which the discrete choice takes place before the realization of the nonseparable shock and the continuous decision. Hence, unobserved selection is not allowed to depend on nonseparable shock realization. Similarly, Iskhakov et al. (2017) break the simultaneity issue by assuming that the discrete retirement choice is taken before and based on the expectations about the continuous consumption choice. Murphy (2018) also imposes that the two choices are taken sequentially. In his paper, parcel owners first decide whether to build or not, and only afterwards, a nonseparable price shock is realized and they decide on their house size accordingly if they chose to build in the first stage. The problem is that the sequentiality of the choices is a strong assumption, and it might lead to biased results if the true decision process is in fact simultaneous. For example, in Murphy (2018), it is likely that small price realization will increase both the house size *and* the probability of building a house. The imposed timing ignores this, as the discrete building choice is only based on expectations about the price shock and corresponding house size choice. Thus it might miss part of what is truly happening in the data. My general simultaneous choice framework nests these different timing assumptions, which have testable implications within the framework. Thus I can verify when the sequentiality assumption is reasonable. Another technique is to discretize the continuous choice such that the model becomes a dynamic discrete choice model. For example, De Groote and Verboven (2019) study the adoption of solar photovoltaic systems and discretize the continuous level of adopted capacities. This is appealing, as it allows the application of known techniques in the dynamic discrete choice literature. However, discretizing the continuous choice is implicitly equivalent to making an assumption about the unobserved selection process via the assumption on the distribution of the additive discrete error terms. I show in this paper that by exploiting the continuous nature of the choice, the unobserved selection process can be identified instead of being assumed. Therefore, one can focus on the true discrete-continuous choice problem without discretizing the continuous choice. Another solution is to completely abstract from the nonseparable shock, i.e., to assume that individuals with the same observed covariates will make the same continuous choice. A more convincing alternative is to reduce the level of unobserved heterogeneity, for example, by including only a finite number of unobserved types (Blundell et al., 2016a). My approach is more general, as I allow for a more flexible distribution of unobserved heterogeneity.

The closest literature for the identification of simultaneous discrete-continuous choice can be found in static reduced-form identification analysis of non-parametric simultaneous equations (Matzkin, 2007, 2008; Imbens and Newey, 2009), nonseparable models (Chesher, 2003; Chernozhukov et al., 2020), the discrete-continuous Roy model (Newey, 2007), treatment effects with endogenous selection into treatment (Heckman and Vytlacil, 2005; Chernozhukov and Hansen, 2005, 2006, 2008), or in reduced-form identification analysis of dynamic treatment effects (Heckman and Navarro, 2007; Heckman et al., 2016). In this literature, the idea of using an instrument for non-parametric identification of simultaneous equations is frequent (Newey and Powell, 2003). However, my main contribution here is that I obtain identification under very weak and testable assumptions on the instrument. I only need a condition that the instrument is *relevant*, except at most at a finite set of points. This relevance yields a non-overlapping condition, similar to what Torgovitsky (2015) and D'Haultfœuille and Février (2015) employ in a different context with continuous treatment. Using the additional assumption that the optimal choice is monotone with respect to the unobserved nonseparable shock (as in quantile regression), the relevance is sufficient to recover identification. Indeed, I show that there exists a unique monotone function identified by the system, while if I had proceeded pointwise, uniqueness would have not held. By proceeding pointwise, other studies mentioned above need either stronger assumptions on the effect of the instrument (often regarding the rank of a matrix of the probabilities of selecting into treatment with respect to the instrument) or a different, less general setup for the selection mechanism (e.g., an additive process). To the best of my knowledge, Vuong and Xu (2017) are the only other authors to exploit the power of monotonicity in a similar fashion as I do for identification. However, they use it to relax strict monotonicity and still maintain a strong rank condition on the effect of the instrument on the selection process, while I am as general as possible with my mild condition of relevance. By developing a framework where the optimal choices take the form of a triangular simultaneous system of equations, I establish a connection and show how one can use the results from this literature on reduced-form identification to identify more general dynamic structural models.

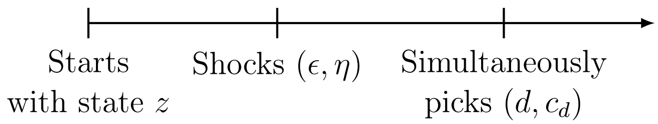
I also contribute to the literature on faster estimation methods that avoid the computation of the value function (Rust, 1987; Hotz and Miller, 1993; Hotz et al., 1994; Carroll, 2006; Arcidiacono and Miller, 2011; Iskhakov et al., 2017). I provide a faster alternative to indirect inference and the most recent developments of endogenous grid methods (Iskhakov et al., 2017). A comparison of different estimation methods can be found in section 1.6.

Finally, my application contributes to a large literature on labor market participation and consumption, focusing on women. For example, see, Heckman and Macurdy (1980); Blundell et al. (2016a). Thanks to my method, I estimate the complete distribution of individual responses.

This paper is organized as follows. Section 1.2 describes a general *simultaneous discrete-continuous choice* framework. Section 1.3 discusses identification. Section 1.4 shows how dynamic models are embedded in the framework. Section 1.5 describes the estimation method built on the constructive identification arguments. Section 1.6 compares my novel method with existing methods using Monte-Carlo simulations. Section 1.7 estimates an empirical discrete-continuous choice model of women's labor supply and consumption. Section 1.8 concludes the paper.

1.2 Framework

I consider the general problem with the following timing where the agent:



The individual simultaneously selects a discrete action $d \in \mathcal{D} = \{0, 1\}$ and accordingly makes one continuous choice $c_d \in \mathcal{C}_d \subset \mathbb{R}$ to maximize his payoff. The simultaneous decision is made given some state $z \in \mathcal{Z}$ observed by the researcher, as well as two random preference shocks $\epsilon = (\epsilon_0, \epsilon_1) \in \mathcal{E} \subset \mathbb{R}^2$ and $\eta \in \mathcal{H} \subset \mathbb{R}$ that are unobserved by the econometrician. ϵ only affects the discrete choice d, while η impacts the continuous choice c and the discrete choice. Note that the *same* η impacts the continuous choice decision in both discrete states (c_0 and c_1). In other words, I have *rank invariance* (Heckman et al., 1997; Chernozhukov and Hansen, 2005), that is, η is not discrete-choice specific.[2] The payoffs of the agent are given by the function $\mathcal{V}_d(c_d, z, \eta, \epsilon)$. The agent simultaneously selects d and c_d to solve:

$$\max_{d, c_d} \quad \mathcal{V}_d(c_d, z, \eta, \epsilon) \tag{1.1}$$

I require additional assumptions for tractability and identification of the model.

Assumption 1 (Additive Separability) *The shock ϵ enters the payoff additively such that $\forall d \in \{0, 1\}$:*

$$\mathcal{V}_d(c_d, z, \eta, \epsilon) = \tilde{v}_d(c_d, z, \eta) + \epsilon_d.$$

The additive separability assumption is usual in the discrete choice model literature (Rust, 1987; Arcidiacono and Miller, 2011). It applies to ϵ, while η can still enter the payoff in a nonseparable manner. A consequence of Assumption 1 is that the optimal continuous policy functions will not depend directly on ϵ. Indeed, c_d are defined as the (interior) solutions to the maximization of the

[2] In theory, the continuous choices could even represent different variables depending on the discrete option selected: for example, if d represents the choice between working and studying, c might represent the amount of time worked and the effort of the student respectively, hence the possibly different support. The main restriction is that even if they represent two different choices, these two continuous choices are impacted by the *same* unobserved shock η.

conditional payoff. Here, because of the additivity, we have that

$$c_d = \underset{c}{argmax}\ (\tilde{v}_d(c, z, \eta) + \epsilon_d) \iff c_d = \underset{c}{argmax}\ \tilde{v}_d(c, z, \eta).$$

Assumption 2 (Instrument) *State z contains two kinds of variables $z = (x, w)$, where $x \in \mathcal{X}$ represent general state variables and w is an instrument such that $\forall d \in \{0, 1\}$:*

$$\tilde{v}_d(c_d, z, \eta) = \tilde{v}_d(c_d, x, w, \eta) = v_d(c_d, x, \eta) + m_d(x, w, \eta).$$

The support $\mathcal{W}$ of w contains two different values, as $\mathcal{D} = \{0, 1\}$.[3]

w is an 'instrument' to recover the optimal continuous policies. On the one hand, with the additive functional form of $m_d(x, w, \eta)$, w is *excluded* from the optimal continuous policy choice. Indeed,

$$c_d \ = \ \underset{c}{argmax}\ (v_d(c, x, \eta) + m_d(x, w, \eta) + \epsilon_d) \ = \ \underset{c}{argmax}\ v_d(c, x, \eta)$$

On the other hand, it might still be *relevant* and impact the discrete choice.

Assumption 3 (Monotonicity) *The conditional payoff functions are twice continuously differentiable such that $\forall d \in \{0, 1\}$*

$$\frac{\partial^2 v_d(c_d, x, \eta)}{\partial c_d \partial \eta} > 0.$$

Assumption 3 implies that, conditional on $(D = d, X = x)$, the conditional optimal policy function $c_d^*(\eta, x)$ is C^1 and strictly increases with respect to η.[4] It ensures that there will be a one-to-one mapping from $\eta \in \mathcal{H}$ to $c_d \in \mathcal{C}_d$ for all d and x. This kind of monotonicity condition has been widely used for identification (Chernozhukov and Hansen, 2005; Bajari et al., 2007; Hong

[3] w being binary is a minimal condition for identification when d is binary. You can have discrete or even continuous w, the identification proof follows the same line, the m_d objects to be identified are just slightly different.

[4] Note Assumption 3 is equivalent to $\partial^2 \tilde{v}_d(c_d, x, w, \eta)/(\partial c_d \partial \eta) > 0$. Indeed, because of additivity in Assumption 2:

$$\frac{\partial^2 \tilde{v}_d(c_d, x, w, \eta)}{\partial c_d \partial \eta} = \frac{\partial^2 v_d(c_d, x, \eta)}{\partial c_d \partial \eta} + \frac{\partial}{\partial \eta} \underbrace{\left(\frac{\partial m_d(x, w, \eta)}{\partial c_d} \right)}_{=0}.$$

and Shum, 2010). In a sense, it means that I only identify monotone effects of the unobserved nonseparable source of heterogeneity (η, here). A very important implication of Assumption 3 is that this framework applies to problems where we observe continuous choices in *each* discrete option. For example, it does not apply directly to the problem of an investor who decides whether to invest ($d = 1$) or not ($d = 0$) and the corresponding investment conditional on investing ($d = 1$) (Hong and Shum, 2010). Indeed, in this case, $c_0^*(h) = 0$ for all h and it is not strictly increasing. However, it would apply to a discrete choice of portfolio and corresponding conditional investment. Similarly, it does not apply directly to the house construction problem of Murphy (2018), where the agent only decides of his house size if he chooses to build one $d = 1$. However, this setup still applies to a slightly modified version of the building problem where the discrete decision would be to build ($d = 1$) or to buy (or rent) an existing house ($d = 0$), and c_d would be the corresponding house size/housing service.

Assumptions 1, 2 and 3 yield the following triangular structure for the optimal choices:

$$\begin{cases} C_d = c_d^*(X, \eta) \\ D = d^*(c_0, c_1, X, W, \eta, \epsilon) \end{cases} \implies \begin{cases} C_d = c_d^*(X, \eta) \\ D = d^*(X, W, \eta, \epsilon) \end{cases}.$$

With this triangular structure, there is a link between my general structure and (reduced-form) systems of simultaneous equations (Chesher, 2003; Matzkin, 2008; Imbens and Newey, 2009), as well as with the related literature on heterogeneous/quantile treatment effects (Chernozhukov and Hansen, 2005; Vuong and Xu, 2017). To identify the structure, one needs to first identify the choices. To identify the system of choices, I need additional assumptions on the shocks.

Assumption 4 (Independent of w) *Conditional on $X = x$, the pair of shocks (ϵ, η) is independent and identically distributed and is independent of W.*

Assumption 5 (Independent Shocks) *Conditional on $X = x$, the discrete choice-specific and the continuous choice-specific shocks are independent of one another: $\eta \perp \epsilon$*

Assumption 6 (Continuous choice shock distribution) *Conditional on $X = x$, the continuous choice-specific shock η has an atomless distribution.*

Normalization 1 (Continuous shock) *Conditional on $X = x$, η is distributed as $\mathcal{U}(0, 1)$.*

Assumption 7 (Discrete choice shock distribution) *Conditional on $X = x$, the discrete choice-specific shock ϵ has continuous support and is independent and identically distributed with continuous distribution $F_{\epsilon|X=x}(\epsilon)$ over the full support $\mathbb{R}$.*

Assumption 8 (Regularity) $\forall d \in \{0, 1\}$,

$$\forall (x, w, \eta) : \quad \max_c \; v_d(c, x, w, \eta) < \infty.$$

Assumption 4 is an independence assumption between the shocks and the instrument, conditional on the other observables X. Assumption 5 assumes independence between the two shocks. Both of these assumptions are not as restrictive as they may appear. Indeed, note that the additive term $m_d(x, w, \eta)$ can be interpreted in two different ways that we cannot identify separately. First, in Assumption 2, I describe m_d as an additive part of the payoff $\tilde{v}_d$. Second, it can also be interpreted as part of a more general additive discrete shock term $\tilde{\epsilon}_d$ where $\tilde{\epsilon}_d$ could depend on w, x and η, i.e., $\tilde{\epsilon}_d(x, w, \eta) = m_d(x, w, \eta) + \epsilon_d$. Therefore, in this sense, the independence assumptions 4 and 5 on ϵ_d are still general: we could have a general $\tilde{\epsilon}_d$ that is not independent of $z = (x, w)$ or η. Then, ϵ_d is the remaining part of the discrete shock that is independent of η and w.
The main restriction is the exclusion restriction that $\eta \perp W|X$. It is crucial for the identification to have the same distribution of η, regardless of the value of w.

The atomless Assumption 6 is made to obtain smooth conditional distributions of continuous choices. Here, I cannot identify the distribution of η separately from the rest of the problem. Therefore, as is standard in the literature (Blundell et al., 1997; Matzkin, 2003), I normalize it to a uniform distribution, which represents the quantiles of any atomless continuous distribution (conditional on X), in Normalization 1.

Assumption 7 is a regularity condition on the distribution of the discrete choice-specific shock. Along with Assumption 7, Assumption 8 is another regularity condition on the functional form that ensures that $0 < Pr(d|\eta, z) < 1$ for all d, η, z. Indeed:

$$Pr(D = 0|\eta, z) = Pr\Big(\epsilon_0 - \epsilon_1 > (\max_c \; \tilde{v}_1(c, z, \eta)) - (\max_c \; \tilde{v}_0(c, z, \eta)) \mid \eta, z\Big).$$

By Assumption 7, $\epsilon_0 - \epsilon_1$ has full support $\mathbb{R}$, independent of η (Assumption 5). By Assumption 8, the payoff functions difference is bounded. Thus, $0 < Pr(D = 0|\eta, z) < 1 \; \forall \eta, z$. Since $Pr(D =$

$1|\eta, z) = 1 - Pr(D = 0|\eta, z)$, we have that:

$$\forall d, \eta, z \qquad 0 < Pr(d|\eta, z) < 1.$$

Similar to the distribution of η, the distribution of ϵ will not be identified in my setup. Thus, I need to assume that this distribution is known. Therefore, in practice, I will later follow the literature on (static or dynamic) discrete choice models (McFadden, 1980; Rust, 1987; Hotz and Miller, 1993; Matzkin, 1993; Magnac and Thesmar, 2002; Arcidiacono and Miller, 2011) and assume (generalized) extreme-value distributions. This family of distributions is convenient as it yields closed-form solutions linking the conditional value functions and the choice probabilities.[5] However, other distributions can be used (Chiong et al., 2016).

We need one last (testable) condition under which the framework is identified.

Assumption 9 (Instrument Relevance)

Assumption 9a *For any $x \in \mathcal{X}$, the additive terms of the payoff are such that there is, at most, a finite set of K (with $0 \leq K < \infty$) values h of η such that*

$$m_0(x, w = 0, h) - m_1(x, w = 0, h) = m_0(x, w = 1, h) - m_1(x, w = 1, h).$$

Assumption 9b *For any $x \in \mathcal{X}$, there exist two different values of w, denoted $w = 0$ and $w = 1$, for which the additive terms of the payoff are such that there is, at most, a finite set of K (with $0 \leq K < \infty$) values h of η such that*

$$Pr(D = 0|\eta = h, x, W = 1) - Pr(D = 0|\eta = h, x, W = 0) = 0.$$

Identification of the optimal policies requires that the instrument is *sufficiently relevant*. As stated in Assumption 9: there must be at most a finite set of values of η at which the instrument is *not* relevant for identification. In other words, $m_0(w = 0, x, h) - m_1(w = 0, x, h) \neq m_0(w = 1, x, h) - m_1(w = 1, x, h)$ or $Pr(D = 0|\eta = h, x, W = 1) \neq Pr(D = 0|\eta = h, x, W = 0)$ except at most at a finite set of values h. If this is the case, the instrument provides sufficient information to identify

[5]Notice that generalized extreme-value distributions implicitely eliminates the possible dependence of the distribution of ϵ on X. Dependence is allowed, but as the distribution is not identified anyway, one usually abstracts from it in practice.

the continuous policies. This condition is fairly intuitive and is considerably less restrictive than *full rank* assumptions and other assumptions made for the identification of heterogeneous/quantile treatment effects (Newey and Powell, 2003; Chernozhukov and Hansen, 2005, 2006, 2008). As shown later in the identification proof, the idea is that by fully exploiting the monotonicity of the conditional continuous choices, full rank conditions on the selection process with respect to the instrument are more restrictive than necessary for their identification. A similar intuition about the power of monotonicity can be found in Vuong and Xu (2017). Note that Condition 9b has *testable implications* for the observed reduced forms. It allows to test whether the structural model is identified, I will discuss this in the next section.

Lemma 1 (Equivalence) *Under Assumptions 2, 4 and 5, Assumptions 9a and 9b are equivalent.*

Proof. By construction:

$$\begin{aligned}
Pr(D = 0|\eta, x, w) = Pr\Big(&\epsilon_0 - \epsilon_1 \\
&> (\underset{c}{max}\ v_1(c, x, \eta)) - (\underset{c}{max}\ v_0(c, x, \eta)) \\
&+ \quad m_1(x, w, \eta) - m_0(x, w, \eta) \ \Big|\ \eta, x, w\Big).
\end{aligned}$$

Since $(\underset{c}{max}\ v_1(c, x, \eta)) - (\underset{c}{max}\ v_0(c, x, \eta))$ is independent of w (Assumption 2) and since $\epsilon_d \perp (w, \eta)|x$ (Assumptions 4 and 5), we have that:

$$\begin{aligned}
& Pr(D = 0|\eta, x, w = 0) \neq Pr(D = 0|\eta, x, w = 1) \\
\iff \quad & m_0(x, w = 0, \eta) - m_1(x, w = 0, \eta) \ \neq \ m_0(x, w = 1, \eta) - m_1(x, w = 1, \eta).
\end{aligned}$$

Thus, Assumption 9a expressed in terms of structural forms is equivalent to Assumption 9b on the optimal conditional choice probabilities. ■

Summary of the setup:

Under the assumptions above, I consider the general problem where an individual selects (d, c_d) to maximize his payoff:

$$\underset{d,c_d}{max} \quad v_d(c_d, x, \eta) \ + \ m_d(x, w, \eta) \ + \ \epsilon_d.$$

The general setup described here can apply not only to a wide range of static but also *dynamic* discrete-continuous choice models. I provide one static example below, and I will describe how it embeds dynamic models in section 1.4. The idea is that, in the dynamic case, v_d represents the current conditional value functions, embedding the expectations about the future, as in Hotz and Miller (1993).

Example 1: Static demand for energy

In the spirit of Dubin and McFadden (1984), consider the demand for energy with discrete appliance choice. The agent simultaneously decides between two energy sources $d = 0$ or 1 and the corresponding amount of energy she will consume (c_d). x contains observable information about the cost of each energy source and possibly the wealth or income of the agents. ϵ_d represents individual-specific unobserved preferences for each energy type. η could represent some other unobserved characteristics of the consumer impacting both her preference for the energy type and the amount of energy she wants to consume. The higher η is, the higher c_d for all d.

In practice, the greatest challenge is to find a good instrument w. Here, a good w could be some variable about the accessibility of each energy alternative. For example, the previous alternative selected by the individual might be a good instrument. First, conditional on the present alternative choice (d) and on current wealth (included in x), the past (w) should have no impact on the current energy consumption level (c_d). Thus, it would be an exogenous instrument. Moreover, changing alternatives is costly in terms of time, so individuals who were previously using energy 0 are less likely to use energy 1 now than their counterparts who were already using it. In this case, the agent incurs some disutility cost of switching from one alternative to the other and no cost if he does not switch. In other words, for all x and h, for alternative 1, $m_1(x, w = 0, h) < 0$ and $m_1(x, w = 1, h) = 0$, and for alternative 0 $m_0(x, w = 0, h) = 0$ and $m_0(x, w = 1, h) < 0$. In this case, $m_0(x, w = 0, h) - m_0(x, w = 1, h) > 0$ and $m_1(x, w = 0, h) - m_1(x, w = 1, h) < 0$, so they are different, and the instrument is relevant (Assumption 9).

Discussion of Simultaneity:

My general simultaneous choice framework nests the non-simultaneous timings where either the discrete or continuous choice is taken before and based on expectations about the other choice (and its shock realization). These two timings have *testable implications* for the optimal choices within the simultaneous framework:

- If the discrete choice is taken first, before the realization of η and the continuous choice, then the CCP $Pr(D = d|\eta, X, W)$ is *independent of* η. Indeed, η is not yet realized. The discrete choice is only based on expectations about η and the corresponding $c_d^*(x, \eta)$.

- Conversely, if the continuous choice is made first, before the discrete choice and the realization of ϵ, then the CCCs $c_d^*(x, \eta)$ are *independent of* d, i.e., $c_0^*(x, \eta) = c_1^*(x, \eta)$ $\forall \eta$.

Since I identify the policy functions c_d^* and $Pr(D = 1|\eta, X, W)$ in the simultaneous framework, I can *test the timing* of the model.

In the next section, I study the identification of the discrete-continuous choice model.

1.3 Identification

I observe data on the variables (D, C_d, X, W). I only observe C_0 if $D = 0$ and C_1 if $D = 1$. For all (x, w, η) in $\mathcal{X} \times \mathcal{W} \times \mathcal{H}$, I study non-parametric identification of the following objects: the optimal *Conditional Continuous Choices* (CCCs) $c_d^*(\eta, x)$, the optimal *Conditional Choice Probabilities* (CCPs) $Pr(d|\eta, w, x)$ for $d = 0$ and $d = 1$, and the indirect payoff functions (taken at the optimal c) $\underset{c}{max}\ v_d(c, x, \eta)$ and $m_d(x, w, \eta)$. Without loss of generality, in this section, I focus on any given x value and omit x from what follows. This is not an issue because x is exogenous in this problem, and my assumptions about the distribution of the shocks are conditional on $X = x$. First, I characterize the reduced forms and constraints imposed by the structure. Then, I discuss the identification of the optimal policies (CCCs and CCPs) and of the payoffs.

1.3.1 Reduced forms and constraints

In the data, I observe (d, c_d, w). w is exogenous in the model while c_d and d are endogenous choices. There is a *fundamental observability problem*, as I only observe one value of c_d depending on the discrete choice selected:

$$c_d = c_0(1-d) + c_1 d.$$

I do not observe both 'potential outcomes', only the selected one. Therefore, from the data, I recover the distribution of c conditional on d and w. I denote it $F_{C_d|d,w}(c_d) = Pr(C_d \leq c_d | D = d, W = w)$. I also recover the conditional probability of selecting d knowing w, denoted as $p_{d|w} = Pr(D = d|W = w)$. In other words, the data provide us with the following *reduced-form* functions, which exhaust all relevant information:

$$R = \{\{p_{d|w}\}_{(d,w)\in\{0,1\}\times\{0,1\}},\ \{F_{C_d|d,w}(c_d)\}_{\forall c_d\in\mathcal{C}_d,(d,w)\in\{0,1\}\times\{0,1\}}\}.$$

An important *vocabulary* remark is in order: in this paper, $Pr(D = d|W = w)$ is part of the *reduced forms*, while $Pr(D = d|\eta = h, W = w)$ is what I call the *conditional choice probabilities* (CCPs) or selection on unobservables (η) that I want to identify. This differs from the dynamic discrete choice literature, where $Pr(D = d|W = w)$ are actually called CCPs (Hotz and Miller, 1993; Arcidiacono and Miller, 2011). However, here, I have simultaneous choices and a nonseparable shock η, which affects both choices. Thus, the true counterparts to the usual CCPs are $Pr(D = d|\eta = h, W = w)$ for all d and not $Pr(D = d|W = w)$, hence the different terminology.

Now let us see the constraints implied by the structure on the reduced forms.

Lemma 2 *Under Assumptions 3-8 of the structural model, the distribution $F_{C_d|d,w}(c_d)$: $\mathcal{C}_d \rightarrow [0,1]$ is C^1 and strictly increasing.*

Proof. The distribution of η is C^1 and strictly increasing (Assumption 6). As previously explained, under Assumptions 5, 7 and 8, the probability of selecting d knowing $\eta = h$ is different from zero (or one) for all h and for both w (i.e., $0 < Pr(d|h,w) < 1$). As a consequence, the distribution function of η conditional on d and w is also C^1 and strictly increasing. Now, note that, by the monotonicity Assumption 3, the distribution functions of c_d (conditional on w) are strictly monotone transformations of the distribution of $\eta|d$. In other words:

$$\underbrace{Pr(\eta \leq \ h \ |d, w)}_{=F_{\eta|d,w}(h)} = \underbrace{Pr(C_d \leq \ c_d^*(h) \ |d, w)}_{=F_{C_d|d,w}(c_d^*(h))} \quad \forall d, w.$$

Therefore, since $F_{\eta|d,w}(h)$ is C^1 and strictly increasing (with respect to h), $F_{C_d|d,w}(c_d^*(h))$ is also C^1 and strictly increasing (with respect to h). Now, since $c_d^*(h)$ are C^1 and strictly increase with respect to h (Assumption 3), $F_{C_d|d,w}(c_d)$ are also C^1 and strictly increase with respect to c_d for all d. ■

Lemma 2 provides some regularity conditions on the distributions generated by the structural form. The fact that $F_{C_d|d,w}(c_d)$ are C^1 is helpful for the testable conditions of our model provided in what follows.

Lemma 3 *Under Assumption 9b in which K is defined, there is the same finite number K of values of c_0 and c_1 such that*

$$\frac{d\Big(\overbrace{F_{C_d|d,W=1}(c_d)p_{d|1} - F_{C_d|d,W=0}(c_d)p_{d|0}}^{\Delta F_{C_d}(c_d)}\Big)}{dc_d} = 0 \qquad \forall d.$$

Proof. Appendix 1.A ■

Under the relevance Assumption 9, there is only a finite number K of values h of η such that the instrument has no effect $Pr(d|\eta = h, w = 1) = Pr(d|\eta = h, w = 0)$. I will show that when this happens, we have $d(Pr(\eta \leq h|d, W = 1) - Pr(\eta \leq h|d, W = 0))/dh = 0$. Now, by the monotonicity of the optimal continuous choice, the observed conditional distributions of $C_d|d$ are transformations of the unobserved conditional distribution of $\eta|d$. Therefore, even if we do not observe the conditional distribution of $\eta|d$, we know that if the instrument is sufficiently relevant (Assumption 9), Lemma 3 will be fulfilled.

Lemma 3 yields observable and *testable* implications on the reduced forms. Indeed, the functions $\Delta F_{c_d}(c_d)$ are directly observable for all d, as is $d\Delta F_{c_d}(c_d)/dc_d$ (the derivative is well defined, cf Lemma 2). It can be used to test the relevance Assumption 9 that is crucial for identification. The idea is that if the function $\Delta F_{c_d}(c_d)$ is *flat* on a segment of values of c_d, then there is a segment of values of η such that the instrument is not relevant. In this case, the instrument has no differential impact on the conditional choice probabilities, so it does not help to identify the optimal continuous policy. If this is the case, the model is not point identified for this segment of η.

Lemmas 2 and 3 fully characterize the impact of my structure on the reduced forms. With these reduced forms, one would like to identify the structural form, i.e., the values of the payoffs $v_d(c_d^*(h), h)$ (taken at the optimal continuous choice) and $m_d(w, h)$.

The difficulty for the identification is that the shock η is unobserved and nonseparable. As a consequence, there is an unobserved variable that affects every structural object we would like to identify: the conditional payoffs $v_d(c_d^*(h), h)$ and $m_d(w, h)$, the optimal discrete choice $d^*(h, w, \epsilon)$, the corresponding conditional choice probabilities (CCPs) $Pr(d|h, w) = \mathbb{E}_\epsilon[d^*(h, w, \epsilon)|h, w]$ and the optimal conditional continuous choices $c_d^*(h)$ (CCCs) for all d. Thus, I first need to back out the value h of η. To do so, I will first identify the conditional continuous choices $c_d^*(h)$ from the reduced forms R by exploiting monotonicity, Bayes' law and the relevant instrument w. Then, I will use monotonicity to identify η from the data by inverting the monotone $c_d^*(h)$: $h = (c_d^*)^{-1}(c_d)$. Once I identify the values h of the shock η, I can identify the conditional choice probabilities (CCPs) of selecting alternative d knowing $\eta = h, w$: $Pr(d|h, w)$. Then, I use these $Pr(d|h, w)$ as in Hotz and Miller (1993) to identify the difference in payoffs between the two alternatives. Finally, I discuss identification of the payoffs under additional structural assumptions in the next section.

1.3.2 Identification of Conditional Continuous Choices (CCCs)

Difficulty: observability problem

As in the literature on continuous choices (Matzkin, 2003; Bajari et al., 2007; Hong and Shum, 2010), I would like to exploit the monotonicity Assumption 3 to identify the optimal continuous choices. For any value of w, by monotonicity, we have that

$$Pr(\eta \leq \ h \ |d) = Pr(C_d \leq \ c_d^*(h) \ |d) \quad \forall d$$

$$\overset{\text{under Lemma 2}}{\Longleftrightarrow} \quad c_d^*(h) = F_{C_d|d}^{-1}(Pr(\eta \leq h|d)) \quad \forall d.$$

Thus, *if* we knew the distribution of η conditional on d, we could recover the optimal conditional continuous choices $c_d^*(h)$ by using the monotonicity of the conditional distribution of C_d knowing d

to invert it. However, here we only know the unconditional distribution of η (by Assumption 6).[6] The conditional distributions of $\eta|d$ are unobserved. They depend on an unobserved selection mechanism: $Pr(\eta \leq h|d) = Pr(d|\eta \leq h)Pr(\eta \leq h)/Pr(d)$. Because of this selection with simultaneous discrete and continuous choices, we cannot use usual inversion methods based on monotonicity for identification.

Another way to see the problem would be the following. Knowing that η is uniform and independent of observables (Assumptions 4 and 6), we have:

$$
\begin{aligned}
Pr(\eta \leq\ h) &= \overbrace{Pr(C_d \leq c_d^*(h))}^{\text{unobserved}} \quad \forall d \\
&= Pr(C_d \leq c_d^*(h); (D = 0 \cup D = 1)) \\
&= \underbrace{Pr(C_0 \leq c_0^*(h);\ D = 0)}_{\text{observed}} + \underbrace{Pr(C_0 \leq c_0^*(h);\ D = 1)}_{\text{unobserved}} \\
&= \underbrace{Pr(C_1 \leq c_1^*(h);\ D = 0)}_{\text{unobserved}} + \underbrace{Pr(C_1 \leq c_1^*(h);\ D = 1)}_{\text{observed}}.
\end{aligned}
$$

Imagine that we observed both c_0 and c_1 for every individual, independently of the discrete choice d, i.e., if $D = 0$ or $D = 1$ is selected, we observe both c_0 and c_1. Then, we observe the unconditional distribution of $c_d^*(h)$: $Pr(C_d \leq c_d^*(h))$. In this case, knowing that η is uniform, one could exploit monotonicity to recover $c_d^*(h)$ by inverting its unconditional distribution: $c_d^*(h) = F_{C_d}^{-1}(Pr(\eta \leq h))$. However, here again, we observe c_0 if $D = 0$ and c_1 if $D = 1$. Because of this selection, we cannot identify the optimal continuous choice policies.

Identification via the instrument:

Instead, to identify $c_d^*(h)$, I use the properties of the instrument (Assumption 2) to obtain structural

[6]Note that instead of normalizing the unconditional distribution, we could normalize the functional form of one of the conditional $\eta|d$ distributions. $\eta|_{D=0} \sim \mathcal{U}(0, 1)$ for example. However, the problem would be the same, as we still would not know the distribution of the other conditional shock $\eta|_{D=1}$.

restrictions. Using Bayes' law we have, $\forall h \in [0,1]$:

$$\begin{aligned} h = \quad & Pr(\eta \leq h) \\ = \quad & Pr(\eta \leq h|w) \\ = \quad & Pr(\eta \leq h \mid D=0,w)Pr(D=0|w) + Pr(\eta \leq h \mid D=1,w)Pr(D=1|w) \\ = \quad & Pr(c \leq c_0^*(h) \mid D=0,w)Pr(D=0|w) \\ & + Pr(c \leq c_1^*(h) \mid D=1,w)Pr(D=1|w) \\ = \quad & F_{C_0|D=0,w}(c_0^*(h))Pr(D=0|w) + F_{C_1|D=1,w}(c_1^*(h))Pr(D=1|w) \\ = \quad & F_{C_0|D=0,w}(c_0^*(h))p_{0|w} + F_{C_1|D=1,w}(c_1^*(h))p_{1|w}, \end{aligned}$$

where the first equality comes from the fact that $\eta \sim \mathcal{U}[0,1]$ by normalization. The second follows because $\eta \perp w$ by Assumption 4. The third equality comes from the law of total probability. The fourth equality comes from the monotonicity of $c_d^*(h)$. The fifth and sixth equalities are just changes in notation.

Thus:

$$h = F_{C_0|D=0,w}(c_0^*(h))p_{0|w} + F_{C_1|D=1,w}(c_1^*(h))p_{1|w} \qquad \forall h \in [0,1] \quad \forall w \in \{0,1\}. \tag{1.2}$$

Take equation 1.2 for both w, which yields the following system $\forall h$:

$$\begin{cases} h = F_{C_0|D=0,W=0}(c_0^*(h))p_{0|0} + F_{C_1|D=1,W=0}(c_1^*(h))p_{1|0} \\ h = F_{C_0|D=0,W=1}(c_0^*(h))p_{0|1} + F_{C_1|D=1,W=1}(c_1^*(h))p_{1|1} \end{cases}.$$

Thanks to the instrument, we have a system of two equations to identify two unknown increasing functions. The role of the instrument and Assumption 2 is now clearer. The instrument being exogenous to c_d is crucial here, otherwise, we would have two equations with four unknown functions: $c_0^*(h, W=0), c_0^*(h, W=1), c_1^*(h, W=0), c_1^*(h, W=1)$, which would not be identified. Similarly, without a relevant instrument (i.e., if $d \perp w$), the distributions conditional on w would be the same (i.e., $p_{0|0} = p_{0|1}$ and $F_{C_d|d,W=0}(c) = F_{C_d|d,W=1}(c)$), so the two equations would in fact contain exactly the same information.

Identification problem: Let the reduced form be described as:

$$R = \{\{p_{d|w}\}_{(d,w)\in\{0,1\}\times\{0,1\}},\ \{F_{C_d|d,w}(c_d)\}_{\forall c_d\in\mathcal{C}_d,(d,w)\in\{0,1\}\times\{0,1\}}\}$$

The conditional continuous choice (CCCs) policy functions are identified if and only if there exists a unique set of structural functions $\{c_d(h)\}_{\forall h\in[0,1],d\in\{0,1\}}$ strictly increasing with respect to h, which satisfies equation (1.2), and is compatible with R.

Theorem 1 (Identification) *For any reduced form drawn from the structural model, there exist unique conditional continuous choice (CCC) functions $c_d(h)$ (for $d = 0$ and $d = 1$) mapping $[0,1]$ into $\mathcal{C}_d$, which are strictly increasing and solve the system of equations (1.2):*

$$h = F_{C_0|D=0,w}(c_0(h))p_{0|w} + F_{C_1|D=1,w}(c_1(h))p_{1|w} \qquad \forall h\in[0,1] \quad \forall w\in\{0,1\}.$$

As a consequence, the optimal CCCs, $c_d^*(h)$ for $d = 0$ and $d = 1$, are point identified from the reduced form R as the unique increasing solutions to the identification problem.

Proof. The complete proof appears in Appendix 1.B.

Sketch of the proof:

Existence of the solution is trivial: since the reduced forms are drawn from the structural model, the true $c_d^*(h)$ will be the solution to our system of equations (1.2) by construction.

What is more difficult to prove is the *uniqueness* of the solution. First, we show that the mapping between the conditional continuous choices, denoted $\tilde{c_0}(c_1)$, is identified from the reduced forms. Once we have it, using system (1.2), it is trivial to show that the continuous policies are also identified.

Combining the two equations in the system of equation (1.2), we have that:

$$\begin{aligned}
& F_{C_0|D=0,W=0}(c_0^*(h))p_{0|0} + F_{C_1|D=1,W=0}(c_1^*(h))p_{1|0} = F_{C_0|D=0,W=1}(c_0^*(h))p_{0|1} + F_{C_1|D=1,W=1}(c_1^*(h))p_{1|1} \\
\iff\ & F_{C_0|D=0,W=1}(c_0^*(h))p_{0|1} - F_{C_0|D=0,W=0}(c_0^*(h))p_{0|0} = -\Big(F_{C_1|D=1,W=1}(c_1^*(h))p_{1|1} - F_{C_1|D=1,W=0}(c_1^*(h))p_{1|0}\Big) \\
\iff\ & \Delta F_{C_0}(c_0^*(h)) = -\Delta F_{C_1}(c_1^*(h)),
\end{aligned}$$

where $\Delta F_{C_d}(c)$ are directly observed from the data, and are C^1 as a sum of C^1 functions (Lemma 2). However, the problem is that h is unobserved. Now, even without observing h, if two conditional choices $\tilde{c_0}$ and $\tilde{c_1}$ correspond to the same unobserved h, we will have: $\Delta F_{C_0}(\tilde{c_0}) = \Delta F_{C_1}(\tilde{c_1})$. Thus, for the true mapping $\tilde{c_0}(c_1)$ between the two continuous conditional choices we will have

$$\forall c_1 \qquad \Delta F_{C_0}(\tilde{c_0}(c_1)) = -\Delta F_{C_1}(c_1). \tag{1.3}$$

The mapping is identified if and only if there exists a unique function $\tilde{c_0}(c_1)$ solution to equation (1.3). What are these $\Delta F_{C_d}(c)$ functions? They are observable C^1 functions (Lemma 2). They are related to the unknown conditional choice probabilities as follows (cf proof of Lemma 3):

$$\forall h \quad \Delta F_{C_d}(c_d^*(h)) = \int_0^h \Big(Pr(D=d|\eta=\tilde{h}, W=1) - Pr(D=d|\eta=\tilde{h}, W=0)\Big)d\tilde{h}.$$

Moreover, since $Pr(D=1|\eta,W) = 1 - Pr(D=0|\eta,W)$, we have by construction that, $\forall h$:

$$\Delta F_{C_0}(c_0^*(h)) = \int_0^h \Big(Pr(D=0|\eta=\tilde{h}, W=1) - Pr(D=0|\eta=\tilde{h}, W=0)\Big)d\tilde{h} = -\Delta F_{C_1}(c_1^*(h)). \tag{1.4}$$

Which is what we had by rewriting system (1.2). However, it is very important: it means that $\Delta F_{C_0}(c)$ and $-\Delta F_{C_1}(c)$ are transformations (through unknown $c_d^*(h)$) of the same underlying object, which is based on the difference in conditional choice probabilities $Pr(D=0|\eta=h, W=1) - Pr(D=0|\eta=h, W=0)$. Thus, by construction, $\Delta F_{C_0}(c)$ and $-\Delta F_{C_1}(c)$ will go *'through the same values, in the same order'*, just not at the same 'speed'. The shape of ΔF_{C_d} is directly determined by the difference in conditional choice probabilities, hence the reason why we make our identification Assumption 9 on these probabilities directly.

Now, take the easier case where $Pr(D=0|\eta=h, W=1) > Pr(D=0|\eta=h, W=0)$ for all h.[7] In other words, the identification Assumption 9 is satisfied with $K=0$. Equation (1.4) implies that $\Delta F_{C_0}(c_0)$ and $-\Delta F_{C_1}(c_1)$ will be *strictly increasing* from $Pr(D=0|\eta=0, W=1) - Pr(D=0|\eta=0, W=0)$ at the minimum values of c_0 and c_1 (corresponding to $c_0^*(0)$ and $c_1^*(0)$) to $\int_0^1 \Big(Pr(D=0|\eta=\tilde{h}, W=1) - Pr(D=0|\eta=\tilde{h}, W=0)\Big)d\tilde{h}$ at the maximum values of c_0 and c_1 (corresponding to $c_0^*(1)$ and $c_1^*(1)$). $\Delta F_{C_d}(c_d)$ are thus C^1 and strictly monotone: they are

[7] Obviously, the same reasoning applies in the reverse case where $Pr(D=0|\eta=h, W=1) < Pr(D=0|\eta=h, W=0)$ for all h.

invertible. In this case, the unique mapping between c_0 and c_1 is obtained by inverting equation (1.3):

$$\forall c_1 \qquad \tilde{c_0}(c_1) = \Delta F_{C_0}^{-1}\Big(- \Delta F_{C_1}(c_1)\Big).$$

The solution exists and is unique. Thus $\tilde{c_0}(c_1)$ is identified in this case.

Now, we can show that the continuous policies are still identified even if $\Delta F_{C_d}(c_d)$ are not strictly monotone but only *piecewise monotone*. This is the general case covered by our identification Assumption 9: if there exists a finite set of $K > 0$ (and $K < \infty$) values of h at which $Pr(D = 0|\eta = h, W = 1) = Pr(D = 0|\eta = h, W = 0)$, then by equation (1.4), we can show that $\Delta F_{C_d}(c_d)$ are *piecewise monotone*. Piecewise monotonicity is not a problem for identification here. We are *not* solving equation (1.3) point by point, in which case there could exist several solutions for some values of c_1. Instead, we are solving for the entire *monotone* policy functions $c_d^*(h)$ directly. Therefore, even if pointwise there might exist several solutions, there exists a unique *monotone* function on the whole support of c_1 that solves equation (1.3). In practice, we first identify these K points at which $d\Delta F_{C_d}(c_d)/dc_d = 0$ (Lemma 3). We know that these points are increasingly matched together by construction. Then, we split the support of c_0 and c_1 accordingly. On the subsegments, $\Delta F_{C_d}(c_d)$ are strictly monotone and C^1, thus invertible. Therefore, we can recover the mapping piece by piece.

The only case in which identification does not hold is when ΔF_{C_d} are *flat* on some segment. This corresponds to the case where our identification assumption 9 is violated, and the instrument is not relevant to a set of nonnull masses. In this case, we only have *partial identification* of the policy functions: they are point identified everywhere outside of the flat segment (on which there is an infinite number of possible mappings between c_1 and c_0).

Once we identify the mapping $\tilde{c}_0(c_1)$, we can recover the policies using any equation of the initial system (1.2), as:

$$\forall c_1 \qquad h(c_1) = F_{C_0|D=0,W=0}(\tilde{c}_0(c_1))p_{0|0} + F_{C_1|D=1,W=0}(c_1)p_{1|0}.$$

Thus we have a unique increasing solution $(h(c_1), \tilde{c}_0(c_1)) \forall c_1 \in \mathcal{C}_1$. Since everything is increasing, we can simply change the arguments to obtain the unique solution $(c_0^*(h), c_1^*(h))\ \forall h \in [0,1]$. ■

One of the main take-aways from this the proof is that, with this setup, by exploiting *knowledge* about the monotonicity of the optimal continuous policies and directly solving for the complete function, I identify the policies with assumptions that are considerably less restrictive than what is usually imposed in related studies. For example, *full rank* assumptions on the effect of the instrument on the selection in identification of IV quantile treatment effects (Newey and Powell, 2003; Chernozhukov and Hansen, 2005, 2006, 2008) are too strong in this framework. In fact, even my subcase where $K = 0$ was already less restrictive than full rank, for example. There is one notable exception of Vuong and Xu (2017), who are also solving for a complete function and not pointwise. However, they choose to use this method to relax strict monotonicity (and still impose some constraint on the conditional choice probabilities), while I use it to be as agnostic as possible about the conditional choice probabilities. My main identification requirement is to have a relevant instrument (Assumption 9), which seems fairly natural. Moreover, it is *testable* by observations of the $\Delta F_{C_d}(c_d)$ functions: as long as they are *not flat*, the policies are identified.

1.3.3 Identification of Conditional Choice Probabilities (CCPs)

Now that the conditional continuous choices (CCCs) are identified, I can directly identify the conditional choice probabilities (CCPs). Indeed, knowing the strictly monotone (and invertible) $(c_0^*(h), c_1^*(h))\forall h$, one can recover h from observing (d, c_d^{obs}). If $D = d$,

$$h = (c_d^*)^{-1}\big(c_d^{obs}\big).$$

From there, it is *as if* $\eta = h$ *were observed*. I observe (d, c_d, w, h) from the data. Thus, I can directly recover the conditional choice probabilities:

$$\forall (d, w, h) \in \{0,1\} \times \{0,1\} \times [0,1]: \qquad Pr(D = d | \eta = h, W = w).$$

Thus, the CCPs are identified once h is recovered from inverting the CCCs.

Inclusion of unobserved types in the model

The fact that η acts as an observed covariate once the CCCs are identified is crucial. Thanks to this, one can apply standard methods from the dynamic discrete choice literature where η would be among the *observed covariates*. This means that once η is identified, one could include *unobserved state variables/types* in the framework as in Arcidiacono and Miller (2011). The non-parametric identification is given by Kasahara and Shimotsu (2009) or Hu and Shum (2012).

1.3.4 Identification of the payoffs

Now that the optimal policy choices are identified, we can proceed to identify the structural model, i.e., the payoff functions $v_d(c_d^*(h), h)$ and $m_d(w, h)$. First, I focus on the identification of the differences in payoff between the discrete alternatives.

Identification of the differences in payoffs:

The conditional choice probabilities are identified in the data. We can use them with our structural assumptions to identify *difference in payoffs* in the model. We know that the CCPs are related to the structure of the model as follows:

$$\begin{aligned} Pr(D=0|\eta=h,w) &= Pr\Big(\epsilon_0-\epsilon_1 > (\underset{c}{max}\ v_1(c,h)+m_1(w,h)) - (\underset{c}{max}\ v_0(c,h)+m_0(w,h))\Big|\ h,w\Big) \\ &= Pr\Big(\epsilon_0-\epsilon_1 > v_1^*(h)+m_1(w,h) - \big(v_0^*(h)+m_0(w,h)\big)\ \Big|\ h,w\Big), \end{aligned}$$

where $v_d^*(h) \equiv v_d(c_d^*(h),h) = \underset{c}{max}\ v_1(c,h)$.

If the distribution of $\epsilon_0-\epsilon_1$ is known (and invertible), given that we know the CCPs, the difference in payoffs will also be identified. As is standard in the discrete choice literature, identification depends on the distribution of the difference in ϵ here.

For example, let us assume that ϵ follows a *Gumbel/extreme-value type-I distribution*(with location 0 and scale 1), as is commonly used in the discrete choice literature (McFadden, 1980; Hotz and Miller, 1993). In this case, we are in the logistic regression scenario and we have:

$$Pr(D=0|\eta=h,w) = \frac{1}{1+exp\Big(v_1^*(h)+m_1(w,h)-\big(v_0^*(h)+m_0(w,h)\big)\Big)}.$$

Thus we identify the *difference in payoffs* as:

$$v_1^*(h) + m_1(w,h) - \big(v_0^*(h) + m_0(w,h)\big) = log\left(\frac{1}{Pr(D=0|\eta=h,w)} - 1\right).$$

Moreover, since $v_d^*(h)$ are independent of w by Assumption 2, we can also identify the *difference in the effect of the instrument*:

$$\begin{aligned}&m_1(w=1,h) - m_0(w=1,h) - (m_1(w=0,h) - m_0(w=0,h))\\ &= log\left(\frac{1}{Pr(D=0|\eta=h,w=1)} - 1\right) - log\left(\frac{1}{Pr(D=0|\eta=h,w=0)} - 1\right).\end{aligned}$$

The differences in payoffs are also non-parametrically identified for other distributions of ϵ. Applications often use generalized-extreme value distributions as they yield easily tractable closed-form solutions (Arcidiacono and Miller, 2011), but other distributions are possible.

Identification of the payoffs:

To non-parametrically identify the *payoffs* directly using the CCPs and CCCs, one needs to add some structure to the problem. In other words, we need additional behavioural conditions to know how the agents behave. For example, by considering the framework applied to dynamic problems, I can use the identification power of the first-order conditions/Euler equation to non-parametrically directly identify the *payoffs* using the identified CCPs and CCCs. This is what I do in the next section by extending the framework to a dynamic setup.

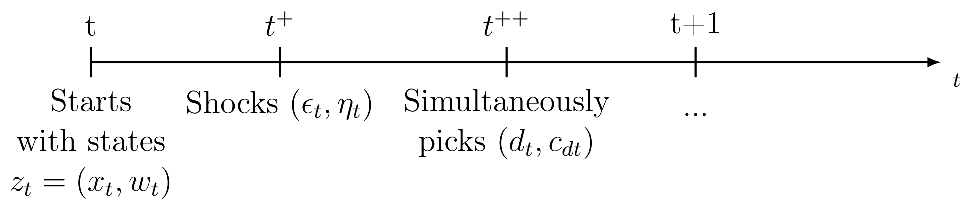
1.4 Extension to Dynamic models

The general framework that I developed embeds dynamic models: v_{dt} must simply be understood as current conditional value functions, embedding expectations about the future. Here, I show how general (non-stationary) dynamic models of agents enter the setup and are non-parametrically identified (in the spirit of Blundell et al., 1997). The model is very general and nests many life-cycle empirical applications of interest (e.g., Blundell et al., 2016a; Iskhakov et al., 2017).

1.4.1 Dynamic Life-Cycle Framework of Labor and Consumption

In this section, I describe how a general dynamic model of labor and consumption choices enters the general framework described in section 1.2.

Each period t until T, the timing of the problem is as follows:



Current period utility:

The *current period conditional utility* for action (d, c_d) at time t is given by:

$$\mathcal{U}_{dt}(c_{dt}, x_t, w_t, \eta_t, \epsilon_t). \tag{1.5}$$

In this example, c_t is consumption, and c_{dt} are conditional consumptions, with $c_t = c_{0t}(1-d_t)+c_{1t}d_t$. d_t is the work decision (Blundell et al., 1997, 2016a). x_t represents all the covariates. These include covariates impacting current utility such as age, education and other demographics. For notational convenience, x_t also include variables such as asset or income that do not necessarily directly impact preferences but still have an impact on consumption choice (and labor choice), notably through their transitions. w_t is again the instrumental variable that must fulfil some conditions I describe below.

I impose some conditions on *current utility* which are necessary (not sufficient) for the dynamic setup described here to fit into the structure described in section 1.2.

Assumption $D1$ **(Additive Separability)** *The shock ϵ_t enters the payoff additively such that:*

$$\mathcal{U}_{dt}(c_{dt}, x_t, w_t, \eta_t, \epsilon_t) = \tilde{u}_{dt}(c_{dt}, x_t, w_t, \eta_t) + \epsilon_{dt}.$$

Assumption $D2$ **(Instrument)** *$w_t \in \mathcal{W} = \{0, 1\}$ is an instrumental variable such that*

$$\tilde{u}_{dt}(c_{dt}, x_t, w_t, \eta_t) = u_{dt}(c_{dt}, x_t, \eta_t) + m_{dt}(x_t, w_t, \eta_t).$$

Assumption $D3$ **(Monotonicity)** *The conditional current utility functions are twice continuously differentiable such that*

$$\frac{\partial^2 u_{dt}(c_{dt}, x_t, \eta_t)}{\partial c_{dt} \partial \eta_t} > 0 \qquad \forall d_t, c_{dt}, x_t, \eta_t$$

Transitions:

In a dynamic context, the individual chooses (d_t, c_{dt}) to maximize not only her current utility but also to maximize her *expected discounted sum of future payoffs.* She discounts the future period utilities at a rate β. In this context, the agent form *rational expectations* about the transition probabilities. These transitions from $(x_t, w_t, \epsilon_t, \eta_t)$ and the current choices (c_t, d_t) to $(x_{t+1}, w_{t+1}, \epsilon_{t+1}, \eta_{t+1})$ matter for the choices. In particular, how the current choices impact these transitions is especially important for the optimal choice decision. The impacts of the choices on the transitions are often expressed through a *budget constraint* like

$$a_{t+1} = (1 + r_t)a_t - c_t + y_t d_t.$$

For now I stay more general and simply assume the existence of general transitions of states and errors which depend on the choices:

$$f_t(x_{t+1}, w_{t+1}, \epsilon_{t+1}, \eta_{t+1} | c_t, d_t, x_t, w_t, \epsilon_t, \eta_t).$$

I need to make additional assumptions on these transitions for the setup to be identified (and to enter the general framework).

Assumption 10 (Conditional Independence) *For all $x_t \in \mathcal{X}$, $w_t \in \mathcal{W}$, $\epsilon_t \in \mathcal{E}$, $\eta_t \in \mathcal{H}$, we have:*

$$f_t(x_{t+1}, w_{t+1}, \epsilon_{t+1}, \eta_{t+1} | c_t, d_t, x_t, w_t, \epsilon_t, \eta_t) = f_t(x_{t+1}, w_{t+1} | c_t, d_t, x_t, w_t) \; f_\epsilon(\epsilon_{t+1}) \; f_\eta(\eta_{t+1}).$$

Assumption 11 (Instrument Transition Exclusion) *For all $x_t \in \mathcal{X}$, $w_t \in \mathcal{W}$, the current instrument is excluded from the transitions, i.e.,*

$$f_t(x_{t+1}, w_{t+1}|c_t, d_t, w_t, x_t) = f_t(x_{t+1}, w_{t+1}|c_t, d_t, x_t).$$

Solution:

Knowing these transition probabilities, the individual chooses (d_t, c_{dt}) to sequentially maximize her expected discounted sum of payoffs. Let us define $V_t(z_t) = V_t(x_t, w_t)$ as the (ex ante) value function of this discounted sum of future payoffs at the beginning of t, just before the shocks (ϵ_t, η_t) are revealed and conditional on behaving according to the optimal decision rule:

$$V_t(z_t) \equiv \mathbb{E}\Big[\sum_{\tau=t}^{T} \beta^{\tau-t} \underset{d, c_{d\tau}}{max}[\ u_{d\tau}(c_{d\tau}, x_\tau, \eta_\tau) + m_d(x_\tau, w_\tau, \eta_\tau) + \epsilon_{d\tau}]\Big].$$

Given the state variable z_t and choice (d, c_{dt}) in period t, the expected value function in period $t+1$ is

$$\mathbb{E}_{z_{t+1}}[V_{t+1}(z_{t+1})|z_t, c_t, d_t] = \int_{z_{t+1}} V_{t+1}(z_{t+1}) f_t(z_{t+1}|z_t, c_t, d_t) dz_{t+1}.$$

By the conditional independence Assumption 10 and instrument exclusion from the transition (Assumption 11), we can remove w_t from the conditioning variables, which yields:

$$\mathbb{E}_{z_{t+1}}[V_{t+1}(z_{t+1})|z_t, c_t, d_t] = \mathbb{E}_{z_{t+1}}[V_{t+1}(z_{t+1})|x_t, c_t, d_t] = \int_{z_{t+1}} V_{t+1}(z_{t+1}) f_t(z_{t+1}|x_t, c_t, d_t) dz_{t+1}.$$

The ex ante value function can be written recursively:

$$V_t(z_t) = \mathbb{E}_{\epsilon,\eta}\left[\underset{d_t, c_{dt}}{max}\Big[\ u_{dt}(c_{dt}, x_t, \eta_t) + m_{dt}(x_t, w_t, \eta_t) + \epsilon_{dt} + \beta \mathbb{E}_{z_{t+1}}[V_{t+1}(z_{t+1})|x_t, c_{dt}, d_t]\Big]\ \right].$$

Thus, in each period, after observing (ϵ_t, η_t), the individual chooses d_t and c_{dt} to maximize her expected payoff:

$$\underset{d_t, c_{dt}}{max}\ u_{dt}(c_{dt}, x_t, \eta_t) + \beta \mathbb{E}_{z_{t+1}}[V_{t+1}(z_{t+1})|x_t, c_{dt}, d_t]\ + m_{dt}(x_t, w_t, \eta_t) + \epsilon_{dt}.$$

Denote the conditional value functions v_{dt} as:

$$v_{dt}(c_{dt}, x_t, \eta_t) \; \equiv \; u_{dt}(c_{dt}, x_t, \eta_t) + \beta \mathbb{E}_{z_{t+1}}[V_{t+1}(z_{t+1})|_{x_t, c_{dt}, d_t}]. \tag{1.6}$$

So that we return to our general setup. Indeed, the dynamic model can be interpreted as a static model, where in every period the agent selects d_t and c_{dt} to solve:

$$\underset{d_t, c_{dt}}{max} \quad v_{dt}(c_{dt}, x_t, \eta_t) \; + m_{dt}(x_t, w_t, \eta_t) + \epsilon_{dt}.$$

Lemma 4 (Dynamic Framework) *Under Assumptions D1, D2, D3, 10 and 11, Assumptions 1, 2 and 3 are satisfied for the conditional value functions defined in equation (1.6) in the dynamic setup.*

The other Assumptions 4, 5, 6, 7, 8 and 9 as well as Normalization 1 are imposed contemporaneously (with index t) and unconditionally on X_t (for simplicity).

If Assumption $D1$ holds for the current utility function, Assumption 1 will hold for the conditional value functions by construction in equation (1.6). Assumptions $D2$ and $D3$ on the current utility do not translate directly into Assumptions 2 and 3 for the conditional value function. One needs additional assumptions about the transitions, i.e., Assumptions 10 and 11.

Conditional independence assumptions are standard for the identification and empirical tractability of dynamic discrete choice models (Rust, 1987; Blevins, 2014). Here, Assumption 10 implies that the transitions of the state variables are independent of the shocks (ϵ_t, η_t). Similarly, the shock transitions are independent of the variables here. There is *no time dependence* on the shocks, which are thus iid every period. Note that one can include some unobserved heterogeneous types correlated over time in the covariates following Arcidiacono and Miller (2011). This allows for some unobserved auto-correlation in the unobservables and attenuates the strength of the conditional independence.

Crucially, here, in addition to the standard conditional independence 10, Assumption 11 also implies that conditional on (d_t, c_t, x_t), the *transitions* are *independent* of the current instrument value w_t. In particular, the instrument is excluded from its own transition to future values, conditional on (d_t, c_t, x_t), i.e.,

$$w_{t+1} \perp w_t \mid c_t, d_t, x_t \quad \forall c_t, d_t, x_t$$
$$\text{or equivalently } f_w(w_{t+1}|c_t, d_t, x_t, w_t) = f_w(w_{t+1}|c_t, d_t, x_t).$$

It implies that instruments that are time independent $w_t = w$ for all t cannot be included. Assumption *D*2 combined with Assumption 11 will satisfy Assumption 2 on the conditional value as stated in Lemma 4 and as shown in the computation above. However, if the exclusion of the instrument from the transition (Assumption 11) does not hold, then w_t affects the expected future value function $\mathbb{E}_{z_{t+1}}[V_{t+1}(z_{t+1}|x_t, w_t, c_{dt}, d_t]$ and enters the conditional value functions v_{dt} in equation (1.6), which are rewritten as $v_{dt}(c_{dt}, x_t, w_t, \eta_t)$ for all d in this case. In this case, it is obvious that the original exclusion of the instrument from v_d in Assumption 2 is violated. Thus, the dynamic setup does fit into the general framework of section 1.2 without Assumption 11.

Similarly, Assumption *D*3 is just a necessary condition for Assumption 3 to hold. I also require the expectations about the future to be independent of current (η_t, ϵ_t). In this case, the monotonicity Assumption 3 in the general framework is also satisfied. Indeed, if the future is independent of the current η_t (Assumption 10), then we can write:

$$\frac{\partial v_{dt}(c_{dt}, x_t, \eta_t)}{\partial c_{dt} \partial \eta_t} = \frac{\partial u_{dt}(c_{dt}, x_t, \eta_t)}{\partial c_{dt} \partial \eta_t} + \underbrace{\frac{\partial \mathbb{E}_{z_{t+1}}[V_{t+1}(z_{t+1})|x_t, c_{dt}, d_t]}{\partial c_{dt} \partial \eta_t}}_{=0}.$$

Therefore, if the conditional independence and monotonicity of the current utility function hold (Assumptions 10 and *D*3), then the monotonicity of the conditional value functions v_{dt} (Assumption 3) also holds.

Instrument example:

The question that remains is, what could be a good instrument satisfying this restrictive conditional independence and exclusion from the transition in practice? In general, a good instrument would be to allow for *switching cost* and to use $w_t = d_{t-1}$ in this setup. Indeed, in this case, the exclusion assumption 11 is easily satisfied for the instrument: w_{t+1} is d_t. Therefore, conditional on the current d_t choice, w_{t+1} is directly known. $w_t = d_{t-1}$ does not provide any additional information, so it can be dropped from the conditioning variables in the transition. Moreover, it is unlikely that w_t provides any information about the other future covariates x_{t+1} after conditioning on the current d_t. Similarly, conditional on x_t, which could include for example, the experience of the individual, it is unlikely that d_{t-1} has an impact on u_{dt}. The exclusion restriction *D*2 is satisfied. Finally, we just need the instrument to be relevant (Assumption 9). This would be the case if one had some utility switching cost from entering or exiting the workforce for example.[8] In this case, we would have: $m_{0t}(x_t, w_t = 0, \eta_t) - m_{1t}(x_t, w_t = 0, \eta_t) \neq m_{0t}(x_t, w_t = 1, \eta_t) - m_{1t}(x_t, w_t = 1, \eta_t)$. And the instrument would be relevant.

Relaxing time independence of η:

One can loosen Assumption 10 and allow for first-order time dependence in η_t in this setup. In other words, I can have $f_\eta(\eta_t|\eta_{t-1})$. In fact, as I identify η_t separately for all t, I can identify these transitions, which are particularly interesting in some applications (e.g., if η represents some unobserved ability or productivity). The only problem is that it is more difficult to find a good instrument in practice in this case. Indeed, in the presence of auto-correlation in η_t, $w_t = d_{t-1}$ is no longer a good instrument, as it violates its independence from η_t in the initial period (Assumption 4). Indeed, in the initial period of the data, η_{-1} is not observed and is correlated with η_0. However, in this case, d_{-1} was a choice taken based on η_{-1} and thus correlated with η_{-1}. Therefore, in the first period, $w_0 = d_{-1}$ is correlated with η_{-1} and thus with η_0. The instrument w_0 is not independent of η_0, which violates Assumption 4. If we were able to condition on η_{t-1}, we could identify η_t: conditional on η_{t-1}, $w_t = d_{t-1} \perp \eta_t$. However, there is no way to recover η_{-1} which is outside the sample. Thus, I cannot allow for transition in η_t with $w_t = d_{t-1}$ as an instrument. Therefore, the best way to account for unobserved auto-correlation with $w_t = d_{t-1}$ as an instrument would be to include unobserved types à la Arcidiacono and Miller (2011) in the model and still impose conditional independence with an iid η_t.

If there exists another instrument satisfying Assumptions 4, *D*2 and 11, then one can allow for auto-correlated η_t. In fact, even if such an instrument is available only in one period t_0 (e.g., a

[8]Alternatively, one could have some auto-correlation in d_t. Recall that $m_{dt}()$ can be interpreted as some observable part of the ϵ_t shocks. Thus, the relevance assumption with $w_t = d_{t-1}$ could also be interpreted as the existence of auto-correlation in a general $\tilde{\epsilon}_t$ term, with

$$\tilde{\epsilon}_{dt} = m_{dt}(x_t, w_t, \eta_t) + \epsilon_{dt}.$$

Thus, the assumption about no correlation in ϵ_t is less restrictive than it seems.

unique unexpected event), then one can still allow for auto-correlation in η_t. Indeed, one can use the instrument in the period to identify the η_{t_0}. For all the following periods, $w_t = d_{t-1}$ can be used as a proper instrument if I include η_{t-1} in the covariates list.

1.4.2 Identification of the dynamic model

First, I show how the CCCs and CCPs are identified in this dynamic model. Then, I show how the payoffs are also non-parametrically identified under additional assumptions.

Optimal choices: CCCs and CCPs

Under Lemma 4, the dynamic framework described in Section 1.4 fits into the general framework described in Section 1.2. Therefore the CCCs and CCPs are identified period by period following exactly the same proof I developed in the previous section. In other words, from data on (D_t, C_t, X_t, W_t, t), I recover reduced forms

$$R = \Big\{ \{Pr(D_t = d | X_t = x, W_t = w, t)\}_{(d,w,x,t)\in\mathcal{D}\times\mathcal{W}\times\mathcal{X}_t\times\{0,\dots,T\}},$$
$$\{F_{C_d|D_t=d,X_t=x,W_t=w,t}(c_d)\}_{\forall c_d\in\mathcal{C}_{dt},\ (d,w,x,t)\in\mathcal{D}\times\mathcal{W}\times\mathcal{X}_t\times\{0,\dots,T\}} \Big\}.$$

From these reduced forms, following Section 1.3, I identify the CCCs and CCPs

$$c^*_{dt}(x, \eta = h) \text{ and } Pr(d|\eta = h, x, w, t) \qquad \forall d \in \{0,1\}, w \in \{0,1\}, h \in [0,1], x \in \mathcal{X}_t, t \in \{0,...,T\}.$$

Special case: identification of the choices with terminal/absorbing actions

Imagine $d_t = 1$ is a *terminal action* or an *absorbing state*. For example $d_t = 1$ if the individual retires, $d_t = 0$ if she stays active. Assuming that an individual cannot go back to the working life, the retirement choice is absorbing (Iskhakov et al., 2017). In this case, assuming all the other modeling assumptions still hold, identification is more direct and simpler. I still use $w_t = d_{t-1}$ as the instrument, so the Assumption 11 on the transitions is still verified. Now, conditional on $w_t = 1$, an individual only has the choice to stay retired, i.e., $d_t = 1$. Thus, focus on previously retired individuals ($W_t = 1$), we have:

$$\begin{aligned}
h &= Pr(\eta_t \leq h | X_t, W_t = 1, t) \\
&= Pr(\eta_t \leq h | D_t = 1, X_t, W_t = 1, t) \\
&= Pr(c \leq c^*_{1t}(h, X_t) | D_t = 1, X_t, W_t = 1, t) \\
&= F_{C_1|D_t=1,X_t,W_t=1,t}(c^*_{1t}(h, X_t)) \qquad \forall X_t \in \mathcal{X}_t, h \in [0,1].
\end{aligned}$$

Since $F_{C_1|D_t=1,X_t,W_t=1,t}(c)$ are invertible (Lemma 2), we recover the continuous choices conditional on being retired as:

$$c^*_{1t}(h, X_t) = F^{-1}_{C_1|D_t=1,X_t,W_t=1,t}(h) \quad \forall X_t \in \mathcal{X}_t, h \in [0,1].$$

It remains to identify the other conditional continuous policy, and to do that one simply needs to take the equation (1.2) at $W_t = 0$, i.e., for individuals who did not select the absorbing state yet. It yields

$$\begin{aligned}
h = \quad & F_{C_0|D_t=0,X_t,W_t=0,t}(c^*_{0t}(h, X_t)) Pr(D_t = 0 | X_t, W_t = 0, t) \\
& + \quad F_{C_1|D_t=1,X_t,W_t=0,t}(c^*_{1t}(h, X_t)) Pr(D_t = 1 | X_t, W_t = 0, t) \\
\iff \quad & F_{C_0|D_t=0,X_t,W_t=0,t}(c^*_{0t}(h, X_t)) = \frac{h - F_{C_1|D_t=1,X_t,W_t=0,t}(c^*_{1t}(h, X_t)) Pr(D_t = 1 | X_t, W_t = 0, t)}{Pr(D_t = 0 | X_t, W_t = 0, t)} \\
\iff \quad & c^*_{0t}(h, X_t) = F^{-1}_{C_0|D_t=0,X_t,W_t=0,t}\left(\frac{h - F_{C_1|D_t=1,X_t,W_t=0,t}(c^*_{1t}(h, X_t)) Pr(D_t = 1 | X_t, W_t = 0, t)}{Pr(D_t = 0 | X_t, W_t = 0, t)}\right).
\end{aligned}$$

Since everything on the right hand side of the equation is known (as I identified c^*_{1t} previously), the other conditional policies $c^*_{0t}(h, X_t)$ are also identified $\forall X_t \in \mathcal{X}_t, h \in [0,1]$. Once the CCCs are identified, we proceed as usual to identify the CCPs.

Transitions

The transitions $f(x_{t+1}|c_t, d_t, x_t)$ are identified directly from the data by observing the conditional transitions of the variables between consecutive periods t and $t+1$. The transition of the instrument is known by construction if $w_{t+1} = d_t$. In other cases, it can also be recovered from the data (and one can test if it is indeed independent from w_t).

As standard in the dynamic model literature, I assume agents are *rational* so that the observed transitions are the same as the one expected by the agents. This way, the transitions recovered from the data can be used to build agents expectations at each time t, and help recover the primitives.

Payoff function

Once the CCCs, CCPs and transitions are identified, I can build upon existing literature to identiy the payoffs (Hotz and Miller, 1993; Blundell et al., 1997; Magnac and Thesmar, 2002; Escanciano et al., 2015). I need to introduce some additional structure to the dynamic model for non-parametric identification: I introduce additional structure on the covariates transition and current utility function.

Budget Constraint: let us introduce additional structure on the transitions via a *budget constraint*:

$$a_{t+1} = (1 + r_t)a_t - c_t + y_t d_t, \tag{1.7}$$

where a_t is the individual asset holdings, y_t is her income and r_t is the interest rate. The asset plays a different role than the other covariates. Indeed, its transition to a_{t+1} is directly impacted by the choice c_t through the budget constraint (1.7). Denote more generally all the covariates $x_t = (\tilde{x}_t, a_t)$ to emphasize the role of the asset.[9]

Assumption 12 (Asset exclusion) *The asset is excluded from the current period utility, i.e.,*

$$u_{dt}(c_{dt}, x_t, \eta_t) = u_{dt}(c_{dt}, \tilde{x}_t, a_t, \eta_t) = u_{dt}(c_{dt}, \tilde{x}_t, \eta_t).$$

Or equivalently, the exclusion Assumption 12 can be stated as:

$$\frac{\partial u_{dt}(c_{dt}, \tilde{x}_t, a_t, \eta_t)}{\partial a_t} = 0.$$

[9]Note that y_t and r_t are included in $\tilde{x}_t$. Even though, in most applications they will also be excluded from the current period utility. For notational simplicity and generality, I let them into the general $\tilde{x}_t$ term which enters in the current utility and represents all the covariates other than asset, i.e., all the covariates whose transitions are not impacted by c_t (Assumption 13).

Assumption 13 (General Covariates Transitions) *For all $\tilde{x}_t \in \tilde{\mathcal{X}}, \forall d_t \in \mathcal{D}, \forall c_t \in \mathcal{C}$, c_t does not impact the $\tilde{x}_t$ and w_t transitions, i.e.,*

$$f_t(\tilde{x}_{t+1}, w_{t+1}|c_t, d_t, \tilde{x}_t) = f_t(\tilde{x}_{t+1}, w_{t+1}|d_t, \tilde{x}_t).$$

I also need some additional structure on the current period utility:

Assumption 14 (Stationary utility) *The current period utility is independent from time*

$$u_{dt}(c_{dt}, \tilde{x}_t, \eta_t) = u_d(c_{dt}, \tilde{x}_t, \eta_t) \ \forall t.$$

Assumption 15 (Monotonicity of c on the current utility) *The current period utility is monotone increasing with respect to c*

$$\frac{\partial u_{dt}(c_{dt}, \tilde{x}_t, \eta_t)}{\partial c} > 0 \qquad \forall d_t, c_{dt}, \tilde{x}_t, \eta_t.$$

Marginal utilities identification:

Lemma 5 (Escanciano et al. (2015)) *Following Escanciano et al. (2015), under Assumptions D1-D3 and 4-15, the conditional marginal utilities at optimal continuous choices*

$$\frac{\partial}{\partial c_{dt}} u_d(c_{dt}, \tilde{x}_t, \eta_t)|_{c_{dt}=c^*_{dt}(\eta_t, x_t)}$$

are identified up to a scale by the Euler Equation for all d, x_t, η_t.

Proof.

Let us define

$$u'^*_d(x_t, \eta_t) = \frac{\partial}{\partial c_{dt}} u_d(c_{dt}, \tilde{x}_t, \eta_t)|_{c_{dt}=c^*_{dt}(\eta_t, x_t)},$$

i.e., the conditional marginal utilities at the optimal CCCs. Notice that these functions depend on $x_t = (\tilde{x}_t, a_t)$. In other words, the optimal conditional marginal utilities depend on the asset, through the optimal CCCs. Then, the *Euler equations* for all d can be rewritten as:

$$u_d'^*(x_t, \eta_t) = \ \beta(1+r_t)\mathbb{E}_t\Big[u_{d_{t+1}}'^*(x_{t+1}, \eta_{t+1}) \ \Big| x_t, c_{dt} = c_{dt}^*(\eta_t, x_t), d_t = d\Big]. \tag{1.8}$$

We have a system of two equations with two unknown functions $u_0'^*$ and $u_1'^*$. Hence the importance of Assumption 14, otherwise we would have a different unknown function on each side of the equation. Now, under Assumption *D*3 and 15, I have

$$\partial u_d'^*(x_t, \eta_t)/\partial \eta_t > 0 \quad \forall d, x_t, \eta_t.$$

In this case, Escanciano et al. (2015) show that these functions are non-parametrically globally point identified by the system (1.8). ■

Conditional values:

Once the marginal utilities are identified through Lemma 5, I follow Blundell et al. (1997) to identify the conditional values.

Lemma 6 (Blundell et al. (1997)) *Under Assumptions D1-D3 and 4-15, the conditional value functions at optimal choices* $v_{dt}(c_{dt}^*(x_t, \eta_t), x_t, \eta_t)$ *are identified up to an unknown constant of integration* K *independent from the asset for all* d, x_t, η_t*. i.e.,*

$$v_{dt}(c_{dt}^*(\eta_t, x_t), x_t, \eta_t) = G_{dt}(\tilde{x}_t, a_t, \eta_t) + K_{dt}(\tilde{x}_t, \eta_t),$$

where G *and* K *are defined in the proof.*

Proof. Recall that $x_t = (\tilde{x}_t, a_t)$. We have the first order conditions, holding at optimal CCCs:

$$\forall d: \ \forall a_t \qquad \frac{\partial}{\partial a_t} v_{dt}(c_{dt}, \tilde{x}_t, a_t, \eta_t) = (1+r_t)\frac{\partial}{\partial c_{dt}} u_d(c_{dt}, \tilde{x}_t, \eta_t) \ |_{c_{dt}=c_{dt}^*(\eta_t, \tilde{x}_t, a_t)}. \tag{1.9}$$

Denote v_d^* is the conditional value taken at the optimal continuous choice, and similarly define $u_d'^*$ as before. We can rewrite the FOC as:

$$\forall d: \ \forall a_t \qquad \frac{\partial}{\partial a_t} v_{dt}^*(\tilde{x}_t, a_t, \eta_t) = (1+r_t) \ u_d'^*(\tilde{x}_t, a_t, \eta_t). \tag{1.10}$$

Crucially, following Assumption 12, the asset is excluded from the current period utilities and marginal utilities. The identification strategy relies on this exclusion. Indeed, from this FOC (1.9) at the optimal CCCs, I can integrate with respect to the continuous asset and obtain

$$\forall d: \ \forall a_t \qquad v^*_{dt}(\tilde{x}_t, a_t, \eta_t) = \int_0^{a_t} (1 + r_t)\ u'^*_d(\tilde{x}_t, a_t, \eta_t)\ da,$$

where the lower bound 0 is taken arbitrarily. Since u'^*_d are identified, we can identify the optimal conditional value functions non-parametrically as:

$$v^*_{dt}(\tilde{x}_t, a_t, \eta_t) \equiv G_{dt}(\tilde{x}_t, a_t, \eta_t) + K_{dt}(\tilde{x}_t, \eta_t).$$

Where the only remaining unknowns are $K_{dt}(\tilde{x}_t, \eta_t)$ which are unknown constant of integration, independent from a_t, and which depends on the arbitrary lower bound of integration. ∎

Additive term $m_{dt}(x_t, w_t, \eta_t)$*:*

It remains to identify the additive terms. I identify the *differences* in total conditional values by relating them to the CCPs using an Hotz and Miller (1993)'s inversion, as in section 1.3.4. In other words,

$$\begin{aligned}\Delta v^*_t(\tilde{x}_t, a_t, \eta_t) + \Delta m_t(x_t, w_t, \eta_t) &= v^*_{1t}(\tilde{x}_t, a_t, \eta_t) - v^*_{0t}(\tilde{x}_t, a_t, \eta_t) + m_{1t}(x_t, w_t, \eta_t) - m_0(x_t, w_t, \eta_t) \\ &= G_{1t}(\tilde{x}_t, a_t, \eta_t) - G_{0t}(\tilde{x}_t, a_t, \eta_t) + K_{1t}(\tilde{x}_t, \eta_t) - K_{0t}(\tilde{x}_t, \eta_t) + m_{1t}(x_t, w_t, \eta_t) - m_0(x_t, w_t, \eta_t)\end{aligned}$$

are identified through the CCPs $Pr(d|\eta_t, x_t, w_t, t)$ for all d, x_t, η_t, w_t, t. Note that I cannot identify K_{dt} separately from m_{dt}. A natural *normalization* is to impose

$$K_{dt}(\tilde{x}_t, \eta_t) = 0 \qquad \forall d, \tilde{x}_t, \eta_t, t$$

Such that the only remaining *additive terms* are m_{dt}. Under this normalization, given that $v^*_{dt}(\tilde{x}_t, a_t, \eta_t) = G_{dt}(\tilde{x}_t, a_t, \eta_t)$ have been previously identified, it means that $\Delta m_t(x_t, w_t, \eta_t)$ are identified for all d, x_t, η_t, w_t, t. Obviously, since these are identified through the discrete choice probabilities, one can only non-parametrically identify the differences in m_d, not their separate values at each d.

About the role of each assumptions:

Assumption 12 excludes the asset from the utility. Having an excluded asset is essential to recover the conditional values once the marginal current utilities are identified.

Assumption 13 implies that the only covariate whose transition is impacted by the choice c_t is the asset, through the budget constraint (1.7). This assumption is made to pin down a simpler Euler equation than with general transitions with several variables impacted by c_t.

To identify the marginal utility non-parametrically from the Euler Equation, one needs to impose some structure on the effect of time in the utility function. I impose that the current period utility is time independent through Assumption 14. Note that, in general, even if the current period utility is time independent, the conditional value functions are still time-dependent, because of a finite horizon, or because of time-dependent transitions. Also note that this assumption is only necessary for non-parametric identification. In parametric models, I can identify time-dependent utilities.

Assumption 15 is a slightly stronger monotonicity condition than the ones I imposed before. In most empirical applications it will be satisfied though.

1.5 Estimation

I build a two-step estimation process in the spirit of Hotz and Miller (1993); Arcidiacono and Miller (2011) in the discrete choice literature. In the first step, I estimate the conditional continuous choices (CCCs) and the conditional choice probabilities (CCPs) based on reduced forms directly estimated from the data. This step is data-driven and is independent from the model specification. In a second step, I use the estimated optimal policies to estimate the structural parameters. Therefore, my estimation method is an analogous to that of Hotz and Miller (1993) and Hotz et al. (1994) but extended to discrete-continuous choices. Its main desirable feature concerns computational gains. By estimating the optimal choices only once, the computational burden of the estimation is significantly reduced. Indeed, one does not need to solve for the value function or the likelihood for each new set of selected parameters. This allows us to estimate models that were previously computationally intractable. It does so at minimal efficiency costs (compared to simulated method of moments, for example). I expose the estimation method in this section, and I compare my estimator's performance with several alternatives in terms of speed and efficiency in the next section 1.6.

1.5.1 1st stage: conditional choices

Reduced forms:

I observe data about (D_t, C_t, X_t, W_t, t). Where $c_t = (1 - d_t)c_{0t} + d_t c_{1t}$. From the data, I estimate the reduced forms:

$$R = \Big\{ \{Pr(D_t = d | X_t = x, W_t = w, t)\}_{(d,w,x,t) \in \mathcal{D} \times \mathcal{W} \times \mathcal{X}_t \times \{0,\dots,T\}}, \\ \{F_{C_d | D_t = d, X_t = x, W_t = w, t}(c_d)\}_{\forall c_d \in \mathcal{C}_{dt},\ (d,w,x,t) \in \mathcal{D} \times \mathcal{W} \times \mathcal{X}_t \times \{0,\dots,T\}} \Big\}.$$

This initial estimation of the reduced forms is crucial, as all the subsequent estimates are derived from it. The reduced forms probabilities $Pr(D_t = d | X_t = x, W_t = w, t)$ can be estimated non-parametrically by kernel or by Sieve logistic or probit regressions. Recall that these probabilities are not the CCPs, as the CCPs are also conditional on η_t.

The continuous choice conditional distributions can also be estimated with non-parametric kernel methods (e.g. Hayfield and Racine, 2008). Another alternative is to first estimate the quantile functions via non-crossing conditional quantile estimation (Muggeo, 2018; Lipsitz et al., 2017, for example), and then invert them to recover the conditional distributions.

In the dynamic setup, the reduced forms also include the *transition probabilities* from t to $t+1$: $f_t(x_{t+1}, w_{t+1} | c_t, d_t, x_t)$ is estimated as usual. Let us distinguish again the asset from other covariates: $x_t = (\tilde{x}_t, a_t)$. Under Assumption 13, $f_t(\tilde{x}_{t+1}, w_{t+1} | d_t, \tilde{x}_t)$ can be estimated using auto-regressive processes for the general covariates $\tilde{x}_t$. In the special case where $w_t = d_{t-1}$, then the transition of the instrument is given by construction. The asset plays a particular role and its transition is given by the budget constraint (1.7): $a_{t+1} = (1 + r_t)a_t - c_t + y_t d_t$.

Conditional Continuous Choices (CCCs):

I estimate the CCCs based on the identification proof. The idea is that we want to solve for the monotone functions $c_{dt}(h,x)$, which solves the empirical counterpart of system (1.2):

$$h = \widehat{F}_{C_{0t}|D_t=0,x_t,w_t}(c_{0t}(h,x_t))\ \widehat{Pr(D_t=0|w_t,x_t)} + \widehat{F}_{C_{1t}|D_t=0,x_t,w_t}(c_{1t}(h,x_t))\ \widehat{Pr(D_t=1|w_t,x_t)} \quad \forall w_t, h, x_t, t,$$

where I replaced the reduced forms by their empirical counterparts. In practice, solving for two functions c_0 and c_1 is not convenient. To simplify, I build upon the identification proof and I first estimate the monotone mapping $\widehat{c}_{0t}(c_{1t},x_t)$ between the two consumptions. Then I will estimate $\widehat{h}(c_1)$. Consider the empirical counterpart to equation (1.3):

$$\widehat{\Delta F}_{C_{0t}|x_t}(c_{0t}(c_{1t},x_t)) = -\widehat{\Delta F}_{C_{1t}|x_t}(c_{1t}) \qquad \forall c_{1t}$$

Thus, for any given x_t, I estimate the conditional consumption mapping $\widehat{c}_{0t}(c_{1t},x_t)$ by solving for the *whole* monotone mapping functions $c_{0t}(c_{1t},x_t)$ minimizing

$$\underset{c_{0t}(c_{1t},x_t)}{argmin} \int_{\mathcal{C}_1} \left(\widehat{\Delta F}_{C_{0t}|x_t}(c_{0t}(c_{1t},x_t)) + \widehat{\Delta F}_{C_{1t}|x_t}(c_{1t})\right)^2 \text{weight}(c_{1t})\ dc_{1t}.$$

It gives a weighted minimum distance estimator to solve for the whole function, instead of proceeding pointwise c_1 by c_1.[10] On a practical note, I resort to *constrained optimization* to solve for the function: select a grid of c_1, and search for the corresponding c_0 by imposing the monotonicity constraint that if $c_1^a < c_1^b$, then $c_0(c_1^a) < c_0(c_1^b)$ for every point in the grid. I repeat this estimation procedure separately for several values of x_t.

Once $\widehat{c}_{0t}(c_{1t},x_t)$ is estimated for all x_t, I can estimate $\widehat{h}_t(c_1,x_t)$ using any equation of system

[10]For the weight(c_{1t}), I put uniform weights on the quantile of c_{1t}. More precisely, I do not solve for the optimal consumption mapping, but for the optimal quantile mapping between the two consumptions. I look for the quantile of γ_0 of c_0 which corresponds to a given quantile of c_1 denoted γ_1. Thus, following the notation in the Appendix 1.B for the identification proof, the objective can in fact be written as:

$$\underset{\gamma_{0t}(\gamma_{1t},x_t)}{argmin} \int_0^1 \left(\widehat{\Delta F}_{0|x_t}(\gamma_{0t}(\gamma_{1t},x_t)) + \widehat{\Delta F}_{1|x_t}(\gamma_{1t})\right)^2 \text{weight}(\gamma_{1t})\ d\gamma_{1t}.$$

In this case, weight$(\gamma_{1t}) = 1$ for all γ_{1t}. And the support of the integral is simply $[0,1]$ since the quantiles are uniform.

(1.2) (with $w_t = 0$ or $w_t = 1$) as:[11]

$$\widehat{h}_t(c_{1t}, x_t) = \widehat{F}_{C_{0t}|D_t=0,w_t,x_t}(\widehat{c}_{0t}(c_{1t}, x_t))\ \widehat{Pr(D_t = 0|}w_t, x_t) + \widehat{F}_{C_{1t}|D_t=0,w_t,x_t}(c_{1t})\ \widehat{Pr(D_t = 1|}w_t, x_t).$$

Once I have estimated the monotone functions $(\widehat{h}_t(c_{1t}, x_t), \widehat{c}_{0t}(c_{1t}, x_t))$ for all $c_{1t} \in \mathcal{C}_1$ and for all x_t, I easily recover the CCCs:

$$\{\widehat{c}_{0t}(h, x_t), \widehat{c}_{1t}(h, x_t)\} \qquad \forall (h, x_t) \in [0, 1] \times \mathcal{X}$$

by flipping the arguments (because everything is monotone).

Conditional Choice Probabilities (CCPs):

Once the monotone CCCs are estimated, I estimate h_t from observed (c_t, d_t, x_t) in the data, by inverting the CCCs.

$$\text{If } d_t = d: \quad \widehat{h}_t = \widehat{c}_{dt}^{-1}(c_t^{obs}, x_t).$$

Then, you can use $\widehat{h}_t$ as if it was *observed* (like a generated covariate), and estimate the Conditional Choice Probabilities

$$\widehat{Pr}(D_t = d|\eta_t = h, X_t = x_t, W_t = w_t).$$

Again, similarly to the reduced forms probabilities, this estimation can be done non-parametrically with kernel or by Sieve logistic or probit regressions.

Alternative methods:

One could resort to estimation methods proposed in the IV-quantile treatment effect literature and based on Chernozhukov and Hansen (2006, 2008), or that based on Vuong and Xu (2017), described in Feng et al. (2020). With respect to these methods, the advantage of the method developed here is

[11] In practice, the two equations might not yield exactly the same results because of the noise in the estimation of the reduced forms. Therefore, I will estimate two different h with each equation (with $w = 0$ and $w = 1$) and obtain my final estimate by weighting the two estimates by the number of observations when $(X_t = x_t, W_t = 0)$ and $(X_t = x_t, W_t = 1)$.

that it is entirely based on the constructive identification proof and does not impose any additional assumptions. The estimation is more flexible, and does not require *full rank* or other assumptions on the conditional choice probabilities (as in the practical estimation paper of Feng et al. (2020)) to hold, for example.

Alternatively, the CCCs and CCPs coud be jointly estimated by Sieve, directly from the data, without estimating reduced forms beforehand. Indeed, for any CCC guess (which has to be monotone in η), one can recover the corresponding η from observing c_d in the data. Joint with a CCP guess, one can derive the likelihood of any data point. Therefore, the CCCs and CCPs can be estimated directly by Sieve maximum likelihood.

1.5.2 2nd stage: structural model

I provide an estimation method for *parametric* models here. I do so for practical reasons since this avoids the curse of dimensionality and because it fits most applications. Assume the model is parametrized by $\theta \in \Theta$. As I did not address the identification of β (Magnac and Thesmar, 2002), I do not estimate it either, so it does not enter θ. The parameters θ can be divided into two parts $\theta = (\theta_0, \theta_1)$: where θ_0 enters the marginal utility and θ_1 does not. In the setup, u is parametrized by θ_0, and denoted $u_d(c_{dt}, \tilde{x}_t, \eta_t, \theta_0)$. The additive term $m_{dt}(x_t, w_t, \eta_t, \theta_1)$ is parametrized by θ_1. More precisely, θ_1 impacts the difference $\Delta m_t(x_t, w_t, \eta_t, \theta_1)$ of $m_{1t}(x_t, w_t, \eta_t) - m_{0t}(x_t, w_t, \eta_t)$, since only the difference is identified by the discrete choices.

I want to estimate θ. To do so, I use the CCCs, the CCPs and the transition estimated in the first stage. My estimation method is based on the minimization of two different objectives identifying different parameters: one based on the Euler equation and the other based on the conditional choice probabilities.

Euler objective:

Recall the notation for the marginal utilities at the optimal CCCs:

$$u_d'^*(\tilde{x}_t, a_t, \eta_t, \theta_0) = \frac{\partial}{\partial c_{dt}} u_d(c_{dt}, \tilde{x}_t, \eta_t, \theta_0)|_{c_{dt}=c^*_{dt}(\eta_t, \tilde{x}_t, a_t)}.$$

Thus, we have the Euler equation:

$$
\begin{aligned}
& u_d'^*(\tilde{x}_t, a_t, \eta_t, \theta_0) = \beta(1+r_t)\mathbb{E}_t\Big[u_{d_{t+1}}'^*(\tilde{x}_{t+1}, a_{t+1}, \eta_{t+1}, \theta_0) \ \Big| x_t, c_{dt} = c_{dt}^*(\eta_t, x_t), d_t = d\Big] \\
\overset{def}{\Longleftrightarrow} \quad & q_1(t, d_t, \eta_t, x_t, \theta_0) \ = q_2(t, d_t, \eta_t, x_t, \theta_0),
\end{aligned}
$$

where $x_t = (\tilde{x}_t, a_t)$.

The CCCs and the CCPs have been estimated in the first stage for all d_t, x_t, η_t, t. I also estimated the transitions. Thus, I can estimate θ_0 as:

$$
\min_{\theta_0} Q^{euler}(\theta_0) = \sum_i \left(q_1(t, d_t, \eta_t, x_t, \theta_0) - q_2(t, d_t, \eta_t, x_t, \theta_0)\right)^2 .
$$

In other words, $\hat{\theta}_0$ minimizes the differences between the two sides of the Euler equation for every observation i in the sample.[12] Now, q_1 is directly given as a function of θ_0 and of the observed characteristics and choices. q_2, on the other hand, contains an expectation and can be computed in several ways.

The first way is to use individuals present for two consecutive periods and to estimate the expectation of future utility using all individuals with the same current states x_t, c_{dt}, d_t. Since c is continuous and x contains continuous covariates, this can be done parametrically or via non-parametric kernel mean regression. This method is the simplest, but it requires many observations. It is close to the idea of Euler-GMM estimation, as pioneered by Hansen and Singleton (1982). The problem is that when the marginal utilities are highly nonlinear, the expectation is poorly estimated and this type of GMM estimation does not work well and needs to be refined Alan et al. (2009).

Hence, I prefer to use an alternative approach based on *forward simulations*, in the spirit of (Hotz et al., 1994). The idea is to use the CCCs, the CCPs and the transition to estimate the expectation term via one-period-ahead simulation. This method is slightly longer but less affected by the nonlinearity problem. It requires to estimate the transitions consistently.

[12] In practice, it is better to rewrite the objective such that it is scale invariant and comparable for all the values of θ_0. For example, if ceteris paribus a specific parameter value in θ_0 scales down everything in q_1 and q_2, the errors will be small at this parameter value, regardless of whether or not this parameter is far from the truth. To avoid that, one need that, for any set of parameters θ_0, the Euler Equation errors are on a similar scale. Log-linearization can be used to achieve this, for example. In the next section the parametric model is such that I can isolate consumption on the left-hand side of the Euler Equation. In this case, the left-hand side of the equation is based on consumption data and is independent of the parameters. Thus I can compare the results between parameters on the same basis.

Probability objective:

Now, the Euler equation does not provide any information about the parameters impacting the differences of the additive term, θ_1. To estimate these parameters, I use the relation between the choice probabilities and the conditional value function (Hotz and Miller, 1993). In particular, if ϵ is extreme value type I, we have:

$$Pr(D = 0|\eta_t, x_t, w_t, \theta) = \frac{1}{1 + exp\left(v^*_{1t}(x_t, \eta_t, \theta) + m_{1t}(x_t, w_t, \eta_t, \theta_1) - (v^*_{0t}(x_t, \eta_t, \theta) + m_0(x_t, w_t, \eta_t, \theta_1))\right)} \quad (1.11)$$

Knowing θ, one can estimate the conditional optimal values $v^*_{dt}(x_t, \eta_t, \theta)$ by *forward simulation* of the life-cycle, for example (Hotz et al., 1994). Note that the value functions are parametrized by θ and not only θ_1. Thus, a way to estimate the parameters is to minimize the differences between the estimated CCPs and the theoretical probabilities (equation (1.11)) with respect to θ for all observations:

$$\underset{\theta}{min} \quad Q^{proba}(\theta) = \sum_i \Big(Pr(D = 0|\eta_t, x_t, w_t, \theta) - \widehat{Pr(D = 0|\eta_t}, x_t, w_t) \Big)^2 .$$

Global objective:

There are two *consistent* ways to estimate θ. The faster one is to perform the estimation in two separate steps: (i) estimate θ_0 from the Euler equation and (ii) estimate the remaining θ_1 from the probability objective (taking $\hat{\theta}_0$ as given). This yields a consistent estimation of θ.

However, the probability objective also depends on (part of) θ_0 which is identified by the Euler equation. An efficient way to account for this information is to perform the estimation in one step and find the parameters θ that minimize a weighted sum of both objectives:

$$\hat{\theta} = \underset{\theta}{argmin} \quad \text{weight}^{euler} Q^{euler}(\theta_0) + \text{ weight}^{proba} Q^{proba}(\theta),$$

where the optimal weights are to be determined. At the optimal weights, the one step method is consistent and more efficient than the two-step estimation.

1.6 Estimator Performance

I test my estimator's performance with Monte Carlo simulations of the estimation of a parametric toy model of simultaneous labor and consumption choices. This model is a simplified version of the application performed in the next section. I provide additional robustness checks in Appendix 1.C.

1.6.1 Toy model

The agent chooses to work (d_t) and consume/save (c_t) from $t = 1$ to $t = T$. Then she retires for one period in $t = T + 1$. She dies in $t = T + 2$.

Working life:

In each period the agent obtains utility:

$$u_{dt}(c_t, x_t, \eta_t) + m_{dt}(x_t, w_t, \eta_t) + \epsilon_{dt} = \begin{cases} c_t^{1-\sigma}/(1-\sigma)\ \tilde{\eta}_t^0(\eta_t, \gamma_0, s_0) + \epsilon_{0t} & \text{if } d_t = 0 \\ c_t^{1-\sigma}/(1-\sigma)\ \tilde{\eta}_t^1(\eta_t, \gamma_1, s_1) + \alpha + \omega(1 - w_t) + \epsilon_{1t} & \text{if } d_t = 1 \end{cases}$$

subject to the budget constraint:

$$a_{t+1} = (1 + r)a_t + d_t y_t - c_t + (1 - d_t)b_t.$$

t is the age of the agent. c_t is the individual consumption. d_t is the labor choice, equal to 1 if she works. w_t is the instrument, equal to the past labor choice d_{t-1}. a_t is the asset holdings. b_t represents benefits earned by unemployed people. y_t represents the earnings. y_t take only two values, y_L and y_H, for low and high income. In this way, the asset is the only continuous covariate, and I can reduce the state space with only two values in the support of y. I observe the income for every individual, even when she does not work. The interest rate r is fixed and equal to 0.05. $\epsilon_t = (\epsilon_{0t}, \epsilon_{1t})$ are additive idiosyncratic shocks impacting preferences for work. They are extreme-value type I. $\tilde{\eta}^d$ are nonseparable taste shocks to utility. $\tilde{\eta}^d(\eta, \gamma_d, s_d)$ is the η^{th} quantile of a lognormal(γ_d, s_d) distribution. In other words, $\tilde{\eta}^d \sim \mathcal{LN}(\gamma_d, s_d)$, so that $\tilde{\eta}^d$ are labor-dependent monotone transformations of the uniform η. Having η as quantiles of some specific distribution is a convenient way of modelling unobserved taste shocks in this type of setup. Thus, $(\gamma_0, \gamma_1, s_0, s_1)$ capture the different effects of unobserved taste shocks on the utility depending on working choice. I normalize $\gamma_0 = 0, s_0 = 0.25$ to interpret the parameters of working individuals with respect to this reference.

The other parameters are more conventional: σ is the risk aversion or intertemporal elasticity of substitution, α is the utility cost of work, and ω is the cost of searching for a job when one was previously unemployed ($w_t = 0$). Thus, $\theta = (\overbrace{\sigma, \gamma_0, \gamma_1, s_1}^{\equiv\theta_0}, \overbrace{\alpha, \omega}^{\equiv\theta_1})$, where θ_1 only impacts the probability of working and not the consumption choices, and θ_0 impacts both.

Transitions:

The asset transition is given by the budget constraint.

In the income transitions, I model gains from working experience: $Pr(y_{t+1} = y_H | d_t = 1, y_t) > Pr(y_{t+1} = y_H | d_t = 0, y_t)\ \forall y_t$. Income is also persistent, so if one had a high income in t, one is more likely to obtain a high income in $t+1$: $Pr(y_{t+1} = y_H | d_t, y_t = y_H) > Pr(y_{t+1} = y_H | d_t, y_t = y_L)\ \forall d_t$. It yields the following transition matrix:

$$Pr(y_1 = y_H | d_0, y_0) = \Pi(d_0, y_0) = \begin{pmatrix} \pi_{0L} & \pi_{0H} \\ \pi_{1L} & \pi_{1H} \end{pmatrix}$$

where $\pi_{1L} > \pi_{0L}$, $\pi_{1H} > \pi_{0H}$, $\pi_{1H} > \pi_{1L}$, and $\pi_{0H} > \pi_{0L}$.
These four parameters are estimated directly from the data by estimating $Pr(y_t = y_H | y_{t-1}, d_{t-1})$ with a bin operator, i.e., by computing the number of observations with $y = y_H$ over the total of observations with each specific y_{t-1}, d_{t-1} combination.

The shocks are iid and uncorrelated over time $\eta_{t+1} \perp \eta_t$ and $\epsilon_{t+1} \perp \epsilon_t$.

The agent discounts the future with discount factor β. I set it to 0.98 and do not estimate it.

Retirement:

At period $T+1$ the woman retires. She only consumes and can no longer work. She obtains the same period utility as when she was unemployed, without the additive ϵ shock. She obtains a $pension(y_T)$, which is a proportion set to 50% of her last income y_T. She lives for only one period in retirement and knows that she will die at $t = T + 2$.[13] There is no bequest motive. As a consequence, she will consume everything, i.e., $a_{T+2} = 0$. Thus, the last period consumption has a closed-form solution:

$$c^*_{T+1} = (1 + r)a_{T+1} + pension(y_T).$$

1.6.2 Comparison

I run Monte Carlo simulations of this toy model and estimate the parameters $\theta = (\sigma, \gamma_0, \gamma_1, s_1, \alpha, \omega)$ using my method. I compare my results with indirect inference Simulated Method of Moments (SMM) where the model is solved using Endogenous Grid Method (Iskhakov et al., 2017).

Results: (Table 1.1)
Table 1.1 shows the estimation results for a model with $T = 2$ periods.[14]

In terms of *speed*, my two-step method (*DCC* for Discrete-Continuous Choices) yields *sizeable computational gains*, even with respect to the state-of-the-art indirect inference method with endogenous grid (Carroll, 2006; Iskhakov et al., 2017). The idea is simple: I have a fixed computational cost of estimating the CCCs and CCPs in the first stage. However, thereafter, when solving for the optimal θ, I do not need to solve the model again, as I already have the optimal choices. I only need to perform some quick computations of the marginal utilities. Concerning the forward simulation, it is also a fixed cost, as the simulated path depends on first-stage CCCs, CCPs and transitions but not on θ. Thus, I only simulate forward once, and I retain the same path (as it is recommended to have the same basis for every set of parameters and to avoid adding some simulation noise to the estimation) for the computation of the expectations and conditional value functions for each tested set of parameters. The other methods, on the other hand, do not have my first-stage fixed cost, but they require considerably more computations in the second stage. If the model is very simple and the second stage is estimated quickly, these methods can perform quicker than mine in theory. However, as is well known, life-cycle models require a long time to solve, and the computational burden increases almost exponentially with the complexity of the model (more covariates, more

[13]This is a simplification; one could easily allow for a known length of retirement and solve the dynamic consumption problem of the retiree accordingly (as I do in the application). However, this does not deliver any particular insights in terms of estimator comparison, so I have agents live for only one period of retirement for simplicity.

[14]Coded in *R*, without parallelization here. Time results obtained from an Intel(R) Core(TM) i7-9750H CPU.

Table 1.1: $T = 2$ periods

	Truth	*Method* *DCC*	*SMM*
N		10,000	10,000
σ	1.60	1.6253	1.5924
		(0.0410)	(0.0156)
γ_1	0.00	0.0070	-0.0052
		(0.0238)	(0.0055)
s_1	0.40	0.4078	0.4001
		(0.0228)	(0.0071)
α	-0.50	-0.4727	-0.5023
		(0.0498)	(0.0348)
ω	-1.00	-0.9982	-0.9972
		(0.0581)	(0.0523)
Average Time taken:			
1st stage: CCPs and CCCs		118s	9s
2nd stage: Structural parameters		170s	14328s
Overall		288s	**14337s**

Other initializations: Number of Monte-Carlo $= 1,000$

$Pr(w_1 = 1) = 0.70$. $y_1 = y_H$ with probability 0.50. $a_1 \sim \mathcal{U}(0, 30)$. $r = 0.05$.

parameters, more periods). It can take several minutes to solve for one tested set of parameters, and finding the optimal parameters may require hundreds or even thousands of tests. Even here, in this very simple example with two periods, solving the model and estimating the moments for one set of parameters takes about 25 seconds with EGM. While with my method, computing the objective for one set of parameter takes less than a second. The more complex the model, the more this gap widens, as solving the model becomes even longer relatively speaking, while the fixed cost of computing the first stage policy only takes slighly longer. As a consequence, my two-stage method yields significant computational gains by reducing the burden of the second stage. Interestingly, *the more complex the model is, the more computation gains from my estimation method* relative to others. Obviously, having more complex models increases my computation time, but not in the same exponential manner as for the alternative methods.

In terms of statistical performances, my two-step method (DCC) estimates the parameters consistently and with small standard errors. As also shown in the simpler case of Appendix 1.C where $T = 1$, Simulated Method of Moments (SMM) is consistent and more efficient for most parameters (if one uses a lot of moments). Both methods build upon the same initial estimation of the reduced forms conditional consumption distribution $F_{C_d|D_t,X_t,W_t,t}(c)$ and probability of working $Pr(D_t = 1|X_t, W_t, t)$. The moments are selected in these two objects.[15] And my first stage estimation of the CCCs is also built on these objects. Therefore, it is not surprising that both methods yield close results. I lose some efficiency due to the two-step nature of my method, similar to the efficiency loss of CCPs estimators. In theory, MLE is more efficient than both (see the $T = 1$ case in Appendix 1.C), but it quickly becomes intractable to compute empirical likelihood with more periods.

Another advantage of my method is that if the model is misspecified, I still recover correct optimal choice estimates because the first stage is independent of the model assumptions (except for the choice of covariates to include). This is not the case for the alternative methods. Also, as I do not solve numerically for the optimal choices, I do not need to smooth potential kinks introduced by joint discrete-continuous choices, contrary to indirect inference with endogenous grid method (Iskhakov et al., 2017).

Overall, I have a method that is statistically consistent, with small standard errors, and considerably faster, by several orders of magnitude, than alternative state-of-the-art indirect inference with endogenous grid method. SMM built upon moments drawn from the reduced forms is also consistent and more efficient but the computational burden is too heavy for complex life-cycle models.

1.7 Application: women's labor and consumption

I illustrate the method developed in this paper with a parametric dynamic model of simultaneous employment and consumption choices for women over their life cycle. I choose a parametric

[15] If we do not pick moments in the reduced forms, i.e., moments conditioned by w, then the model is not identified and the SMM estimation will not be consistent.

application for practical reasons: to avoid the curse of dimensionality and to be able to compare my parameter estimates with the literature. This model especially matters for understanding how different benefit schedules affect the careers of women, particularly mothers, who are known to be the most responsive to incentives (Blundell and Macurdy, 1999; Blundell et al., 2016a). It allows us to understand the mechanism underlying individual choices and thus to carry out counterfactual policy analysis in the long run.

My method is of particular interest for two main reasons here. First, life-cycle models such as that presented here are extremely computationally intensive to estimate, to the extent that one often needs to restrict the complexity of the model for the estimation to be tractable. By first estimating the optimal choices (CCCs and CCPs) and only then the structural parameters in a second step, I do not need to solve the model, and I am able to drastically reduce the computational cost, in the spirit of Hotz and Miller (1993), Hotz et al. (1994) and Arcidiacono and Miller (2011). Faster computation means that one can include more features in existing models, for example, more heterogeneity, observed or unobserved, and still be able to estimate them in a reasonable time. The complete estimation of this complex model only takes me a few hours here, while it could take weeks or months with alternative methods.

In addition to the speed increase, I also include more unobserved heterogeneity in the model with the η term. Thus, by construction, I estimate the *distribution* of consumption choices and working probabilities at any given set of observed covariates, and not only the average choices. This yields new insights in this literature.

1.7.1 Model

Overview:

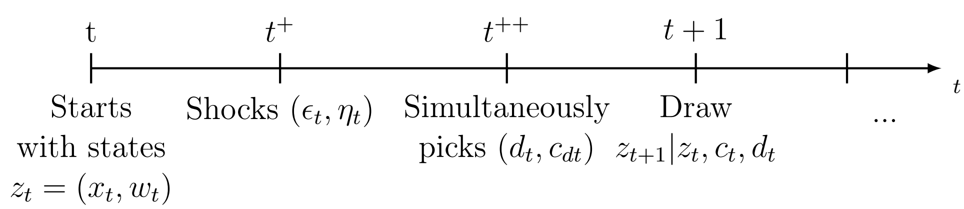
The parametric model enters the general dynamic framework described in section 1.4.1. It is a more realistic version of the toy model described previously, with the same key features. I model the annual consumption and labor supply choices of women from $t = 26$ to $t = 60$ years of age. Each period, women determine their household consumption c_t, and whether they work d_t.[16] At the age 60, they retire and live for 15 more years on their accumulated savings and their pension, which depends on their last income. Throughout their life, women may bear children. Fertility oc-

[16] I focus on the extensive margin, not on the number of hours worked. Individuals are assumed to either work full time or be unemployed. This might be restrictive, particularly for single mothers, who are known to resort more to part-time jobs.

curs randomly following the trend observed in the data, and is not explicitly modelled as a choice. Couples do not divorce, and new couples are not formed in the model. This is for simplicity to avoid dividing the assets or modelling individual husbands' assets. Women's productivities (and thus wages) evolve over their careers. Labor supply choice plays a key role in this evolution, as working experience increases expected future wages, while productivity can depreciate for unemployed women. Similarly, asset holdings evolves over the life cycle following a budget constraint that depends on previous asset holdings, consumption, women's productivities and labor choices (they are paid only if they work), their potential husband's annual income and the tax schedule to which they are subject. The benefit/tax schedule is simplified and estimated based on observed data. It differs depending on the individual's family situation, wealth and labor choice. Finally, women are subject to unobserved preference shocks η and ϵ. η is their unobserved taste shock for consumption, and ϵ represents their unobserved preference for work. With η, I can estimate heterogeneous consumption choices for individuals who are identical as measured by their covariates.

I now describe the model in greater detail.

Working life:



From age $t_0 = 26$ to age $T = 60$, a woman is in her working life. She makes her decision (d_t, c_t) to maximize her expected lifetime utility given her characteristics. These characteristics include her age (t), her income (y_t), her assets (a_t) and some demographics $\tilde{x}_t$: her number of children ($nchild_t$), whether she is in a couple ($couple_t$), and if so, her partner's annual income (y_t^p) and labor force participation d_t^p. All these covariates are included in $x_t = (a_t, y_t, \tilde{x}_t)$.[17] Her decision to work is also influenced by whether she worked before $w_t = d_{t-1}$, for which we observe $w_0 = d_{-1}$. w_t matters because of the utility cost of switching from being unemployed to employed. She also makes her decision based on two idiosyncratic shocks η_t and ϵ_t, unobserved by the econometrician. To satisfy

[17] A small detail: y_t is now excluded from $\tilde{x}_t$, at minimal risk of confusion. It is because y_t does not enter the current period utility directly.

the distributional Assumptions 4-7, we have $\eta_t \sim \mathcal{U}(0,1)$ iid over time, and ϵ_t is i.i.d. extreme-value type I.

Each period, the agent obtains utility:

$$u(c_t, d_t, w_t, x_t, \eta_t, \epsilon_t) = \begin{cases} (c_t/n_t)^{1-\sigma}/(1-\sigma)\ \tilde{\eta}_t^0(\eta_t, couple_t, nchild_t) + \epsilon_{0t} & \text{if } d_t = 0 \\ (c_t/n_t)^{1-\sigma}/(1-\sigma)\ \tilde{\eta}_t^1(\eta_t, couple_t, nchild_t) + \alpha + \omega(1-w_t) + \epsilon_{1t} & \text{if } d_t = 1 \end{cases}$$

$$\equiv \begin{cases} u_0(c_t, \tilde{x}_t, \eta_t) + \ \epsilon_{0t} \\ u_1(c_t, \tilde{x}_t, \eta_t) + \ \underbrace{\alpha + \omega(1-w_t)}_{=m_1(w_t)} + \epsilon_{1t} \end{cases}.$$

where c_t is the total household consumption over the period, n_t is an equivalence scale, which depends on the number of consumption units in the household, i.e., $n_t(couple_t, nchild_t)$, with $n_t(0,0) = 1$, $n_t(1,0) = 1.6$, $n_t(0, 1 \text{ or more}) = 1.4$, $n_t(1, 1 \text{ or more}) = 2$ (Blundell et al., 2016a). Thus, c_t/n_t represents *individual consumption*. σ is the elasticity of intertemporal substitution/risk aversion parameter. The effect of the unobserved shock η_t varies depending on the work choice (d_t) and family situation ($couple_t, nchild_t$). $\tilde{\eta}_t^d$ are transformations of η_t, where $\tilde{\eta}_t^d \sim \mathcal{LN}(\gamma_d + \gamma_d^c couple_t + \gamma_d^n nchild_t, s_d)$. $\tilde{\eta}_t^d$ are the η^{th} quantiles of these distributions. This is a convenient way to include covariates in this setup. Since $\eta_t \sim \mathcal{U}(0,1)$, the transformation to $\tilde{\eta}_t^d$ allows for a wide range of effects of η_t. Note that with this functional form, the monotonicity Assumption *D*3 is satisfied. The parameters (γ_d, s_d) represent the baseline effect of unobserved heterogeneity depending on working behaviour, for single women without children. The parameters (γ_d^n, γ_d^c) determine the effect of the family situation. I set γ_0 to 0 and s_0 to 0.5 so that the other coefficients are interpreted with respect to this baseline. The agents incur a utility cost α from working. w_t is the instrument that corresponds to the past labor choice, $w_t = d_{t-1}$. The agents incur a an utility cost ω from searching for a job (if they were previously unemployed). Thus, $m_{dt}(w_t, x_t, \eta_t) = \alpha d_t + \omega(1-w_t)d_t$, and it is independent of x_t, η_t and t in this application (which is stronger than necessary for the identification). By the additivity of the instrument w_t, Assumption *D*2 is satisfied. Similarly, additive separability of ϵ_t (Assumption *D*1) is satisfied. Note that I have time independent current utility. However in the parametric framework I can identify time-varying utility (Assumption 14 is only necessary for non-parametric identification). Thus one could include and estimate time fixed effects for example.

Transition:

The woman makes her choice of d_t, c_t subject to the *household budget constraint*:

$$a_{t+1} = (1+r)a_t - c_t + d_t y_t + \text{couple}_t d_t^p y_t^p + T(d_t, x_t).$$

This budget constraint describes the asset transition over time. r is the real interest rate. If the woman works $d_t = 1$, she obtains a wage y_t. If she has a husband ($couple_t = 1$) who works ($d_t^p = 1$), the household also obtains his total income ($= 0$ if he does not work). $T(d_t, x_t)$ is the *benefit-tax* schedule. It is a function of the covariates and labor choice. I estimate it directly from the data.

Earnings y_t and husband's earnings y_t^p evolve over time according to an auto-regressive process:

$$y_{t+1} = (\rho_y y_t + \rho_d d_t + \rho_{age} t) \times \text{educ}_t + u_t$$
$$y_{t+1}^p = \rho_y^p y_t^p + \rho_d^p d_t^p + v_t,$$

where u_t, v_t may be correlated. Working ($d_t = 1$) allows individuals to change their expected earnings and potentially increase them. Unemployment will decrease productivity if $\rho_y < 1$, i.e., human capital depreciates. Therefore, working is important not only for current consumption and savings but also for its impact on human capital accumulation. All these coefficients vary with the education level ($\leq$secondary, high school or university) of the woman, educ_t. I do not include the education of the partner to avoid having too many variables in the model, since I focus on the woman. The earning process is estimated directly from the data on observed transitions.

Auto-regressive processes are also estimated for fertility (having a new-born child) and for the husband's work decisions. These depend on past x_t only. d_t and c_t do not enter the transitions here.

Finally, by construction, the next value of the instrument $w_{t+1} = d_t$ and the other covariates than d_t are irrelevant for its transition. Since conditional on current (d_t, c_t, x_t), w_t does not enter the transition of the other variables, the conditional independence Assumption 10 is satisfied.

Retirement:

At age T, the woman retires, and can no longer decide to work. She obtains the same utility as when she did not work, with $d_t = 0$, without the additive ϵ shock.[18] She lives for another 15 years on her accumulated assets and receives a pension that is a proportion of her last income y_T. I include no bequest motive in the model. One can easily solve the consumption problem of retirees, which depend on their last income and assets, to obtain the expected *retirement utility*: $R(x_T) = R(a_T, y_T, y^p_T, couple_T, nchild_T)$.

Life-cycle problem:

The working life decision problem is the one that interests us. Given the development above, at any age t during her working life, the woman decision problem can be written as:

$$V_t(z_t) \equiv \mathbb{E}\Big[\sum_{\tau=t}^{T} \beta^{\tau-t} \underset{d,c_{d\tau}}{max}[\ u_d(c_{d\tau}, \tilde{x}_\tau, \eta_\tau) + m_d(\tilde{x}_\tau, w_\tau, \eta_\tau) + \epsilon_{d\tau}] + \beta^{T-t}R(x_T)\Big],$$

where the future is discounted at a rate β.

For notational simplicity, denote $V_{T+1}(z_{T+1}) = R(x_T)$, a special form of the value function for the retirement period. We have already verified that the identification assumptions hold. Therefore, following computations performed in section 1.4.1, we return to the general setup, where the woman selects d_t and c_t at each age t to solve:

$$\underset{d_t,c_{dt}}{max} \quad v_d(c_{dt}, x_t, \eta_t) \ + m_d(\tilde{x}_t, w_t, \eta_t) + \epsilon_{dt}, \tag{1.12}$$

where the conditional value function is given by:

$$v_{dt}(c_{dt}, x_t, \eta_t) \ \equiv \ u_d(c_{dt}, \tilde{x}_t, \eta_t) + \beta \mathbb{E}_{z_{t+1}}[V_{t+1}(z_{t+1})|_{x_t, c_{dt}, d_t}].$$

The agent internalizes the effect of her choice on her future, discounting it at a rate β. Note that even here where the current utility has a known parametric form, the conditional value's form is generally more complex, with no closed form solution. It depends on complex transitions and expectations about the future. Therefore, the advantage of my method, i.e., that I am able to estimate

[18] Instead, I can normalize the utility of retired women with parameters γ_r and s_r. In which case, I could estimate the γ_0 and s_0 of unemployed women with respect to the retirees baseline. However, my estimation would then be driven by the data comparison of the Euler equation at the retirement age, which represents only a small subset of my panel. Therefore, I prefer to set the utility of retirees equal to the utility of unemployed individuals.

both the optimal policies (CCCs and CCPs) and the parameters of interest without numerically solving for the conditional value, also applies to parametric dynamic models, hence the sizeable computational gains.

1.7.2 Data

To estimate the model, I use EU-SILC French survey data. It is a survey conducted every year and follows households from 2004 to 2015. The data contain information about the labor market status (income, job tenure), asset holdings (financial and housing), tax paid and benefits received, and personal characteristics of the individuals (family situation, education, etc.). Data are available for all the individuals within the household, which is why I also have detailed information about the partner.

Consumption is not directly available in the data, I reconstruct it for households present over two consecutive years based assets evolution and savings.

I set that a woman *works* ($d = 1$) if she worked more than 6 months during the year.

Income is only reported for employed women and husbands. I estimate y and y^p based on the income information of workers using the standard Heckman correction (Heckman, 1979) beforehand. For simplicity, I assume that income is observed for everyone using these estimations when I estimate the model.[19] I estimate this on the subsample of full-time working individuals so that I obtain a productivity y_t representing full-time equivalent productivity. In this estimation, I include covariates other than those used in the model, including education, experience, some parent background information, and zone and year fixed effects.

After cleaning the data for outliers and missing values, I end up with an unbalanced panel of $7,391$ women between 26 and 60 years of age, yielding a total of $21,945$ observations over 11 years.

I fix the real interest rate r at the average of the period ($= 0.05$), as given by the IMF French data.

Descriptive statistics:

Table 1.2 describes the sample of data I use; 76% of women work, with a strong auto-correlation in

[19]Obviously, given that the probability of working is a key part of the model, ideally one would prefer to build our own Heckman correction within the model here, with some kind of nested iteration with the CCP, CCC and productivity estimation, in the spirit of Aguirregabiria and Mira (2002) for example. Another good way to deal with it would be to include *unobserved types*, as in Arcidiacono and Miller (2011), and have wage depend on these types.

Table 1.2: EU-SILC unbalanced panel, 2004 − 2015, 7391 women

Statistic	N	Mean	St. Dev.	Min	Median	Max
Choices:						
Annual household c (k euros)	21,945	36.58	20.99	3.88	32.54	211.54
$c \vert d = 0$	5,330	30.04	19.32	4.02	25.58	204.48
$c \vert d = 1$	16,615	38.67	21.07	3.88	34.78	211.54
d	21,945	0.76	0.43	0	1	1
w	21,945	0.76	0.43	0	1	1
$d \vert w = 0$	5,354	0.14	0.35	0	0	1
$d \vert w = 1$	16,591	0.96	0.20	0	1	1
Covariates:						
Age	21,945	42.37	9.39	26	42	60
Annual Income y (Heckman)	21,945	19.74	5.29	8.10	19.07	43.32
Asset	21,945	108.29	118.55	−32	69.0	528
Nb of children	21,945	1.71	1.09	0	2	4
Couple	21,945	0.75	0.43	0	1	1
Working partner|Couple	16,442	0.93	0.25	0	1	1
Partner's income y^p |Couple	16,442	26.41	13.21	4.02	23.20	147.54
Completed Education	21,945					
≤ Secondary	5,240	0.24				
High School	9,999	0.46				
University	6,706	0.30				
Other:						
Receives Benefits	21,945	0.66	0.47	0	1	1
Benefits|Benefits > 0	14,478	5.16	4.46	0.002	3.60	23.07

c, y, y^p, asset and benefits expressed in real terms (base 2010) and in thousands of euros.

employment: if a woman worked before ($w_t = d_{t-1} = 1$), $Pr(D_t = 1|w_t = 1)$ is very high = 0.96, while if she did not, $Pr(D_t = 1|w_t = 0) = 0.14$ is low. This suggests that w should be a *relevant* instrument for d. On average, households consume $36k5$ euros per year. Observed consumption conditional on working ($c_{1t}|d_t = 1$) is higher than consumption conditional on being unemployed ($c_{0t}|d_t = 0$). However, we do not yet know how much of this is due to the selection: it is possible that women with high η_t select more into employment, boosting the average consumption conditional on working. Regarding the covariates, there is considerable variance in asset holdings. Most women (75%) are in couples, with a median number of 2 children. Their partner is generally working (93%), far more frequently than the women. The partner's income is also larger than the woman's annual income. Finally, approximately 66% of the households received some kind of benefit. These benefits include not only unemployment benefits, but also family benefits, for example. This is why there are more people receiving benefits than the number of unemployed people.

1.7.3 1st stage: Optimal Choices Estimation

Estimation

I follow the procedure described in section 1.5. First, I estimate the reduced-form probabilities and conditional distributions. The probabilities are estimated with a sieve logistic regression. The conditional distributions are estimated via non-parametric kernel methods (R package *np*, Hayfield and Racine, 2008).

Then, I estimate the CCCs $c_{dt}(h_t, x_t)$ for all h_t, x_t accordingly. Once I have the CCCs, I estimate $\hat{h}_t$ from the observed (c_t, d_t, x_t). I then recover the CCPs $\widehat{Pr}(D_t = d|\eta_t = h, X_t = x, W_t = w_t)$ using a sieve logistic regression.

I also estimate the transitions to $X_{t+1}|X_t, D_t, C_t$ according to the description provided previously.

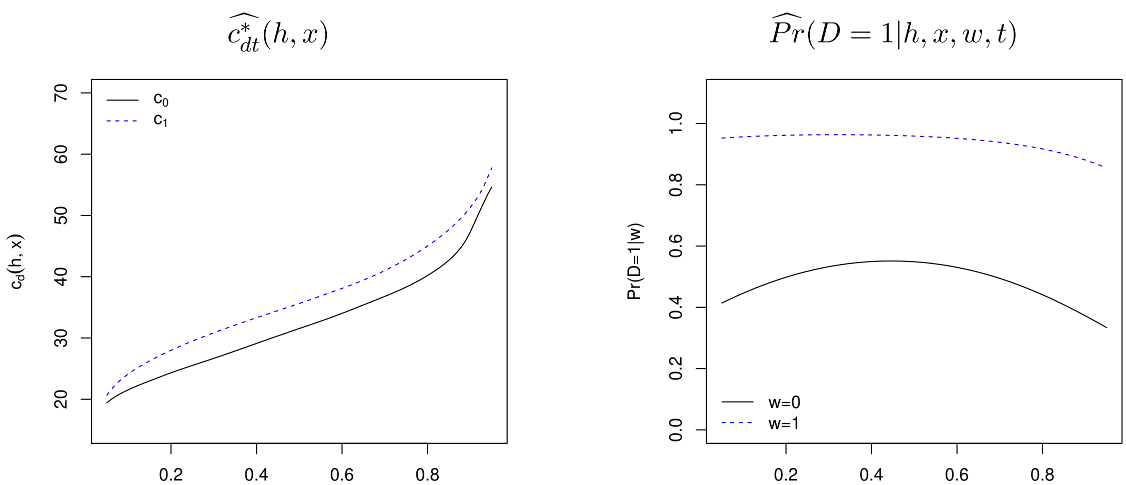
Results

Optimal Choices:

Figure 1.1 shows the optimal choice estimates for the median 26 year-old woman. Potential consumption when working is always greater than alternative consumption when unemployed, with an average difference of approximately $3,930$ euros of consumption per year. By construction, these functions are monotone with respect to the taste shock h.

The conditional choice probabilities are more complicated. First, note that the probability of working is always greater and close to 1 for individuals who were employed previously. If the median woman was previously unemployed, however, her probability of working today is less than 50%. The relation between the employment decision and the taste shock is complex. By working, the woman will obtain an income that she will be able to consume. However, at the same time, she will have less leisure time. There is a trade-off between a substitution and wealth effect. If she was previously employed ($w_t = 1$), the higher her taste shock (η_t) was, the less likely the median

Figure 1.1: CCCs and CCPs estimates



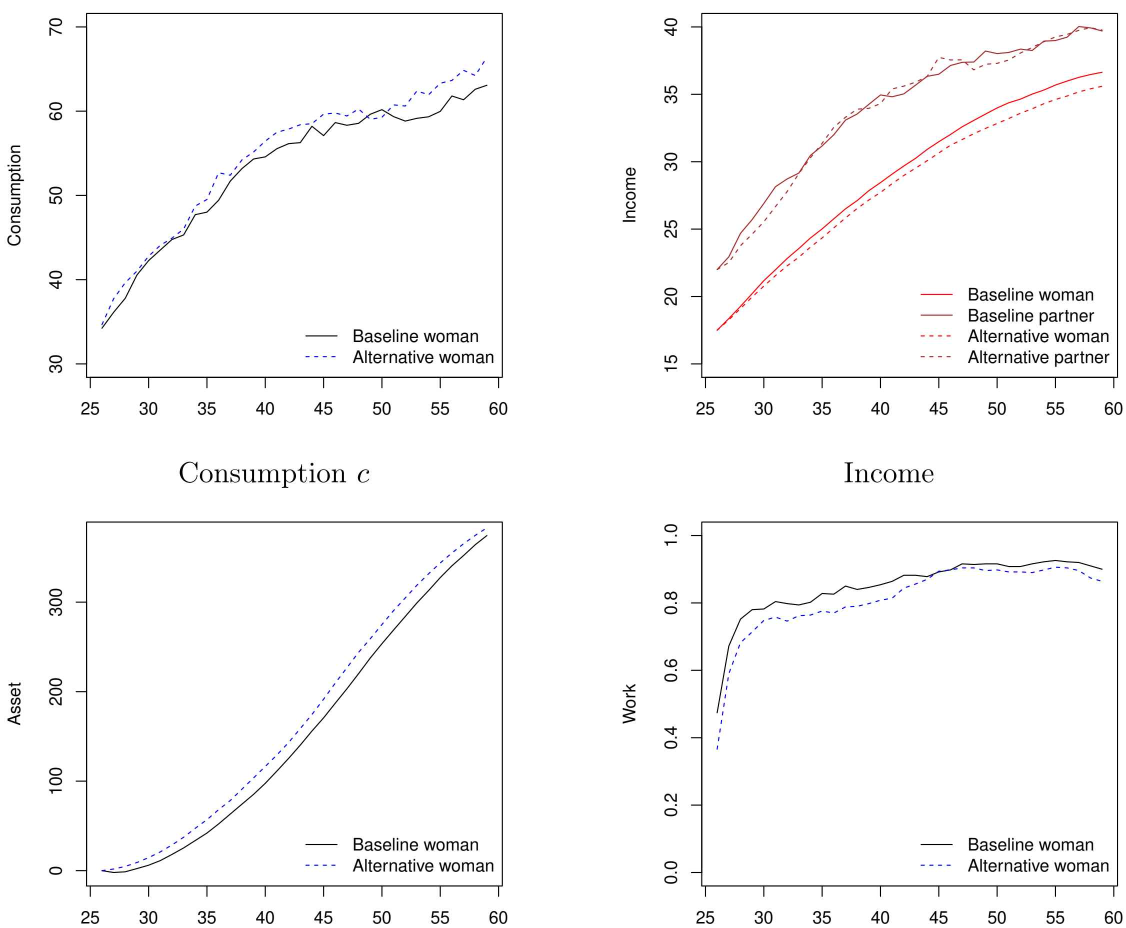
Average evolution of a *baseline woman* with median characteristics: $26y.o.$ woman with income of $17k5$ euros, no assets, in a couple, no children, with a partner earning $22k$ euros.

woman was to work. The substitution effect dominates. Note that even is she is less likely to work, in any case, the higher η_t is, the more she consumes. If she was previously unemployed, the case is more complicated. Up to approximately the median taste shock, the wealth effect dominates, and she will choose to work more to consume more. After this threshold, it decreases, and the more she wants to consume, the less she will work.

Life-cycle simulation:

Figure 1.2 shows the average results over $1,000$ life-cycle simulation for the median 26 year-old woman and for an alternative woman with the same characteristics but two children. First, consumption, income and asset all increase throughout the life cycle. Her partner's income also increases in a similar fashion. Once they enter the labor force, women are increasingly likely to work until retirement. By having two children, the alternative woman is less likely to work initially, and this persists throughout her life cycle. As a consequence, she on average has an income disadvantage, while her husband seems to suffer no particular penalty. However, with two children she will initially obtain more benefits and be able to accumulate slightly more assets. The households with two children consume only slightly more, which suggests that they obtain considerably lower utility from their individual consumption.

Figure 1.2: Life-cycle simulations



Consumption c — Income — Asset — Job tenure d

Average evolution of a *baseline woman*: $26y.o.$ working ($w = 0$) woman with income of $17k5$ euros, no assets, in a couple, no children, with a partner earning $22k$ euros.
And an *alternative woman* with the same characteristics but 2 children.

1.7.4 2nd stage: Structural model Estimation

Estimation

Now, I want to estimate the set of structural parameters: $\theta = (\underbrace{\sigma, \gamma_0^c, s_0^c, \gamma_0^n, s_0^n, \gamma_1^c, s_1^c, \gamma_1^n, s_1^n}_{\equiv \theta_0}, \underbrace{\alpha, \omega}_{\equiv \theta_1})$.

Following section 1.4.2, denote the marginal utilities at optimal CCCs as:

$$u_d'^*(\tilde{x}_t, a_t, \eta_t) = \frac{\partial}{\partial c_{dt}} u_d(c_{dt}, \tilde{x}_t, \eta_t)|_{c_{dt}=c^*_{dt}(\eta_t, \tilde{x}_t, a_t)}.$$

Thus I have the Euler equation for all d:

$$u_d'^*(\tilde{x}_t, a_t, \eta_t, \theta_0) = \beta(1+r_t)\mathbb{E}_t\Big[u_{d_{t+1}}'^*(\tilde{x}_{t+1}, a_{t+1}, \eta_{t+1}, \theta_0)\ \Big| x_t, c_{dt}, d_t\Big].$$

Here, the functional form of $\partial u_d(c_d, \tilde{x}, \eta)/\partial c_d$ is known and depends on the parameter θ_0. Thus, θ_0 are estimated in a first step by minimizing the differences between the two sides of the equation. For the left-hand side of the equation, I use every observation in the data, including the estimated $\hat{\eta}$ as if it was observed and the corresponding observed c_{dt}. For the right-hand side, one can either take the empirical expectation about the future, or simulate it using the estimated CCCs, the CCPs and the transitions. Given the small number of observations I have, I prefer to use the former solution in this application.

The other parameters α and ω (in γ_1) additively enter the utility, so they are not in the Euler equation. They are identified via the CCPs. To recover α and ω, I choose to simulate complete life cycles for each set of parameters θ using the reduced forms. In this way, I obtain estimates of the conditional value functions $v_d()$, and using extreme-value type-1 form of ϵ, I can recover the theoretical $Pr(D_t = 1|\eta_t, X_t, W_t)$ and compare them to the CCPs. The optimal parameters θ minimize these differences. I run the two-stage estimates, so I estimate $\widehat{\theta}_1$ by minimizing the difference in probabilities with respect to θ_1 with θ_0 fixed to the $\widehat{\theta}_0$ estimated in the first stage.

Results

Structural Parameters: (Table 1.3)
I find a coefficient for risk aversion (and the elasticity of intertemporal substitution) similar to the literature: 1.63 versus 1.56 in Blundell et al. (1994), or 1.53 in Alan et al. (2009). It suggests my method yields consistent estimations, with more complex model and faster computation.

As expected, the utility cost of searching for a job when previously unemployed is high. In comparison, the utility cost of working is almost null. For the effect of the taste for consumption, note first that the smaller the coefficient is, the higher the utility since $1 - \sigma < 0$. Thus, note that, for a given consumption level, single working women without children have higher utility (on average) than if they were unemployed ($\gamma_1 = -1.04 < \gamma_0 = 0$). When they are in couples without children, their utility is similar ($-1.04 - 0.65$ versus -1.80). Additional children yield more disutility for employed women ($-0.10 > -0.31$). The variances are similar ($s_1 = 0.54$ close to the fixed

Table 1.3: Structural parameter estimates

	Parameter estimates	
	Parameter	Estimate
Discount factor	β	0.98 (fixed)
Constant Relative Risk Aversion	σ	1.63
Effect of η by family...		
... when unemployed:		
$\mathcal{LN}(\gamma_0^c couple + \gamma_0^n nchild,\ s_0)$	γ_0	0 (fixed)
	γ_0^c	-1.80
	γ_0^n	-0.31
	s_0	0.50 (fixed)
... when employed:	γ_1	-1.04
$\mathcal{LN}(\gamma_1 + \gamma_1^c couple + \gamma_1^n nchild,\ s_1)$	γ_1^c	-0.65
	γ_1^n	-0.10
	s_1	0.54
Additive terms:		
Utility cost of working	α	-0.04
Utility cost of search	ω	-2.14

$s_0 = 0.50$). e

One could use this estimated model to perform counterfactual analysis and study the effect of different labor market reforms on women's consumption and career choice, such as, the effect of increasing the age of retirement or changing the benefits given to single mothers.

1.8 Conclusion

This paper develops a general class of discrete-continuous choice models and provides a list of conditions to achieve non-parametric identification. The identification proof is original as it solves for a unique monotone function instead of proceeding pointwise, which allows identification under weaker relevance conditions than in the existing literature.

Given the identification, I provide a new estimation procedure yielding sizeable computational gains with respect to the existing alternatives for the estimation of dynamic models. The gains are so large that they should facilitate the use of complex dynamic discrete-continuous models in the future and offer greater latitude to researchers to test for several model specifications. This will allow us to find new results in several fields where my methodology can be applied: labor, housing, education, industrial organization, etc.

On a final note, part of the method described here applies more broadly to discrete-continuous dynamic *processes*, choices or not. This yields additional interesting dynamic applications.

Appendix

1.A Proof: Lemma 3

I prove Lemma 3.

Proof. First, let us relate $Pr(d|\eta = h, W = 1) - Pr(d|\eta = h, W = 0)$ to the distributions/quantiles. Recall that using Bayes and $\eta \perp w$, we have $\forall w$:

$$\begin{aligned} h = \quad & Pr(\eta \leq h) \\ = \quad & Pr(\eta \leq h|w) \\ = \quad & Pr(\eta \leq h \mid D = 0, w)Pr(D = 0|w) + Pr(\eta \leq h \mid D = 1, w)Pr(D = 1|w) \\ = \quad & F_{\eta|D=0,w}(h)p_{0|w} + F_{\eta|D=1,w}(h)p_{1|w}. \end{aligned} \tag{1.13}$$

Then, combining (1.13) at $w = 1$ and $w = 0$, we obtain $\forall h$:

$$\begin{aligned} F_{\eta|D=0,W=1}(h)p_{0|1} - F_{\eta|D=0,W=0}(h)p_{0|0} \;&=\; -\Big(F_{\eta|D=1,W=1}(h)p_{1|1} - F_{\eta|D=1,W=0}(h)p_{1|0}\Big) \\ \overset{def}{\Longleftrightarrow} \Delta F_{\eta_0}(h) \;&=\; \Delta F_{\eta_1}(h). \end{aligned} \tag{1.14}$$

Moreover, notice that we can rewrite $F_{\eta_d|w}(h)$:

$$\begin{aligned} F_{\eta|d,w}(h) &= Pr(\eta \leq h|d, w) \\ &= Pr(\eta \leq h,\ d|w)/Pr(d|w) \\ &= Pr(d|\eta \leq h, w)Pr(\eta \leq h|w)/Pr(d|w). \end{aligned} \tag{1.15}$$

Let us focus on the choice $D = 0$ (by symmetry it will be the same for $D = 1$) and rewrite (1.14) by plugging (1.15) into it:

$$\Delta F_{\eta_0}(h) = \big[Pr(D = 0|\eta \le h, W = 1)Pr(\eta \le h|W = 1)/\cancel{Pr(D = 0|W = 1)}\big]\cancel{Pr(D = 0|W = 1)}$$
$$- \big[Pr(D = 0|\eta \le h, W = 0)Pr(\eta \le h|W = 0)/\cancel{Pr(D = 0|W = 0)}\big]\cancel{Pr(D = 0|W = 0)}.$$

Moreover, since $W \perp \eta$: $Pr(\eta \le h|W = 1) = Pr(\eta \le h|W = 0) = Pr(\eta \le h) = h$, we have:

$$\Delta F_{\eta_0}(h) = \big[\ Pr(D = 0|\eta \le h, W = 1) - Pr(D = 0|\eta \le h, W = 0)\ \big]h. \tag{1.16}$$

Now, note that if $\eta \sim \mathcal{U}(0, 1)$:

$$Pr(D = 0|\eta \le h_0, w) = \int_0^{h_0} Pr(D = 0|\eta = h, w)/F(h_0)dh = \int_0^{h_0} Pr(D = 0|\eta = h, w)/h_0 dh. \tag{1.17}$$

Thus, we can rewrite (1.16) $\forall h_0$ as:

$$\begin{aligned}\Delta F_{\eta_0}(h_0) &= \Big[\ \int_0^{h_0} Pr(D = 0|\eta = h, W = 1)/h_0 dh - \int_0^{h_0} Pr(D = 0|\eta = h, W = 0)/h_0 dh\ \Big]h_0\\ &= \int_0^{h_0} \Big(Pr(D = 0|\eta = h, W = 1) - Pr(D = 0|\eta = h, W = 0)\Big)dh. \end{aligned} \tag{1.18}$$

Which leads to:

$$Pr(D = 0|\eta = h, W = 1) - Pr(D = 0|\eta = h, W = 0) = \frac{d\Delta F_{\eta_0}(h)}{dh}.$$

Moreover, if $c_d(h)$ is a strictly monotone solution to our problem, we have $F_{\eta|d,w}(h) = F_{C_d|d,w}(c_d(h))$, and thus $\forall h$:

$$\begin{aligned}\Delta F_{\eta_0}(h) &= F_{\eta|D=0,1}(h)p_{0|1} - F_{\eta|D=0,0}(h)p_{0|0}\\ &= F_{C_0|D=0,W=1}(c_0(h))p_{0|1} - F_{C_0|D=0,W=0}(c_0(h))p_{0|0}\\ &\overset{def}{=} \Delta F_{C_0}(c_0(h)).\end{aligned}$$

Thus:

$$\begin{aligned} Pr(D=0|\eta=h,W=1)-Pr(D=0|\eta=h,W=0) &= \frac{d\Delta F_{\eta_0}(h)}{dh} \\ &= \frac{d\Delta F_{C_0}(c_0(h))}{dh} \\ &= \frac{d\Delta F_{C_0}(c_0)}{dc_0}\underbrace{\frac{dc_0(h)}{dh}}_{>0}. \end{aligned}$$

So, under assumption 9b, there is a finite set of K points $h \in [0,1]$ such that $Pr(D=0|\eta=h,W=1)-Pr(D=0|\eta=h,W=0)=0$. Then there is a finite set of K points $c_0(h_k)$ such that $d\Delta F_{C_0}(c_0(h))/dh=0$. And since $c_0'(h)>0$ because of the monotonicity of the quantiles, it implies that there is a finite set of K points such that $d\Delta F_{C_0}(c_0)/dc_0=0$.

We can follow exactly the same reasoning for $D=1$. We obtain that, if there is a finite set of K points $h \in [0,1]$ such that $Pr(D=1|\eta=h,W=1)-Pr(D=1|\eta=h,W=0)=0$, then there is a finite set of K points $c_1(h_k)$ such that $d\Delta F_{C_1}(c_1(h))/dh=0$. Since $c_1'(h)>0$ because of the monotonicity of the quantiles, it implies that there is a finite set of K points such that $d\Delta F_{C_1}(c_1)/dc_1=0$. ■

1.B Proof of Identification Theorem 1

I develop the complete proof of Theorem 1 about the identification of the continuous choice policies.

Proof. For a given increasing solution $c_d(h)$, let us first introduce the notation:

$$\begin{aligned} p_{d|w} &\equiv Pr(D=d|W=w) \\ \gamma_d(h) &\equiv F_{C_d|D=d,W=0}(c_d(h)) \\ \Psi_{d1}(\gamma_d(h)) &\equiv F_{C_d|D=d,W=1}(c_d(h)) = \underbrace{F_{C_d|D=d,W=1}(F^{-1}_{C_d|D=d,W=0}(}_{\equiv\Psi_{d1}()}\underbrace{F_{C_d|D=d,W=0}(c_d(h))}_{\equiv\gamma_d(h)})). \end{aligned}$$

Recall from Lemma 2 that $F_{C_d|d,w}(c)$: $\mathcal{C}_d \to [0,1]$ is C^1 and strictly increasing function of c. Now, under assumption 3, a solution $c_d(h)$ is also a strictly increasing and C^1 function of h. Thus $\gamma_d(h)$ are C^1 and strictly increasing functions of h, from $\gamma_d(0)=0$ to $\gamma_d(1)=1$. The mappings $\Psi_{d1}(\gamma_d)$

give us the mapping from the quantiles of c_d with instrument value $W = 0$ to their counterpart with instrument value $W = 1$. Similarly, given Lemma 2, we have that Ψ_{d1} are also C^1 and strictly increasing functions of γ_d from $\Psi_{d1}(0) = 0$ to $\Psi_{d1}(1) = 1$. These mappings are directly reconstructed from the data since the data identifies $F_{C_d|d,w}$ $\forall d, w$. So, from data on (c_d, d, w) we now recover: $\forall (d, w)$ $p_{d|w}$ and $\forall d$ $\Psi_{d1}(\gamma_d)$.

Under this reparametrization, the system described in equation (1.2) rewrites, $\forall h$, with increasing $\gamma_d(h)$:

$$\begin{cases} h = \gamma_0(h)p_{0|0} \quad + \quad \gamma_1(h)p_{1|0} \\ h = \Psi_{01}(\gamma_0(h))p_{0|1} \ + \ \Psi_{11}(\gamma_1(h))p_{1|1} \end{cases} . \tag{1.19}$$

The conditional distribution functions $F_{C_d|d,w=0}$ are known, strictly increasing and invertible (Lemma 2). So if there is a unique solution $\{\gamma_d(h)\}_{d\in\{0,1\}}$ to system (1.19), there is a unique solution $\{c_d(h)\}_{d\in\{0,1\}}$ to the original system (1.2). Thus, we first show uniqueness of $\{\gamma_d(h)\}_{d\in\{0,1\}}$ to system (1.19), and then we will come back to $\{c_d(h)\}_{d\in\{0,1\}}$.

Lemma 7 (Identification) *(in the reparametrized problem)*
Under assumption 9, there exists a unique strictly increasing $\gamma_d(h)$ solution to system (1.19) starting from $(\gamma_0(0), \gamma_1(0)) = (0, 0)$ to $(\gamma_0(1), \gamma_1(1)) = (1, 1)$.

Proof of Lemma 7.
We prove that there exists a unique increasing solution to system (1.19).

Existence: *existence* is straightforward. Indeed, we are only focusing on images of the structural model. So by construction, with the true optimal policies, $\gamma_d^*(h) = F_{C_d|D=d,W=0}(c_d^*(h))$ are solutions to the system.

Uniqueness: we need to show this is the *unique* strictly increasing solution to this problem. To do this, we procede in two-steps. First we show that there is a unique increasing mapping between the two conditional quantiles, denoted $\gamma_1^*(\gamma_0)$, which solves the system. Then this mapping yields a unique increasing solutions $\gamma_d^*(h)$. The idea is that, in the end, we want to identify which conditional quantiles γ_0 and γ_1 corresponds to a given h. But to do that, we will first recover the *conditional quantile mapping*, i.e., which γ_1 corresponds to a given quantile γ_0 in choice 0. And then we recover to which h they both corresponds.

<u>*Step 1:*</u> Let us recover the *conditional quantile mapping*: $\tilde{\gamma}_1(\gamma_0)$, i.e., which γ_1 corresponds to a given γ_0, without knowing to which h they correspond. We want to show that there exist a *unique* conditional quantile mapping solution to our problem under assumption 9. First, note that $\gamma_0(h)$ and $\gamma_1(h)$ are C^1 and strictly increasing functions of h. Thus, a higher γ_0 corresponds to a higher h and thus to a higher γ_1. As a consequence, the mapping $\tilde{\gamma}_1(\gamma_0)$ will also be C^1 and strictly increasing function of γ_0 starting from $\tilde{\gamma}_1(0) = 0$ (since $\gamma_1(0) = \gamma_0(0) = 0$) to $\tilde{\gamma}_1(1) = 1$ (since $\gamma_1(1) = \gamma_0(1) = 1$). As a consequence, we need to show that there exists a unique *increasing* mapping $\tilde{\gamma}_1(\gamma_0) : [0, 1] \to [0, 1]$, with $\tilde{\gamma}_1(0) = 0$ and $\tilde{\gamma}_1(1) = 1$, solution to our system of equations (1.19).

We can get restrictions on our mapping using our structural system (1.19). Subtract the first line from the second line in the system of equations (1.19).

$$\begin{aligned} \Psi_{01}(\gamma_0(h))p_{0|1} - \gamma_0(h)p_{0|0} &= -\left(\Psi_{11}(\gamma_1(h))p_{1|1} - \gamma_1(h)p_{1|0} \right) \\ \overset{def}{\Longleftrightarrow} \qquad \Delta F_0(\gamma_0(h)) &= \Delta F_1(\gamma_1(h)) \qquad \forall h \end{aligned} \tag{1.20}$$

Notice that the functions ΔF_0 and ΔF_1 are directly observed from the data as both Ψ_{d1} and $p_{d|w}$ are known $\forall d, w$. We also know they are C^1 functions of γ_d as the sum of C^1 functions (since $\Psi_{d1}(\gamma_d)$ are C^1).

Now, even if we do not observe h, if γ_1 and γ_0 correspond to the same unobserved h, we have: $\Delta F_0(\gamma_0) = \Delta F_1(\gamma_1)$ by equation (1.20). As a consequence, a conditional quantile mapping $\tilde{\gamma}_1(\gamma_0)$ solution to the system (1.19) must solve the equation:

$$\Delta F_0(\gamma_0) = \Delta F_1(\tilde{\gamma}_1(\gamma_0)) \qquad \forall \gamma_0 \in [0, 1] \tag{1.21}$$

Now we show that there exists a unique *increasing* mapping $\tilde{\gamma}_1(\gamma_0) : [0, 1] \to [0, 1]$, with $\tilde{\gamma}_1(0) = 0$ and $\tilde{\gamma}_1(1) = 1$, solution to this equation (1.21) under assumption 9b.

First, let us see the implications of Lemma 3 in our reparametrized problem.

Lemma 3 bis: There is the *same finite number K of values of γ_0 and γ_1* such that

$$\frac{d\Delta F_d(\gamma_d)}{d\gamma_d} = 0 \quad \forall d$$

Proof. This is just a consequence of Lemma 3. First, notice that using our reparametrization:

$$\begin{aligned}
\Delta F_d(\gamma_d(h)) &= \Psi_{d1}(\gamma_d(h))p_{d|1} \; - \; \gamma_d(h)p_{d|0} \\
&= F_{C_d|D=d,W=1}(c_d(h))p_{d|1} - F_{C_d|D=d,W=0}(c_d(h))p_{d|0} \\
&= \Delta F_{C_d}(c_d(h))
\end{aligned}$$

So:

$$\begin{aligned}
&\frac{d\Delta F_{C_d}(c_d(h))}{dh} = \frac{d\Delta F_d(\gamma_d(h))}{dh} \\
\iff &\frac{d\Delta F_{C_d}(c_d)}{dc_d}\,\frac{dc_d(h)}{dh} = \frac{d\Delta F_d(\gamma_d)}{d\gamma_d}\,\frac{d\gamma_d(h)}{dh}
\end{aligned}$$

Now, $dc_d(h)/dh > 0$ and $d\gamma_d(h)/dh > 0$ by strict monotonicity of the solution. So, if $d\Delta F_{C_d}(c_d)/dc_d = 0$ then $d\Delta F_d(\gamma_d)/d\gamma_d = 0$. As a consequence, Lemma 3 implies Lemma 3 bis in our reparametrized problem. ■

<u>Case $K = 0$:</u> in the particular case where $K = 0$, there exists no point such that $\frac{d\Delta F_d(\gamma_d)}{d\gamma_d} = 0 \; \forall d$. ΔF_d are C^1 with no points at which the derivative is zero, so they are monotone and invertible. As a consequence, we can easily recover the unique quantile mapping by inverting ΔF_1 in system (1.21). We have:

$$\tilde{\gamma}_1(\gamma_0) \;\; = \;\; (\Delta F_1)^{-1}(\Delta F_0(\gamma_0)) \quad \forall \gamma_0 \in [0, 1]$$

<u>General case $K > 0$:</u> There is a finite number $K < \infty$ of γ_0 and γ_1 such that $d\Delta F_d(\gamma_d)/d\gamma_d = 0$. Let us denote $\gamma_0^1 < \gamma_0^2 < ... < \gamma_0^K$ the ordered K γ_0 such that $d\Delta F_0(\gamma_0)/d\gamma_0 = 0$. And similarly, denote γ_1^k from $k = 1, ..., K$ the ordered K γ_1 such that $d\Delta F_1(\gamma_1)/d\gamma_1 = 0$.

First, we want to show that if the mapping $\tilde{\gamma_1}(\gamma_0)$ solves the system, then $\gamma_1(\gamma_0^k) = \gamma_1^k$. Let us take the derivative version of the system (1.21):

$$\frac{d\Delta F_0(\gamma_0)}{d\gamma_0} = \frac{d\Delta F_1(\tilde{\gamma_1}(\gamma_0))}{d\gamma_0} = \frac{d\Delta F_1(\gamma_1)}{d\gamma_1}\,\frac{d\tilde{\gamma_1}(\gamma_0)}{d\gamma_0} \qquad \forall \gamma_0$$

Since the mapping $\tilde{\gamma_1}(\gamma_0)$ must be increasing, we have: $d\tilde{\gamma_1}(\gamma_0)/d\gamma_0 > 0$. As a consequence, if a mapping $\tilde{\gamma_1}(\gamma_0)$ is a solution, then when γ_0 is such that $d\Delta F_0(\gamma_0)/d\gamma_0 = 0$, the mapped γ_1 must also have a null derivative, i.e., $d\Delta F_1(\tilde{\gamma_1}(\gamma_0))/d\gamma_1 = 0$. So, in a solution, the K points such that $d\Delta F_0(\gamma_0)/d\gamma_0 = 0$ are mapped to the K points such that $d\Delta F_1(\gamma_1)/d\gamma_1 = 0$.
Moreover, since we are looking for an *increasing* solution mapping $\tilde{\gamma_1}(\gamma_0))$, we necessarily have that these K points are sorted, i.e.:

$$\tilde{\gamma_1}(\gamma_0^k) = \gamma_1^k \qquad \forall k \in \{1, ..., K\}$$

If this was not the case, the $\tilde{\gamma_1}(\gamma_0)$ would not be increasing. Thus we have a unique solution for the K γ_0^k points.

Now, we show that $\tilde{\gamma_1}(\gamma_0)$ is also uniquely defined at other points than the γ_0^k. We use that the function ΔF_d are piecewise monotone and invertible (because C^1 with finite number of points with null derivatives) between the points of null derivative. It is similar to the $K = 0$ case, except that here we can only use piecewise monotonicity and partition the set accordingly.
Formally we procede as follows:

- Split the compact set $[0, 1]$ of γ_0 into $K + 1$ sub-intervals Γ_0^k:
  $\Gamma_0^1 = [0, \gamma_0^1],\ \Gamma_0^2 = [\gamma_0^1, \gamma_0^2],\ \dots\ ,\ \Gamma_0^{K+1} = [\gamma_0^K, 1]$ such that $[0, 1] = \underset{k\in\{1,\dots,K+1\}}{\cup} \Gamma_0^k$
  We denote $\mathcal{S}_0^k$ the image of those subsets by ΔF_0. We have $\Delta F_0 : \Gamma_0^k \to \mathcal{S}_0^k$.

- Do the same with the set of γ_1: split the compact set $[0, 1]$ of γ_0 into $K + 1$ sub-intervals Γ_1^k:
  $\Gamma_1^1 = [0, \gamma_1^1],\ \Gamma_1^2 = [\gamma_1^1, \gamma_1^2],\ \dots\ ,\ \Gamma_1^{K+1} = [\gamma_1^K, 1]$ such that $[0, 1] = \underset{k\in\{1,\dots,K+1\}}{\cup} \Gamma_1^k$.
  We denote $\mathcal{S}_1^k$ the image of those subsets by ΔF_1. We have $\Delta F_1 : \Gamma_1^k \to \mathcal{S}_1^k$.

- Since ΔF_d are C^1, in between the points of null derivative, ΔF_d are strictly monotone and invertible. It implies that $\mathcal{S}_d^k$ are compact sets, as image of compact sets by strictly monotone functions. Moreover, $\Psi_{d1}(0) = 0$ for all d. Thus, $\Delta F_d(0) = 0$ for all d. We also have

$\Psi_{d1}(1) = 1$ for all d. Thus, $\Delta F_0(1) = p_{0|1} - p_{0|0} = (1 - p_{1|1}) - (1 - p_{1|0}) = -(p_{1|1} - p_{1|0}) = \Delta F_1(1)$. Moreover, since we showed that a solution must have $\tilde{\gamma_1}(\gamma_0^k) = \gamma_1^k$, and given that we know a solution exists, then the K points must satisfy our original equation (1.21). Which means that $\Delta F_0(\gamma_0^k) = \Delta F_1(\gamma_1^k)\ \forall k$. It implies that $\mathcal{S}_0^k = \mathcal{S}_1^k$ and we denote them $\mathcal{S}^k$ for all $k \in \{1, ..., K+1\}$. We have: $\mathcal{S}^0 = [0, \Delta F_0(\gamma_0^1)]$, $\mathcal{S}^1 = [\Delta F_0(\gamma_0^1), \Delta F_0(\gamma_0^2)]$, ..., $\mathcal{S}^{K+1} = [\Delta F_0(\gamma_0^K), \Delta F_0(1)]$. Notice that we could have defined the image sets based on ΔF_1 instead of ΔF_0, as $\Delta F_0(\gamma_0^k) = \Delta F_1(\gamma_1^k)\ \forall k$.

Now, we are looking for an increasing mapping solution to the system. By *monotonicity*, we know that for a solution $\gamma_1 : \Gamma_0^k \to \Gamma_1^k$ since the upper bounds (γ_d^k) of these sets are image of each other. On each subintervals Γ_d^k, the corresponding function ΔF_d is strictly monotone and $C^1\ \forall d$. And we have that $\Delta F_0 : \Gamma_0^k \to S^k$ and $\Delta F_1 : \Gamma_1^k \to S^k$. So we can invert it segment by segment and get for any given k:

$$\tilde{\gamma_1}(\gamma_0) = (\Delta F_1)^{-1}(\Delta F_0(\gamma_0)) \quad \forall \gamma_0 \in \Gamma_0^k$$

This uniquely define the solution $\tilde{\gamma_1}(\gamma_0)$ on Γ_0^k.

- We repeat this $\forall k \in \{1, ..., K+1\}$ to obtain a unique mapping $\tilde{\gamma_1}(\gamma_0)$ covering the whole set of γ_0, i.e., $\underset{k\in\{1,...,K+1\}}{\cup} \Gamma_0^k = [0, 1]$.

So, we have a *unique mapping* $\tilde{\gamma_1}(\gamma_0)$ solution to equation (1.21).

<u>*Step 2:*</u> From this unique mapping between the conditional quantiles, we would like to recover the unique quantile functions $\gamma_d(h)$. To recover the functions $\gamma_d(h)$, we just need to use any equations of our original system (1.19) (the first one, for example) to obtain the $h(\gamma_0)$ corresponding to a given $(\gamma_0, \tilde{\gamma_1}(\gamma_0))$ as

$$h(\gamma_0) = \gamma_0 p_{0|0} + \tilde{\gamma_1}(\gamma_0) p_{1|0}$$

So we have a unique increasing solution $(\tilde{\gamma_1}(\gamma_0), h(\gamma_0))\ \forall \gamma_0 \in [0, 1]$. By changing the arguments it means that there exists a unique increasing solution $(\gamma_0(h), \gamma_1(h))$ to the system (1.19), starting from $(\gamma_0(0), \gamma_1(0)) = (0, 0)$ to $(\gamma_0(1), \gamma_1(1)) = (1, 1)$. We denote this unique solution $\gamma_d^*(h)$. Therefore, we proved Lemma 7. ■

Thus, we have a unique increasing solution $\gamma_d^*(h)$ to system (1.19). Now recall that $\gamma_d(h) = F_{C_d|d,w=0}(c_d(h))$. By Lemma 2, $F_{C_d|d,w=0}(c_d)$ are strictly increasing and C^1, thus invertible. As a consequence:

$$c_d^*(h) = F^{-1}_{C_d|d,w=0}(\gamma_d^*(h))$$

So, if there exists a unique set of solution $\{\gamma_d(h)\}_{d\in\{0,1\}}$ to the rewritten system (1.19), there exists a unique set of increasing functions $\{c_d(h)\}_{d\in\{0,1\}}$ solution to the original system (1.2).
This unique set of functions identify the optimal CCCs $c_d^*(h)$ from the data (c_d, d, w). ■

1.C Appendix of Estimator Performance

1.C.1 $T = 1$ Special case

Let us focus on the one-shot decision problem with $T = 1$. This case is interesting because I can obtain closed-form solutions to the problem, and easily compare true Maximum Likelihood Estimator with my estimator. Obviously, because of the existence of this closed-form solution, the time comparison between the methods is irrelevant. But this $T = 1$ example is useful for efficiency comparison with maximum likelihood.

Closed form solution:

The agent works in $t = 1$, retires in $t = 2$ and dies in $t = 3$. The retiree consumes everything she has left, to obtain $a_3 = 0$. Thus the consumption of the retiree is $c_2 = (1+r)a_2 + pension(y_2)$, and is independent from η_2. Moreover, by the budget constraint, $a_2 = (1+r)a_1 + y_1 d_1 - c_{d1} + (1-d_1)b_1$. I set the benefits b_1 equal to 0 in this example. Thus in period $t = 1$, the only period of her working life, conditional on d, the agent solves:

$$\begin{aligned}\underset{c_{d1}}{max} \quad & \frac{c_{d1}^{1-\sigma}}{1-\sigma}\tilde{\eta}^d(\eta_1,\gamma_d,s_d) + \alpha d_1 + \omega(1-w)d_1 + \epsilon_{d1} \\ & + \beta\mathbb{E}\Big[\frac{((1+r)^2 a_1 + (1+r)y_1 d_1 - (1+r)c_{d1} + pension(y_2))^{1-\sigma}}{1-\sigma}\ \tilde{\eta}^d(\eta_2,\gamma_0,s_0)\Big].\end{aligned}$$

where $\mathbb{E}[\tilde{\eta}^d(\eta_2,\gamma_0,s_0)] = e^{\gamma_0+s_0^2/2}$, also there is no ϵ_2 shock in the retirement period, and the retirement utility is the same as the utility when unemployed.

It yields the closed form solution for the conditional consumption in $t=1$:

$$c_{d1} = \frac{1}{(1+r) + \Big(\beta(1+r)e^{\gamma_0+s_0^2/2}/\tilde{\eta}^d(\eta_1,\gamma_d,s_d)\Big)^{1/\sigma}}\Big((1+r)^2 a + (1+r)y_1 d + pension(y_2)\Big).$$

The agent consumes a share of available income which depends on the decision. Since the retiring utility is the same as the unemployed one, I only identify γ_1, s_1 with respect to γ_0, s_0 and β.[20] Thus, the parameters to estimate are $\theta = (\sigma,\gamma_1,s_1,\alpha,\omega)$. β is fixed at 0.98, $\gamma_0 = 0$, $s_0 = 0.25$.

Results: (Table 1.C.1)

First, the speed comparison is irrelevant here. Indeed with one period one does not need to solve for the value function so it is considerably easier, especially since we also have closed-form solutions to simulate the model and compute the likelihood. On average here SMM already takes longer than my DCC method but only because it requires to test more set of parameters to find the optimum, as the objective are different between the two functions. It could be the reverse, and one could expect both methods to go at relatively similar speed when $T = 1$ in general. The real benefits of my method are when it allows to avoid solving for the value function, i.e., as soon as $T > 1$.

Concerning the statistical efficiency, as expected when you have a closed form solution for the likelihood, MLE is always the most efficient. It is also the quickest as I do not need to estimate any reduced form in a first stage and I am using a known closed form solution in this $T = 1$ case. Obviously once I go to more period, MLE becomes the longest method, and is becoming untractable.

My method (DCC) is consistent and relatively efficient, but less than the MLE benchmark.

[20]Notice that if I had a specific retirement utility, different from the unemployment one, I could also identify parameters of unemployed with respect to retirees.

Table 1.C.1: Comparison of the estimators when $T = 1$

	Truth	*Method*					
		DCC		*MLE*		*SMM*	
		1,000	10,000	1,000	10,000	1,000	10,000
σ	1.60	1.5806 (0.1759)	1.5782 (0.0827)	1.6042 (0.0444)	1.5992 (0.0137)	1.6135 (0.0560)	1.5970 (0.0211)
γ_1	0.00	0.0071 (0.0714)	0.0040 (0.0286)	-0.0061 (0.0205)	0.0007 (0.0072)	-0.0269 (0.0213)	-0.0009 (0.0078)
s_1	0.40	0.4246 (0.0747)	0.4043 (0.0366)	0.4005 (0.0187)	0.4001 (0.0060)	0.3926 (0.0245)	0.3857 (0.0073)
α	-0.50	-0.4782 (0.3266)	-0.5092 (0.1016)	-0.4928 (0.0852)	-0.5000 (0.0268)	-0.4986 (0.0989)	-0.4850 (0.0401)
ω	-1.00	-1.0689 (0.1715)	-1.0044 (0.0484)	-1.0115 (0.1577)	-0.9931 (0.0441)	-1.0308 (0.2919)	-1.0029 (0.0665)
Avg Time taken:		16s	32s	1s	9s	16s	50s

Other initializations: Number of Monte-Carlo $= 1,000$.
$Pr(w_1 = 1) = 0.7$. y_H is set to 20 and in this case I impose $Pr(y = y_H) = 1$.
$a_1 = 12.5$ for everyone here. Benefits $b = 0$. Pension Percentage of income = 50%.
Fixed parameters: $\gamma_0 = 0$, $s_0 = 0.25$, $\beta = 0.98$, $r = 0.05$.

Simulated Method of Moments (SMM) with moments drawn from the reduced forms conditional distributions of $c_d|d, w, x$ and conditional probabilities $p_{d|w,x}$ is also consistent. It is also more efficient (except for the additive parameter ω). I lose efficiency because of the two-step nature of my method, since I'm always computing the second step using first step optimal choices estimates. But overall, the efficiency loss is largely compensated by time gains in more complicated models.

Chapter 2

Imperfect Information, Learning and Housing Market Dynamic

Christophe Bruneel-Zupanc[1]

Abstract

This paper examines the decision problem of a homeowner who maximizes her expected profit from the sale of her property when market conditions are uncertain. Using a large dataset of real estate transactions in Pennsylvania between 2011 and 2014, I verify several stylized facts about the housing market. I develop a dynamic search model of the home-selling problem in which the homeowner learns about demand in a Bayesian way. I estimate the model and find that learning, especially the downward adjustment of the beliefs of sellers facing low demand, explains some of the key features of the housing data, such as the decreasing list price overtime and time on the market. By comparing with a perfect information benchmark, I derive an unexpected result: the value of information is not always positive. Indeed, an imperfectly informed seller facing low demand can obtain a better outcome than her perfectly informed counterpart thanks to a delusively stronger bargaining position.

[1]E-mail address: christophe.bruneel@tse-fr.eu.

JEL classification: D83, R2, R3.

Keywords: housing, pricing, imperfect information, Bayesian learning.

2.1 Introduction

Real estate transactions involve large financial amounts ($206k, on average, in Pennsylvania between 2011 and 2014) and take time (109 days). A lengthy or negative outcome of the home-selling process can make a substantial difference to the seller's well-being, and making a successful sale can be challenging. Contrary to many other markets, the homeowner does not post a take-it or leave-it price. Instead, she posts a list price which impacts transaction outcomes less directly (Han and Strange, 2016). Indeed, even though the list price influences negotiations between sellers and buyers, the transaction frequently occurs at a different sale price (85% of the time in my sample). Of utmost importance to the seller is that a higher list price should yield a higher sale price but at the expense of a longer time on the market (Miller and Sklarz, 1987). This trade-off is complex, especially because the seller has imperfect information about the demand (Salant, 1991). Indeed, houses are highly differentiated assets and the seller is often unable to observe more than a few recent transactions of similar properties nearby. This lack of information impacts the seller's welfare (Anenberg, 2016).

In this article, I investigate the home-selling decision of imperfectly informed sellers. I build a single-agent dynamic search model of the housing market in which a rational seller is uncertain and learns about demand in a Bayesian way. I estimate the model using an original dataset of real estate transactions in Pennsylvania between 2011 and 2014. The estimated model yields insights on the optimal list pricing strategies and how information frictions affect them. In particular, I estimate the cost of uncertainty for the seller.

The literature establishes several stylized facts about the home-selling problem, and my model aims to explain three of them. First, we observe that transactions are occurring at a price mostly below, but also sometimes above or exactly equal to, the list price (Merlo and Ortalo-Magne, 2004). Second, the housing market is illiquid. Third, the list prices are duration-dependent and adjust downwards throughout the listing process, even when market conditions seem stable (Salant, 1991).

My main contribution is to make sense of the two first facts by modeling the number of buyers

on the market explicitly, independently from their valuations for the house.
First, there can be several buyers on the market. This allows me to explain the relationship between the list and the sale prices. Indeed, as explained earlier, the sale price is determined after bargaining between the seller and the buyer(s). In these negotiations, the list price serves as a ceiling: the seller commits to accept any offer equal to or above it (Horowitz, 1992; Chen and Rosenthal, 1996a,b). Thus, if there is only one buyer for the house, he will never make an offer greater than the list price. Sales will occur either below, or exactly at, the list price (Arnold, 1999), but never above it. I endogenize such above list prices by introducing competition between several buyers, as in Han and Strange (2016). More precisely, I derive the solution to a sequential bargaining game à la Rubinstein (1982) with several buyers and one seller who commits to accepting offers higher than her list price. In this setup, the seller can obtain a price above her list price. Indeed, when she bargains with the highest valuation buyer, the second-highest buyer's valuation serves as an outside option for the seller (Shaked and Sutton, 1984). This second-highest valuation can be above the list price, which sometimes allows the seller to obtain a price above the ceiling set by the list price. Through this mechanism, modeling the number of buyers allows us to endogenize sales above the list price in order to gain a better understanding of the phenomenon. This is especially relevant since the proportion of sales greater than the list price has been non-negligible in recent years: from only 4% in the mid-1990s (Han and Strange, 2014) to 15% in the US during the 2000s boom, 10% after the bust (Han and Strange, 2016) and 11.42% in my sample.

Second, modeling the number of buyers in addition to their valuations helps to understand the market's illiquidity and how much of it is caused by the list price. Indeed, the list price serves as a signal to attract buyers. The higher this price, the more buyers expect to pay and the less likely they are to visit the house and bargain to buy it. If we observe no transaction, it is either because there are no buyers on the market, or because the list price is too high and repels buyers. My specification of the demand helps to disentangle the 'list price induced illiquidity' (low buyers valuations) from demand illiquidity (no buyer on the market). For example, even with an extremely low list price (such that even low valuation individuals would be willing to buy), the seller is not guaranteed to sell her home, as there may simply be no buyer on the market during this period. My model explains this kind of illiquid demand situation observed in the data. Moreover, using data on time on the market, list and sale prices, I can identify the valuation process independently from the market thinness.

The third key fact is that the list price is non-stationary. To model this, I assume that uncertain sellers dynamically learn about the demand. The gradual acquisition of information about an uncertain demand can explain why a list price varies, even if market conditions are unchanged (Anenberg, 2016). More precisely, the seller is uncertain about the liquidity process (the number of buyers on the market) and progressively learns about it by observing visits. Recall that a buyer only visits the house if his valuation is high enough with respect to the signal given by the list price. Thus, by only observing visits, the seller does not necessarily observe all of the buyers on the market. If she sets a low list price all buyers will visit and she directly observes the number of buyers on the market. Otherwise, by setting a higher list price she may select some buyers, and learn less about the true number of buyers on the market. In this original application of Bayesian learning, the decision variable (list price) influences the informational flow (by influencing the odds of observable entries) and thus the learning pace. In addition to the usual trade-off between high price and short time on the market, the list price also embeds a learning externality here: the lower the list price, the faster you learn.

Estimating the parameters of such a model in which optimal strategies are time-dependent (non-stationary) requires detailed microdata. I use an original dataset of approximately 100,000 complete listing histories (dated initial list price and revisions up to the final sale) of sold (with the help of a realtor) single-family homes in Pennsylvania between 2011 and 2014. I collected the data from the American real estate website, *zillow.com*. I study the home-selling problem in this new context (Pennsylvania 2011-2014) and observe similar stylized facts as the one described previously (see Knight, 2002; Merlo and Ortalo-Magne, 2004; Han and Strange, 2014; Anenberg, 2016).

In terms of results, my learning model (seven structural parameters only) closely matches the data. I find that progressive learning (in particular the downward adjustment of the initial rational beliefs of sellers facing low demand) is key to explaining the observed decline of the list price. The model also reproduces the distribution of time spent on market.

Finally, this paper contributes to the literature on the role of overconfidence (Odean, 1998; Piazzesi and Schneider, 2009). I estimate the cost of uncertainty or value of information. To do so, I simulate the model and compare it to a perfect information benchmark. Counterintuitively, I find that being misinformed is not necessarily bad for the seller, at least for sellers facing low demand who are thus initially 'overconfident'. For them, the value of additional information may even be negative, as an overconfident seller may obtain a further discounted sale price than her perfectly informed counterpart. Indeed, by being wrongfully overconfident, she overestimates her reservation value. She has a genuinely stronger position in the bargaining game, resulting in a higher sale price. However, she also refuses offers more easily and sets higher list prices, implying fewer visits, leading to a longer time spent on the market because of the overestimation. There is an 'overconfidence area', where the gain in sale price offsets the overly long time spent on the market, resulting in a better outcome. Thus, sellers facing low demand can be better off by not knowing this and starting with imperfect information rather than being perfectly informed.

On the other hand, sellers facing high demand suffer twice from the imperfect information: they do not pick the list price that maximizes their true outcome, and they have a weaker bargaining position.

Related Literature: The role of the asking price has been extensively studied in the literature (see Han and Strange, 2015, for a complete survey).

First, the literature establishes some key stylized facts about the relationship between sale and list prices. On average, the ratio of sale to list price is around 96% (Case and Shiller, 1988, 2003; Merlo and Ortalo-Magne, 2004; Han and Strange, 2014). Sales frequently occur below the asking price: generally close to 75% of the recent sales in the US (Carrillo, 2013; Anenberg, 2016). Sometimes transactions occur at a price greater than the list price. As mentioned, this was very uncommon in the 1990s: around 4% of all transactions in England (Merlo and Ortalo-Magne, 2004) and in the US (Han and Strange, 2014). However, this has recently become more frequent (15% during the 2000s boom, and is currently around 10%).

There also exists a mass point of sales (from 10% to 25%) occurring exactly at the list price. To explain this, most housing search models treat the list price as a binding ceiling (Horowitz, 1992; Chen and Rosenthal, 1996a,b).

Yet, as already mentioned, these models cannot explain why some sales occur above the list price. Han and Strange (2016) fill the gap by modeling a bargaining search model between one seller and several buyers. As in my model, competition can generate prices above the list price. In fact, their bargaining rule is a particular distinction of this study: I allow for any valuations (for the seller and buyers) while they specify a discrete distribution with only two types (high and low valuations buyers). They picked this bargaining rule based on common sense to fit the data. I go further and show theoretically that it is the solution to a sequential bargaining problem à la Rubinstein (1982).

The impact of the list price on time on the market has also been studied. A lower list price increases the probability of visits and sale of the house (Salant, 1991; Horowitz, 1992; Carrillo, 2012). Similarly, within a listing process, a downward adjustment of the list price increases the probability of sale (de Wit and van der Klaauw, 2013). Using unique survey data about a buyer's search behavior, Han and Strange (2016) show that the list price directs the buyer in his search.

In addition, sellers of atypical houses are more likely to spend a longer time on the market (Haurin, 1988). This is consistent with the imperfect information story developed in my paper. Indeed, sellers of atypical houses observe fewer past transactions of houses similar to theirs, thus, they are less informed about the demand they face than their neighbors are. This may explain the difference in behavior: as shown in reduced form evidence by Anenberg (2016), information frictions impact the seller's decision.

Fewer studies have focused on the dynamics of list prices. In order to address this, it is necessary to build a model with some time-dependence. Only two recent papers have done this, to date: Merlo et al. (2015) and Anenberg (2016).

Merlo et al. (2015) focus on dynamics, with their main objective to explain list price stickiness: 77.3% of the sales occur without any list price adjustment, and 20.8% with only one adjustment. They show that an extremely small menu cost can generate the observed stickiness. Because it is already well explained in their paper, I abstract from price stickiness in my model. To make sense of the optimal decline in list price, they estimate a model with a rich time-dependent arrival probability function. Overall, they fit their data very closely at the expense of a heavily parametrized model, while I get a decent fit using only seven structural parameters by introducing learning and imperfect information in the model.

Anenberg (2016) is the closest paper to mine. He builds a home-selling model with imperfect information and learning. He formulates and estimates a model in which the seller is uncertain about the buyer's valuation for her property. However, she has a prior about the mean of the valuations distribution. In this context, as in my work, the gradual acquisition of information by the seller can explain the time-varying list price choices (which declines, but also rarely increases, for example). Because of information frictions, a short-run aggregate price may take a longer time to adjust.

The main contrast with this paper is that I model the number of buyers explicitly, and not only their valuations for the house. This allows the study of some aspects of the home-selling problems, which were not the focus of Anenberg's paper. First, I can give a micro-foundation to the bargaining side of the problem. I endogenously generate sales above the list prices, while models with only valuations cannot. Anenberg (2016) generates them with an exogenous probability, for instance. Given the important share of sales above the list price (more than 10% of sales in the US after 2010), being able to endogenize and explain them is crucial: even more so because of their pro-cyclicality (Liu et al., 2014). Moreover, my micro-founded bargaining rule also allows me to derive the paradoxical result that uncertainty can be beneficial to the seller. Finally, using data on list prices, sale prices and time on the market, I can separately identify low valuations (list-price induced illiquidity) from illiquidity (no buyer).

This paper proceeds as follows. Section 2 describes the data. Section 3 develops a dynamic micro-search model of the home-selling problem in which sellers learn. Section 4 details the estimation methods and Section 5 presents the estimation results. Section 6 uses the estimated model to analyze the value of information. I conclude the paper in Section 7.

2.2 Data

2.2.1 Source

My data contain transaction records of properties sold between August 2011 and July 2014 in Pennsylvania, gathered from the real estate platform, *zillow.com*. The website is one of the leading online real estate marketplace sites. It is a real estate listing aggregator which gathers listings from real estate agents, in partnership with MLS services and from private national companies (Century 21, Coldwell Banker and Sotheby's, for example). Consequently, this is a collection of some of the most exhaustive data about real estate transactions in Pennsylvania.

I focus on the sales of single-family homes for which there is a complete history record of the transaction available. For each of these sales, the data include the usual property attributes (square footage, lot size, number of beds, number of baths, year when the house was built, etc.). For approximately 33% of all transactions, a complete record of the last transaction history is available.

As displayed in Figure 2.2.1, this history contains all of the seller's decisions: the initial list price, its eventual adjustments through the sale process, potential intermediary listing removals, and the final sale price.

Figure 2.2.1: Example of transaction history record

DATE	EVENT	PRICE	$/SQFT	SOURCE
11/20/13	Sold	$328,000 -0.6%	$140	Public Record
11/18/13	Listing removed	$329,900	$141	BHHS Fox & Roa...
10/02/13	Price change	$329,900 -5.7%	$141	Prudential Fox...
08/29/13	Price change	$349,900 -5.2%	$150	Prudential Fox...
08/06/13	Price change	$369,000 -2.4%	$158	Prudential Fox...
07/12/13	Price change	$378,000 -5.5%	$162	Prudential Fox...
06/08/13	Listed for sale	$399,900	$171	Prudential Fox...

This property was sold after five months on market in 2013, after four list price adjustments. It was finally sold at a price of $\$328,000$, slightly below the final list price of $\$329,900$.

In order to study list price dynamics and time on the market, I focus only on transactions for which the history is available. It yields a standard selection bias. Indeed, the detailed history is available for properties of better quality, resulting in a significantly higher sale price than the complete pool of transactions on average. However, this selection bias is present in most data used in the reference literature, which are obtained from real estate agencies. In theory, zillow.com (hereafter, Zillow) allows homeowners to list their house on their own ('for sale by owner'). Unfortunately, these listings represent less than 1% of the observations. As a consequence, these data cannot be used to understand the seller's choice to resort (or not) to using a real estate agent (as modeled by Salant (1991)), and I focus only on transactions in which a real estate agent has intervened (as in Merlo et al. (2015) or Anenberg (2016)).

Zillow's data exhibit two main flaws. First, the data are right-censored: I only observe transactions which ended up in a success (sale). I do not observe properties that are still on the market or were withdrawn by the sellers (except if they were relisted and sold afterwards). In the transaction history of sold properties, I know whether or not the seller previously decided to pull her property off the market (before relisting it for the final sale that I observe). As I observe the complete sale history, I know the price choices/adjustments made before a potential withdrawal. A total of 17% of the final sales have been withdrawn from the market and relisted later before being sold.[2] One could imagine that a seller who has already tried to sell her property (and failed) differs from a 'new' seller. In particular, these two types of sellers should differ in terms of the information they have about the market conditions (one of the two having accumulated information about the market demand with her failed listing). In order to avoid these differences, and since I do not focus on the withdrawal decision here, I drop properties which were withdrawn at least once from my sample and focus only on 'first time' sellers.

Second, as in most of the real estate data, I lack information about buyers' offers and the eventual rejection of these offers by the sellers. Thus, I am forced to keep the buyer's side of my model relatively simple (contrary to Merlo et al. (2015) who have information about buyers' bids and can use it to model their bargaining process, for example).

After selecting the observations and cleaning the data, I have 97,451 real estate transactions in Pennsylvania between August 2011 and July 2014. The data are described in the next section.

[2] This implies that the proportion of sellers who fail to sell their property is larger than 17% overall, since I only observe a part of the withdrawals: the properties that were relisted and sold afterwards.

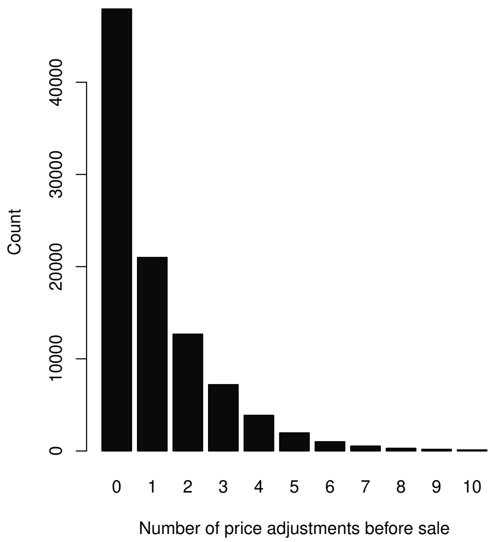
2.2.2 Summary statistics

Table 2.2.1: Summary Statistics

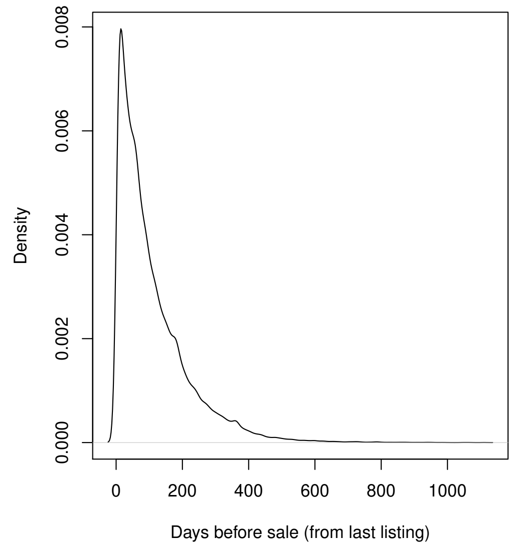
Variable	Mean	Std. Dev.	Min	Median	Max
Price and Timing:					
Price	206512	122209	15000	179000	745000
Days before sale (from last listing)	109	108	0	76	1112
Days before sale (from first listing)	118	110	7	84	1141
Number of adjustment before sale	1.12	1.56	0	1	10
Proportion of listing with adjustments	0.5046				
Ratio sale/Final list price	0.9546	0.0847	0.1316	0.9679	4.3337
Ratio sale/Initial list price	0.9121	0.1089	0.1316	0.9365	4.3337
Proportion of sale price > Final list price	0.1142				
Proportion of sale price = Final list price	0.1495				
Properties characteristics:					
Living area (sqft)	1850	711	780	1682	4572
Number of beds	3.34	0.74	1	3	7
Number of baths	2.1	0.81	0.5	2	5.5
Number of transactions	97451				
Number of census tracts	1806				

Table 2.2.1 and Figure 2.2.2 present the summary statistics of my sample. The Pennsylvanian data exhibit common stylized facts to those observed in the literature and which have been extensively detailed by Merlo and Ortalo-Magne (2004). First, prices are sticky and often not adjusted (see Figure 2.2.2a). However, in my sample they are not as 'sticky' as usual: only 50% of the sellers sell their properties without changing their list price at least once. This percentage is generally closer to 75%: equal to 76.79% in Merlo and Ortalo-Magne (2004) in the UK for example, and also closer to 75% in another Zillow data set of properties on the East Coast of the United States. I choose not to focus on the stickiness and not to model it. It has already been well explained by Merlo et al. (2015) and including a menu cost in my model would be too computationally costly (I would need to add the previous period list price as a state variable) and provide no interesting new insight.
The average time on the market is about 16 weeks (Figure 2.2.2b). This is within the usual range: higher than that observed by Merlo et al. (2015) in the UK (about 10 weeks), and slightly lower than that observed by Anenberg (2016) in San Francisco and Los Angeles (about 18 weeks on market).

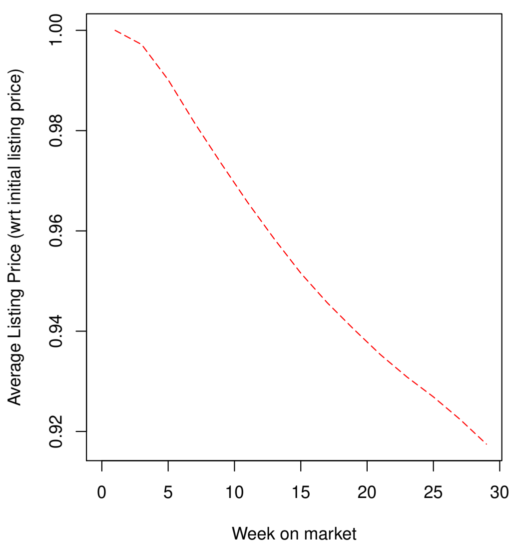
Figure 2.2.2: Descriptive Statistics



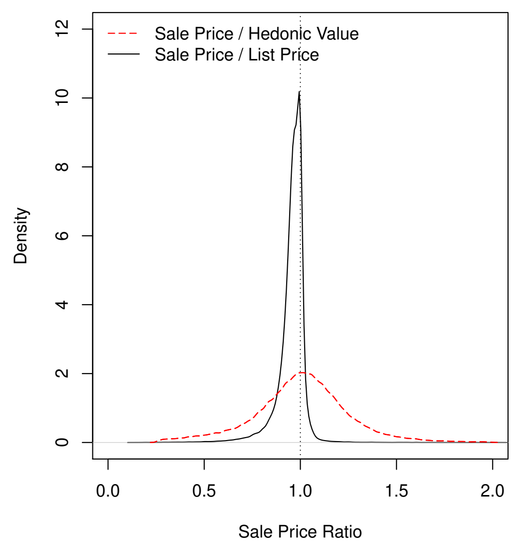
(a) Distribution of the number of price adjustments before sale



(b) time on the market distribution



(c) List price dynamics
(as ratio of initial list price)



(d) Ratio Sale to list price distribution

The list price is decreasing through the selling process: minus 8% between the first and the 30^{th} week (Figure 2.2.2c). Most sales happen below the list price (73.63%), many occur exactly at the list price (14.95%) and the 11.42% remaining occur above it. This number of sales above the list price is considerably greater than in the English data (3.9% in Merlo et al., 2015): this is why my model emphasizes the endogenization of sales above the list price.

Unobserved private information: Figure 2.2.2d shows the distribution of sale prices normalized either by the final list prices or by the predicted prices (hedonic values). The distribution of prices normalized by the hedonic values of the property is less concentrated than the one normalized by list prices. The list price contains extra private information about the property value, which is not contained in the explanatory variables used in the hedonic estimation (number of bathrooms, number of beds, living area, census tracts, etc.). I find that the list price is a better predictor of the sale price than the hedonic fitted price, as is well known in the literature (Horowitz, 1992; Merlo et al., 2015). To soften the impact of this private information unobserved by the econometrician, I use mostly 'relative moments' normalized by the list price when I estimate the model, for example, sale price distribution relative to the final list price, or list price dynamics relative to initial list price.

2.3 Model

I model a discrete-time (with periods of two weeks), infinite horizon problem of a homeowner deciding to sell her property with the objective of maximizing the sale price. As in Anenberg (2016), I introduce uncertainty and Bayesian learning into this framework. To explain the previously described stylized facts, one of the main features of the model is that the seller is imperfectly informed about the true demand. More precisely, she has rational expectations about the distribution of the number of buyers in the population, but she does not know the distribution of buyers interested in her house specifically for each period, and she is learning about it progressively.

Each period t, given her information set, the seller picks a list price p_t^L in order to maximize her expected gains from the sale. The chosen optimal list price will balance a classic trade-off between high sale price and short time on the market. To this classic trade-off, the model adds a learning externality to the list price decision: ceteris paribus, a lower list price choice allows the seller to learn more quickly about the market conditions.

In addition to her dynamic list price decision, the seller also decides whether or not to accept an offer, knowing that if she refuses (or if she receives no offer), she will incur a holding cost of keeping her house for sale one more period, modeled as a discount factor δ.

In what follows, I first describe the within-period game: the explicit bargaining rules defining the sale price, the demand modelization and the seller's learning process. Next, I focus on the seller's dynamic optimization problem specification.

2.3.1 Bargaining rules

The sale price is determined after a sequential bargaining game (Rubinstein, 1982) of offers/counteroffers with complete information between one seller (valuation v^s) and n inspecting/visiting buyer(s) (with ordered valuations $v^b_{(1)} < v^b_{(2)} < ... < v^b_{(n)}$). The list price p^L serves as a commitment device in the model: if a buyer makes an offer greater or equal to it, the seller has to accept it and sell him the property. This bargaining game between n buyers and one seller with complete information has a unique subgame-perfect equilibrium outcome (see Appendix 2.A) which is my *bargaining rule*:

- if $v^s > v^b_{(n)}$ (or if $n = 0$): no sale.
- Otherwise, gains from trade exist with at least one buyer, thus under complete information the seller will sell her house to the highest valuation buyer at the price p^S, where

$$p^S = max\Big\{v^b_{(n-1)}\ ,\ min\Big(p^L\ ,\ v^s + \frac{1}{2}(v^b_{(n)} - v^s)\Big)\Big\}$$

  If $n = 1$, remove the $v^b_{(n-1)}$ part (or consider it $= 0$).

In words, trade only occurs if there are gains from trade ($v^b_{(n)} \geq v^s$). If this is the case, under complete information the seller will sell her property to the highest valuation buyer (buyer$_{(n)}$) from whom she can extract the higher sale price.

Without the presence of other buyers and without a list price, bilateral sequential bargaining between the seller and buyer$_{(n)}$ would yield the classic Rubinstein (1982) outcome whereby they share the 'transaction gains'. The seller gets a portion ϕ of $(v^b_{(n)} - v^s)$ while the buyer gets the rest (the portion $1 - \phi$). Thus the sale price would be $p^S = v^s + \phi(v^b_{(n)} - v^s)$. For simplicity, I make assumptions such that they share the transaction gains equally ($\phi = 0.5$).[3]

Moreover, the seller has to accept any offer higher or equal to the list price. Thus, if the bilateral bargaining price is greater than p^L, this gives an opportunity for the buyer to make a lower offer, equal to p^L, that the seller must accept. As a consequence, bilateral bargaining between the seller and buyer$_{(n)}$ with a list price would yield the sale price $p^S = min\Big(p^L\ ,\ v^s + \frac{1}{2}(v^b_{(n)} - v^s)\Big)$.

Now, if another buyer (the second-highest valuation buyer, denoted buyer$_{(n-1)}$) is present on the

[3] To end up with an equal sharing of the pie as in Rubinstein (1982) sequential bargaining, I assume either that the bargaining period length tends to zero or that all the agents are equally impatient with impatience factor tending to one (see Appendix 2.A).
An alternative to the equal share would be to give all the bargaining power to the seller (i.e. having her infinitely more patient than the buyer), as done implicitly in Anenberg (2016).

market, his valuation $v^b_{(n-1)}$ can serve as an outside option for the seller.[4] If $v^b_{(n-1)}$ is high enough (greater than the bilateral bargaining with list price outcome), the seller can threaten buyer$_{(n)}$ to sell her property to buyer$_{(n-1)}$ instead. Competition between the two buyers forces buyer$_{(n)}$ to offer at least $p^S = v^b_{(n-1)}$ in order to ensure that the seller sells him the property. Adding this competitive outside option to the bilateral bargaining of the seller with only buyer$_{(n)}$ yields the general sale price formula.

The main advantage of this bargaining rule is that it can endogenously generate prices below the list price ($v^s + 1/2(v^b_{(n)} - v^s)$), at the list price ($p^L$), and above it (in case of competition, with $v^b_{(n-1)}$ which can be higher than p^L sometimes). The challenge here is to reproduce it in proportions comparable to the data (15% equal to p^L, 11.5% above and the rest below).

The bargaining rule is known by the seller (and the buyers). This implies that, though the seller does not know the exact buyers' valuations before meeting them, knowing the bargaining rule allows her to compute what the hypothetical bargaining outcomes may be (sale or not, and eventual sale price) for any scenario. Therefore, for a given number of buyers, if she knows the buyers' valuations distribution, the seller is able to build an expectation about the bargaining outcomes (using order statistics of the highest and second-highest valuation for the given number of buyers). If she also has an idea of the distribution of the number of buyers on the market, she can compute her expected profit from sale as a function of the list price. Then she can pick the list price to maximize it: this is what the seller does in this model.
The demand side of the model is detailed in the next section.

2.3.2 Buyer

The demand side is split in two main components: the number of buyers on the market during the period ($N_t^{market} \sim Poisson(\lambda)$) and the valuations of each of these buyers for the given property ($v^b \sim \mathcal{LN}(\mu, \sigma)$).

For each property s and for each period t of two weeks, the number of buyers on the market potentially interested in the property (N_t^{market}) is drawn from a $Poisson(\lambda_s)$ distribution. The rate of arrival λ is a key demand parameter in the model. Each seller faces a specific λ_s drawn from the true distribution $Gamma(\alpha_0, \beta_0)$. They have rational expectation and know that $\lambda_s \sim$

[4]Notice that only the second-highest valuation can impact the sale price: ceteris paribus the other buyers' valuations are irrelevant.

$Gamma(\alpha_0, \beta_0)$, but they do not know the value of their individual draw λ_s. With the information they obtain through their own listings, the sellers progressively update their initial rational belief about λ_s using Bayesian learning.

The reservation values of every buyer on the market for a given property are defined as follows:

$$
\begin{aligned}
& V^b = \eta_s \ exp(\theta_b) \\
\iff & v^b := \frac{V^b}{\eta_s} = exp(\theta_b) \text{ with } \theta_b \sim \mathcal{N}(\mu, \sigma) \\
\iff & v^b \sim \mathcal{LN}(\mu, \sigma)
\end{aligned}
$$

where θ_b represents the buyer specific taste for a given property and η_s represents the property intrinsic/predicted value (estimated via hedonic regression). I assume that the buyer knows his taste θ_b for the property, as well as the property intrinsic value η_s. Thus, he also knows his reservation value (V^b) and his reservation value normalized by the hedonic value (v^b). I focus on normalized values (v^b) rather than real monetary values (V^b) in order to compare any type of homes on the same scale. By doing this, I implicitly assume a linear homogeneity of the home-selling problem between the different properties, as in Merlo et al. (2015). In other words, I assume that a $\$10,000$ deviation for a property worth $\$100,000$ is perceived similarly to a $\$20,000$ deviation for a property worth $\$200,000$. In this way, I build a single representative problem for every seller, independently of the 'quality' of their properties. Then, under the linear homogeneity assumption, I can compare data counterparts to the model price outcomes; that is, the data prices normalized by their hedonic values.

The seller knows the buyer's reservation value distribution $v^b \sim \mathcal{LN}(\mu, \sigma)$ (but she does not know the buyer's exact taste shock realization before entering into bargaining with him). Thus, to build expectations about the demand at the start of each period, her only unknown demand parameter is λ_s.

Inspection rule: Depending on his valuation, each buyer on the market will choose to inspect the property (or not). An 'inspection' means that the buyer 'visits the property and bargains with the seller': once he visits he always bargains in the model. Before inspecting, the buyer observes a detailed advertisement about the property on the listing website and already knows his own valuation for it (v^b) without inspecting it.[5] However, he only discovers the seller's and potentially other inspecting buyers valuations if he meets and start to bargain with them when he inspects the house. Since the buyer suffers a cost of inspecting the property (I assume this cost to be infinitesimal for simplicity), he only does so if he expects to have a chance to buy it (and thus benefit from his inspection). With an infinitesimal inspection cost, it will be the case as long as $v^b > v^s$ (as there is always a chance for him to have no better competitor and to be able to buy the home in this case). Thus, to determine whether or not it is beneficial for him to enter, the buyer must build expectation about unknown v^s. The buyer has limited rationality and uses the list price p^L as a *signal* about v^s to build a naive conjecture that $\hat{v}^s = g(p^L)$.[6] I use a simple affine functional form $g(x) = a_0 + a_1 x$ with $0 < a_1 < 1$. It yields the following simple *inspection/entry rule* that all buyers follow:

$$\text{a buyer inspects the property if } v^b > g(p^L) = a_0 + a_1 p^L$$

The seller only observes the number of inspections (N_t) and not the latent number of buyers on the market (N_t^{market}). Obviously, the seller can only sell her property to the inspecting buyers. She knows the buyers inspection rule $g()$, and since she knows that $v^b \sim \mathcal{LN}(\mu, \sigma)$, she is able to compute any $Pr(v^b > g(p^L))$ for any $p^L > 0$. As a consequence, the homeowner faces a classic trade-off when she sets her list price; ceteris paribus, a high list price allows her to 'sort' buyers with a higher taste for her property, leading to a higher expected sale price. However, this also signals a higher reservation value to the buyers and thus reduces the probability that a buyer will visit and enter the bargaining process ($Pr(v_b > g(p^L))$), leading to a longer time on the market. As staying on the market is costly for the seller (she has to keep her home tidy, spend time for

[5] An alternative story would be to say that visits are not costly at all (so all buyers visit and discover v^b when they do), but it is the decision to make a first offer which is costly. In this case the 'inspection' event would correspond to an 'entry in bargaining' or 'offer to the seller' instead of corresponding to a 'visit'. Because of its cost (going to the bank and dealing with all financial details), buyers also need to decide whether or not to 'inspect' (make an offer). If one chooses to enter the bargaining, he learns about the seller and his competitors values. The final outcome is determined via the bargaining rule.

[6] Buyers are very naive and uninformed in this model. This can be justified by the fact that they are simple one-shot buyers, staying on the market only for one period. They do not have the time to gather information, thus they have a naive expectation.
Obviously, the seller's list price choice is more complex than the buyers' conjecture and depends only partially on v_s. In fact, the list price choice even depends on the buyers' believed functional form $g()$ itself. I show in the results that this belief turns out not to be self-fulfilling: sellers reservation values differ from the buyers' simple conjecture. Since the conjecture $g(p^L)$ might not be correct, i.e. $\hat{v}_s \neq v_s$, a buyer entry in the bargaining process does not necessarily result in a sale: in particular if $v^s > v^b > \hat{v}^s$, the buyer enters in a bargain with the seller, but both agents quickly realize that there will be no profitable trade for both of them, and no transaction occurs (as specified in the bargaining rule).

potential visits, etc.), the optimal list price, which maximizes the seller's expected profit from sale, balances two opposite objectives: short time on the market and high sale price.
In addition to this classic trade-off, the list price also embeds an *informational externality*: ceteris paribus, a lower list price allows faster learning about λ_s. I detail the seller's learning process in the next section.

2.3.3 Seller's information and learning

When the seller decides to set her list price, she knows most parameters of the problem: she knows the bargaining rules, the buyers' inspection rule, the distribution of buyers' valuations (she does not know the realized value at the start of the period, but she knows that $v^b \sim \mathcal{LN}(\mu, \sigma)$ and knows μ and σ values), and that the number of buyers on the market will follow a $Poisson(\lambda_s)$ distribution. However, she has imperfect information about the demand since she *does not know the value of* λ_s: she only knows that in the population, $\lambda_s \sim Gamma(\alpha_0, \beta_0)$.

In period 0 (at the start of the listing), the seller forms an initial belief about λ_s, based on her rational expectation that $\lambda_s \sim Gamma(\alpha_0, \beta_0)$. Then for each period, she will update this belief via Bayesian learning rules, using the information she will acquire.
Let us drop the index s from λ_s and denote it only λ from now on (but remember it is an individual draw which is seller specific).
I determine the general *learning rule* for any period t. Suppose that the seller enters any period t with the prior belief that $\lambda \sim Gamma(\alpha_t, \beta_t)$ (i.e. $f_\lambda(\lambda) = \lambda^{\alpha_t - 1}\frac{\beta_t^{\alpha_t} e^{-\beta_t \lambda}}{\Gamma(\alpha_t)}$, $\mathbb{E}[\lambda] = \alpha_t/\beta_t$ and $\mathbb{V}[\lambda] = \alpha_t/\beta_t^2$).[7] To learn about λ, the seller will observe the number of inspections N_t (and not the latent number of buyers on the market N_t^{market} directly) in period t, and update her belief using this information.
First, recall that $N_t^{market} \sim Poisson(\lambda)$, and each of these buyers choose to inspect the property if their valuation is greater than $g(p^L)$. Thus, the process for the number of inspections depends on the list price of the period, $N_t \sim Poisson\big(\lambda \; Pr(v^b > g(p_t^L))\big)$: to learn about λ, the seller does not directly observe the latent process determined solely by λ, but by another process which is more or less close to it depending on the choice of list price. Nonetheless, computing the posterior distribution remains quite simple in this case. For simplicity, we denote $Pr(v^b > g(p_t^L)) = c_t$. Then,

[7] In this model, λ is fixed and does not vary with time. However, making λ vary through time in an unknown way for the seller would not change the learning rule at all (the seller would still only be able to observe the number of inspections).

the likelihood of observing $N_t = k$ inspections follows a $Poisson(\lambda c_t)$ distribution and is given by $f(N_t = k|\lambda c_t) = (\lambda c_t)^k e^{-\lambda c_t}/k!$.

Observing N_t, the seller updates her initial belief using Bayes formula to compute the *posterior belief*:

$$
\begin{aligned}
f_\lambda(\lambda|N_t = k) &= \frac{f(N_t = k \mid \lambda c_t)\ f_\lambda(\lambda)}{f(N_t = k)} \\
&= \quad ... \\
&= \frac{(\beta_t + c_t)^{\alpha_t+k}}{\Gamma(\alpha_t + k)}\ \lambda^{\alpha_t+k-1} e^{-(\beta_t+c_t)\lambda} \\
\iff \text{Posterior belief:} \quad & \lambda \sim Gamma\Big(\alpha_{t+1} = \alpha_t + N_t,\ \beta_{t+1} = \beta_t + Pr(v^b > g(p_t^L))\Big)
\end{aligned}
$$

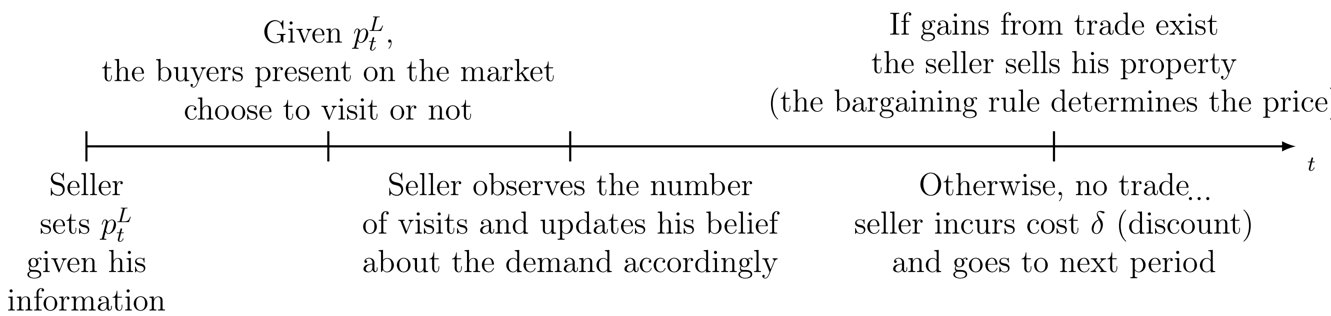
The learning rule is fairly simple: the α parameter of the prior is updated by adding to it the observed number of inspections, while we add the individual probability of inspection to the β parameter.

In this context, the list price p_t^L embeds a *learning externality*. To see this, note that when the list price is so high that $Pr(v^b > g(p^L)) \to 0$, the seller will always observe $N_t = 0$, no matter what λ is. Her listing does not provide her any information about λ in this case, and the seller does not learn anything, her belief stays the same over the period (since $Pr(v^b > g(p^L)) = 0$, she will observe $N_t = 0\ \forall\ N_t^{market}$ and thus $\alpha_{t+1} = \alpha_t$ and $\beta_{t+1} = \beta_t$). As p^L decreases, $Pr(v^b > g(p^L))$ increases and non-entries of some buyers are more and more likely due to the fact that there was indeed no buyer on the market (instead of being likely caused by a too low $Pr(v^b > g(p^L))$). This continues up to the opposite extreme scenario where p^L is so small that $Pr(v^b > g(p^L)) \to 1$, in which case a 'non-entry' only occurs when there is no buyer and $N_t = N_t^{market}$ (and the learning rule is actually equivalent to simple Bayesian updating with the basic $Poisson(\lambda)$ distribution where one adds 1 to β each period).

In general, a smaller list price leads to a higher probability of inspection, which makes the 'number of inspection process' closer to the latent 'number on market process' and allows us to learn faster about the demand parameter (λ) determining the number of buyers on the market.

2.3.4 Seller's optimization problem

Figure 2.3.1: Timeline of events in the model



The timing of the model is summarized in Figure 2.3.1. At the start of each period, the seller sets an optimal list price p_t^L in order to maximize her expected profit from the sale given her information set. Her information set consists of her belief about λ at the start of the period (which can be summarized by the two parameters (α_t, β_t)), and the knowledge of all the other parameters of the problem.

The number of buyers on the market is then drawn (from $Poisson(\lambda)$), as well as their valuations for the property. Given p_t^L and the inspection rule, each buyer chooses to inspect the property (or not). The seller then observes the number of inspections and updates her belief about λ according to the Bayesian learning rule: it determines her updated reservation value for the bargaining.

Once the seller has updated her v^s and all the buyers are entered, if at least one of the buyer has a valuation greater than the seller's, trade will occur according to the bargaining rule.[8] Otherwise, there is no room for beneficial trade and no trade occurs, the seller incurs a cost of keeping her property on the market (δ under the form of a discount factor) and goes to the next period where she repeats the same process, starting with her updated belief.

I repeat this game over an infinite horizon up to the point where the seller sells her property.[9]

Denote $\Omega_t = (\alpha_t, \beta_t)$ the seller information set at time t. Also denote the seller value $v^s(\Omega_t, \mu, \sigma, a_0, a_1, \delta)$ simply as $v^s(\Omega_t)$. Notice that the seller's valuation does not depend on λ directly and only depends

[8]The bargaining game is done within-period, meaning that future periods buyers can never enter it before the end of the process: it is as if I assumed that the sequential bargaining period was infinitely smaller than the dynamic game period of two weeks. The dynamic game only impacts the seller's reservation value that is built based on her expected gain (in the present or future): thus it only impacts her choice to leave the table in the bargaining process.

[9]In practice, to reduce the computational burden of the simulation, we choose a maximum number of weeks (e.g. two years) on the market, after which we stop the simulation if no sale has occurred.

on what the seller believes λ to be. The true λ will only impact the updating process of this belief (by generating the true number of buyers on the market that the seller will observe).

This value is pinned down by the following *Bellman's equation* which represents the problem of the homeowner when she sets her list price optimally given her information set at the start of each period:

$$v^s(\Omega_t) = \underset{p_t^L}{max} \quad \sum_{k=0}^{\infty} \mathbb{E}\Big[Pr(N_t = k \ |p_t^L, \lambda)\Big|\Omega_t\Big] \ \mathbb{E}\Big[\Pi^s(N_t = k, \ p_t^L, \ \Omega_t) \ \Big]$$

where the expected probabilities of receiving k visits based on the starting beliefs (Ω_t) are

$$\mathbb{E}\Big[Pr(N_t = k \ |p_t^L, \lambda)\Big|\Omega_t\Big] = \int (\hat{x} \ P(v^b > g(p_t^L)))^k e^{-\hat{x} \ P(v^b > g(p_t^L))}/k! \ \ f_\lambda(\hat{x})d\hat{x}$$
$$\text{with } f_\lambda(x|\Omega_t) = \frac{\beta_t^\alpha}{\Gamma(\alpha_t)} x^{\alpha_t - 1} e^{-\beta_t x}$$

and the corresponding profit function depends on the number of inspections and known updating in the case where this number of inspection indeed happens (i.e. $v^s(\Omega_{t+1})$ instead of Ω_t), as defined below:[10]

$$\Pi^s(N_t = k, \ p_t^L) = \begin{cases} \delta v^s\big(\Omega_{t+1}\big) & \text{if } v^s(\Omega_{t+1}) > v^b_{(k)} \\ \\ \underbrace{p^s\Big(v^b_{(k)}, v^b_{(k-1)}, v^s(\Omega_{t+1}), p_t^L\Big)}_{\text{bargaining rule function}} & \text{otherwise} \end{cases}$$

with $\Omega_{t+1} = \Big(\alpha_t + N_t, \ \beta_t + Pr(v^b > g(p_t^L))\Big)$. With the special case that $\Pi^s(N_t = 0, \ p_t^L) = \delta v^s(\Omega_{t+1})$. The expectation of seller profit is taken with respect to the two highest buyers' values (which are the only ones which potentially matter in the bargaining rule, and which are unknown to the seller when she sets her list price) using the joint density of the two highest order statistics among k. This joint density of two order statistics is in general given for any $i < j \in 1, 2, ..., n$, $\forall x < y \in \mathbb{R}$

[10]For any number of visits that occur when the seller chooses to set a given p_t^L, she knows how she will update Ω_t to Ω_{t+1}, thus she knows the corresponding updated value that she would bargain with in each specific entry case $(v^s(\Omega_{t+1}))$. The probability of observing each specific number of entries are computed using the period starting belief Ω_t.

by:

$$f_{(i,j):n}(x,y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j}f(x)f(y)$$

where f is the truncated lognormal(μ, σ) on $x > g(p^L)$, and F its cdf.

This value function is estimated via value function iteration. The iteration is done on a discrete grid of α and β values. Values for points inside the state space but out of the grids are approximated via bilinear interpolation between the four surrounding points.

2.4 Estimation method and identification

The structural parameters that I want to estimate are: $(\mu, \sigma, a_0, a_1, \delta, \alpha_0, \beta_0)$. Denote the structural parameters vector θ. θ is estimated via simulated method of moments (SMM). The idea of this estimation method is to find the set of parameters for which the simulated sellers' behavior will be the closest to the observed sellers' behavior. To do so, I select features from the empirical data that I want to reproduce by picking a vector of N empirical moments of interest. I denote m^d this $N \times 1$ vector of moments. Then, for a given θ, I construct the corresponding counterpart vector of simulated moments $m^{sim}(\theta)$. These simulated moments are computed on the selling outcomes data of S (=100000) iid simulations of my model with underlying structural parameters θ.[11]

I estimate the seven unknown parameters by minimizing a distance function between empirical and simulated moments such that:

$$\hat{\theta} = \underset{\theta}{argmin}\Big[m^d - m^{sim}(\theta)\Big]' W \Big[m^d - m^{sim}(\theta)\Big]$$

where W is a $N \times N$ positive definite weighting matrix equal to the inverse of the variance-covariance matrix of my moments (computed using bootstrap). To find $\hat{\theta}$ I use a controlled random search algorithm (Price, 1983).

To estimate the seven structural parameters I pick a list of $N = 61$ moments representing the features I want to reproduce with my model. These moments can be categorized in three main

[11] Recall that I normalized the data by the predicted sale price, which allows me to compare everything on the same scale by using a single representative problem for every seller (independent of the 'quality' of the property). Hence, I can run iid simulations of the representative problem and compare it to the normalized real data.

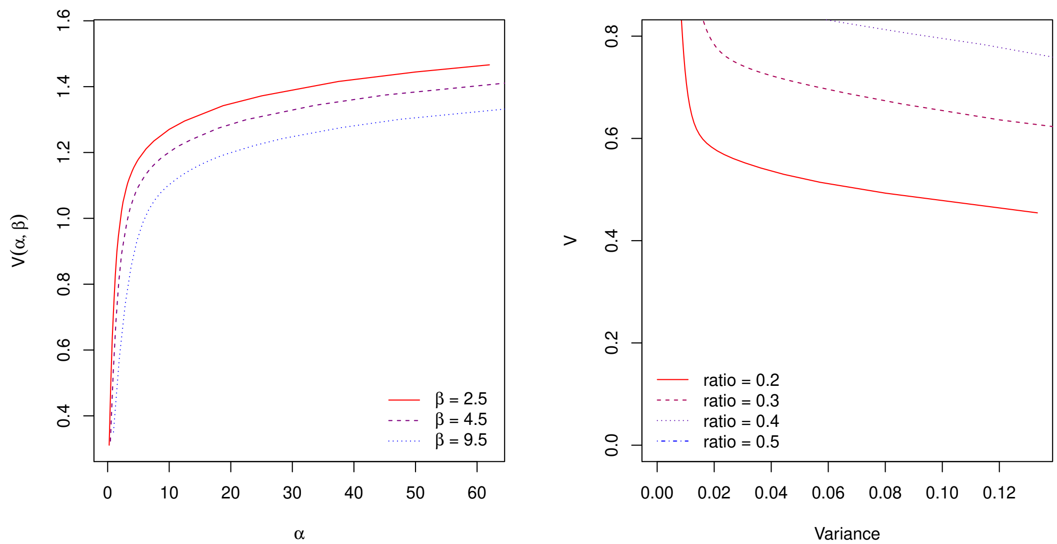
dimensions of the selling process: the time on the market, the distribution of sale price and the list price dynamics (see Appendix 2.B for a list of all the moments). Most of the price moments are relative to the initial list price. This allows me to reduce issues caused by the 'scaling' of the problem or to soften the impact of unobserved heterogeneity not accounted for in the hedonic value estimation. I have only two 'non-relative' price moments: the average sale price and the average list price. They are compared to their counterparts normalized by the predicted financial value (estimated by hedonic regression) in the data.

These moments allow us to identify the parameters. Intuition about identification is non-trivial as most parameters influence several features of the model simultaneously. (α_0, β_0) determines the initial value and list price choice. By pinning down the λ distribution, they also determine the list price dynamics and have a strong influence on the time on the market. (μ, σ) pin down the buyer valuations, and thus the seller value and the list price level he can set. They also influence the final sale price through the buyer and the seller value. In particular, (μ, σ) will directly determine the distribution of sale price when it is above the final list price (as the model imply that in this case $p^s = v^b_{(n-1)}$ and these two parameters solely determine the distribution of $v^b_{(n-1)}$). a_0 and a_1 are also identified via the sale price distribution. In particular, they determine the minimum level of entry of buyers and thus the minimum ratio of sale over list prices. δ only impacts the seller valuation and allows to adjust it better than (a_0, a_1, μ and σ) which face more restrictions (as they determine more specific moments).

2.5 Results

2.5.1 Value function and optimal list price

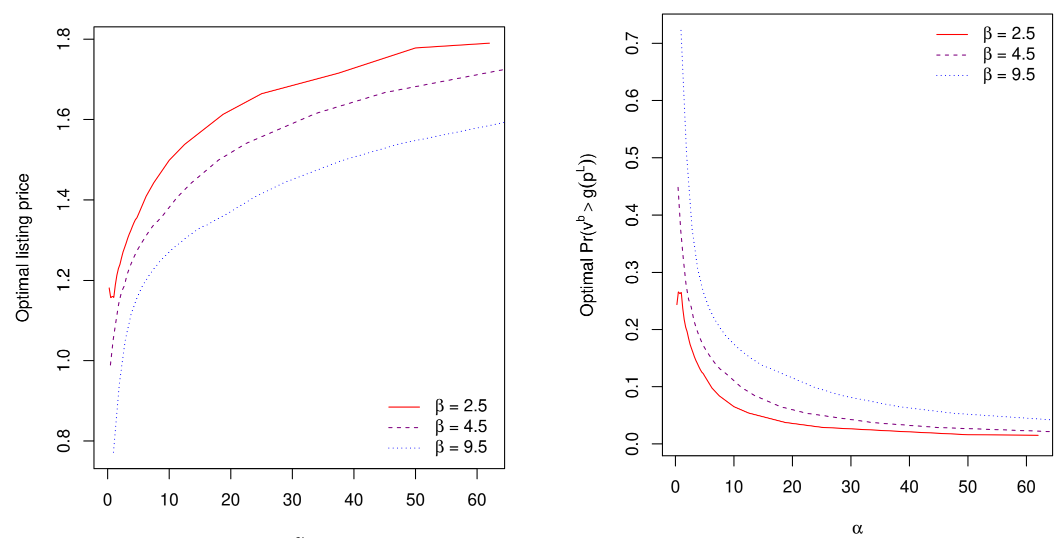
Figure 2.5.1: $v^s(\alpha, \beta)$



(a) Value as function of α for fixed β

(b) Value as function of variance α/β^2 for fixed expectation α/β.

Figure 2.5.2: Optimal list price $p^L(\alpha, \beta)$ and corresponding $Pr\big(v^b > g(p^L)\big)$



(a) p^L as function of α for fixed β

(b) $Pr\big(v^b > g(p^L)\big)$ as function of α for fixed β

I obtain intuitive results for the value function (Figure 2.5.1) and optimal list price (Figure 2.5.2). For the interpretations, recall that if $\lambda \sim Gamma(\alpha, \beta)$, then $\mathbb{E}(\lambda) = \alpha/\beta$ and $V(\lambda) = \alpha/\beta^2$. First, the value increases with the expected number of buyers on the market. Second, at the fixed average belief, the less uncertain the seller is (i.e., the smaller variance of the belief), the higher the value she obtains: as the uncertainty decreases, the value converges to the perfect information benchmark (as if the seller knew its own λ draw). As for the optimal list price, the higher the expected latent number of buyers on the market (higher expected λ), the smaller the chosen probability of inspection (via a higher list price), which balances the expected number of inspections overall (depending on $\lambda Pr\big(v^b > g(p^L)\big)$.

2.5.2 Parameters and Moments

Table 2.5.1: Parameter Estimates of the Structural Model

Parameter	Description	Estimate	Std. Errors
δ	Subjective discount factor (and cost of listing)	0.986	0.00009
Demand parameters			
Valuation process:	$v^b \sim \mathcal{LN}(\mu, \sigma)$		
μ	Mean of buyer valuation	-0.0425	0.00044
σ	Standard deviation of buyer valuation	0.1902	0.00039
Number of buyers: (*and initial rational belief*)	$N^{market} \sim Poisson(\lambda)$ and $\lambda \sim Gamma(\alpha_0, \beta_0)$		
α_0	α prior belief and true distribution	4.48	0.005
β_0	β prior belief and true distribution	11.21	0.0057
Inspection rule	buyer inspects if $v^b > a_0 + a_1 p^L$		
a_0	Buyer's conjecture about seller reservation value: constant	0.409	0.0018
a_1	Buyer's conjecture about seller reservation value: slope	0.58	0.0015

Table 2.5.1 reports the parameter estimates. To understand the mechanism at play, we focus on the screening of buyers implicitly done by the seller at the optimal parameters. Take the screening in the initial period for example. The seller face a demand $v^b \sim \mathcal{LN}(\mu = -0.0425,\ \sigma = 0.19002)$. She sets an initial listing price of 1.106 (no heterogeneity in the first period since it is a representative

agent model and everyone starts with the same rational expectation). In this case, $g(p^L) = 1.05021$, $Pr(v^b > g(p^L)) = 31.52\%$: the seller aims for high quantiles of the demand (top 31.52%). For example, if λ is drawn at its average value ($\alpha_0/\beta_0 = 0.4$), then the true rate of inspection $\lambda Pr(v^b > g(p^L))$ is equal to 12.61%.

The sellers with a high draw of λ will have a considerably higher probability of sale than they expect, and will thus spend a shorter time on the market than they would like to if they had perfect information (however, some will stay on the market and increase their price in the next period if they received multiple visits, they often choose to sell at a price they would have refused with perfect information in the end). On the contrary, sellers with a low draw of λ have a very small probability of sale and are aiming at too high quantiles of the demand (compared to what they would do if they had perfect information about λ): they will 'survive' longer on the market (than people with high λ). As they stay on the market and obtain information that their λ is low, they will decrease their list price to aim at lower quantiles of the buyer valuations in order to compensate for their lower λ and still have a chance to sell their property.

The model is able to match the decreasing list price dynamics observed in the data thanks to the fact that sellers who stay longer on the market ('survivors') are sellers with low draw of λ. These sellers progressively learn and adjust their expectation about the demand (λ) downwards, and thus decrease their list price choice. For these individuals with low draws of λ, the initial belief was too 'optimistic': the correction of their initial belief via learning is the reason for decreasing list price. In addition to this, the decreasing list price is also due to some selection/survivor effect (even for high draws of λ): those who are unlucky (no matter what λ is) and observe no entries will stay longer on market and have a decreasing belief. Thus, in general, those who stay longer on the market are likely to have observed fewer entries (even though some stay because they refused offers, the majority stays because they received none), and thus have a more 'downward updated' belief.

Table 2.5.2: Actual and Simulated Moments

Moment	Actual	Simulated
Mean sale price	1.008	1.02
Mean ratio sale/final listing price	0.955	0.962
Mean initial list price	1.107	1.106
% of accepted offers equal to list price	0.15	0.235
% of accepted offers below list price	0.734	0.712
Mean week on the market (knowing that <52 weeks)	14.817	14.736

Table 2.5.2 and Figure 2.5.3 illustrate how the model matches the moments. Overall, for a small number of parameters (7), it fits the data reasonably well. The list price dynamics and the distribution of the time on the market are well fitted. The sale price distribution (relative to the list price) is matched correctly, except for the tails. In particular, I fail to reproduce the number of sales above the list price (only 5.3% in the simulation against 11.6% in the data). This is because, even with the split demand, the model is still unable to match the time on the market distribution and the sales above the list price at the same time.

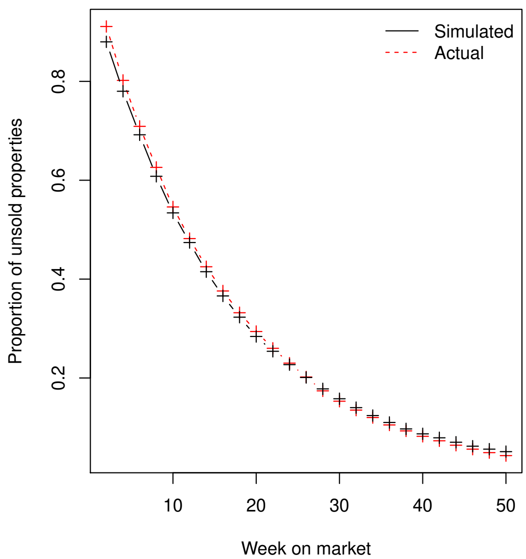
2.6 Value of information

All of the comparisons in this section are completed using the model at the estimated optimal parameters given previously.

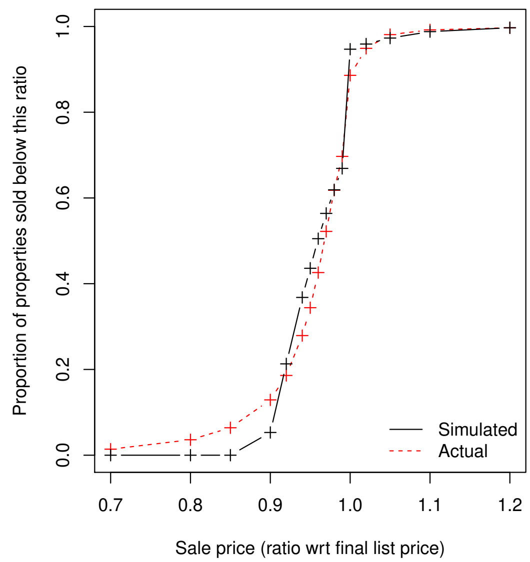
I use my estimated model to get an idea of the *value of information*: how much the seller would gain from being better informed? To answer this, I compare the imperfect information outcomes to the outcomes obtained by a *perfectly informed* (denoted PI) seller, who would know the value of her λ draw.

I compare this benchmark to the value obtained by an imperfectly informed (denoted II) seller who does not know her draw λ and starts from rational expectation $Gamma(\alpha_0, \beta_0)$. To compute the 'realized value' of this imperfectly informed seller, I simulate the complete selling process (with corresponding dynamic belief updating) and observe the price p^S and time spent on the market by the seller. From period 0, the average (over several simulations) discounted sale price ($\delta^t p^S$) gives the empirical counterpart to the theoretical seller value.

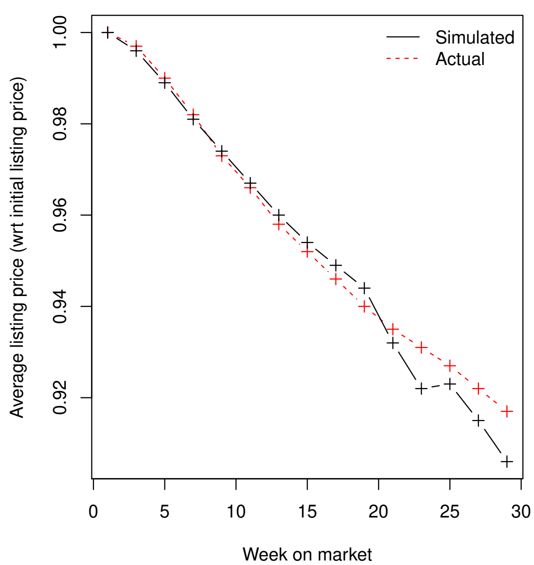
Figure 2.5.3: Actual and Simulated Moments



(a) Proportion of unsold properties as a function of time on the market



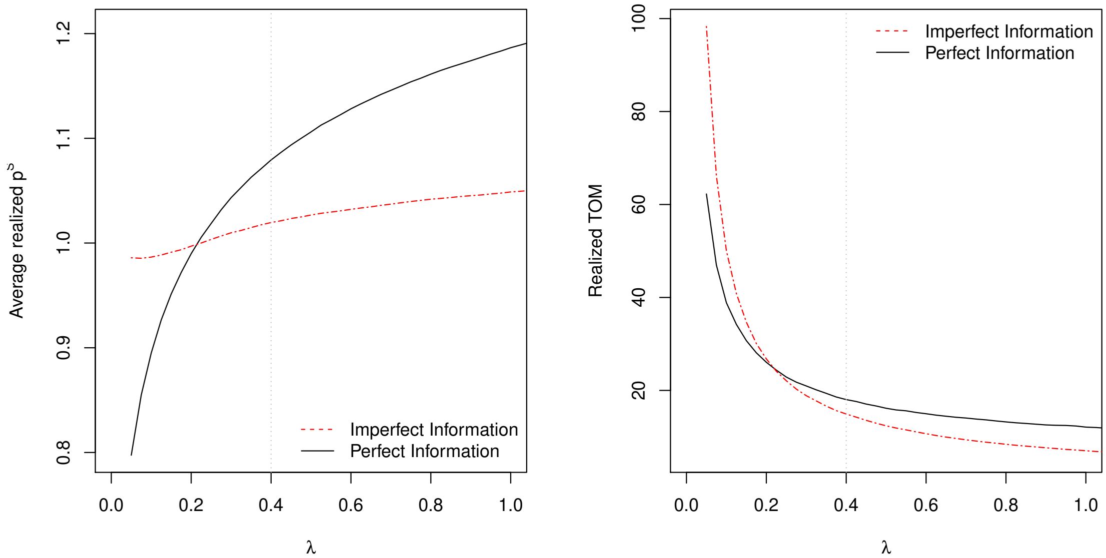
(b) Distribution of sale prices (ratio to final list price)



(c) Average list price as a function of time on the market (ratio to initial list price)

To establish an idea of the value of information, I compare the perfectly informed benchmark valuations to the counterparts' realized outcomes of imperfectly informed sellers. In particular, I do this for several values of λ. In this way, I can see how far II sellers end up from the perfect information benchmark as a function of their initial information error (how far the initial rational belief was from their true λ).

Figure 2.6.1: Value of Information: Sale Price and TOM
(average computed over 10000 simulations)



(a) Average obtained p^S (not discounted) as a function of λ (for fixed initial belief (α_0, β_0))

(b) Average time spent on market (TOM) as a function of λ (for fixed initial belief (α_0, β_0))

Figure 2.6.1a shows the average sale price and time on the market (simulated) for different values of λ. This is done for a PI seller who knows λ, and for a II seller who starts from initial rational belief that $\lambda \sim Gamma(\alpha_0, \beta_0)$ (calibrated at the model estimates) and who progressively updates it.

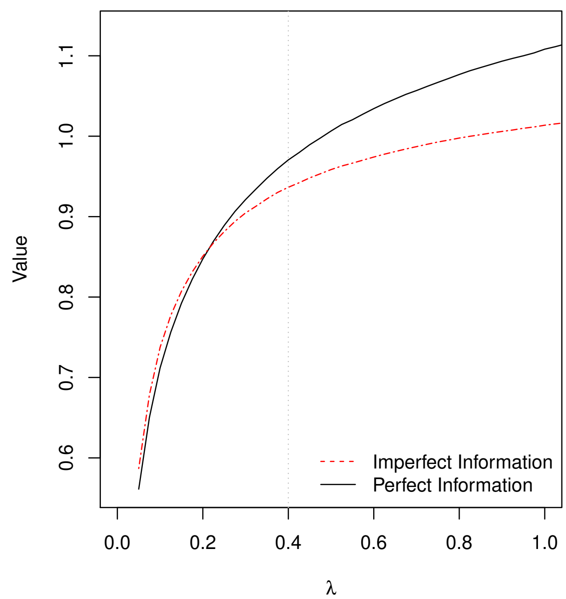
First for both types of sellers, I observe that the higher the λ the higher the sale price obtained and the shorter the time spent on market.

More interestingly, we see that the sale price reacts more to λ for perfectly informed sellers. Indeed, the average sale price obtained by a PI seller varies from 0.797 when $\lambda = 0.05$, to 1.187 when $\lambda = 1$. While the average sale price obtained by a II seller (with initial belief $Gamma(\alpha_0, \beta_0)$) varies from 0.986 when $\lambda = 0.05$, to 1.049 when $\lambda = 1$. The fact that imperfectly informed agents start from the same initial belief smooths the obtained sale price: at the time they sell, despite their updating, they are generally still far from perfect information. Thus, their list price is far from the optimal one (if perfectly informed they would set a lower price to sell faster), hence the final sale price difference. In terms of sale price, sellers with bad draws of λ benefits from this, while sellers with high draws of λ get a smaller price than what they could.

However, as shown in Figure 2.6.1b, imperfectly informed sellers with bad draws are able to earn

more only because they spend too much time on the market (about 99 weeks on average for a II seller with draw $\lambda = 0.05$, compared to 62 weeks for her PI counterpart). In terms of value obtained by the II seller, this overly-long time on the market attenuates the higher sale price obtained.
Similarly, imperfectly informed sellers with high draws spend a shorter time on the market than their perfectly informed counterparts, since they put their listing at a suboptimal list price (and screen less the demand than they would if they were perfectly informed). This could offset the loss in the sale price they endure compared to the PI benchmark.
The question now is to determine how the effects on sale price and time on the market translates into the average 'value' (average discounted sale price) obtained (from period 0).

Figure 2.6.2: Value of Information: Information Paradox
(average computed over 10000 simulations)



Value as a function of λ (for fixed initial belief (α_0, β_0)). The vertical grey line represents $\mathbb{E}[\lambda]$ at the initial rational belief (i.e. α_0/β_0).

Figure 2.6.2 shows how these two effects translate in terms of value. Obviously, the higher the demand λ, the higher the value obtained for perfectly and imperfectly informed agents. However, one can notice an *informational paradox*.
Indeed, one would expect the perfectly informed agent to always outperform the imperfectly informed counterpart because she is solving the 'true' problem (knowing λ), and thus optimizes correctly. However, this is not always the case, and the imperfectly informed agent can perform better

than her perfectly informed counterpart in this model. Indeed, there is an area where the value obtained by the imperfectly informed agent who starts from the belief that $\lambda \sim Gamma(\alpha_0, \beta_0)$ is higher than that obtained by her perfectly informed counterpart who knows λ (i.e. the area where the red-dotted curve is above the black curve in Figure 2.6.2). This happens only to agents with bad draws (i.e., imperfectly informed agents who expect λ to be considerably higher than what it truly is). The realized outcomes (discounted sale price) with an 'overoptimistic' belief can be higher than the ones the agent could obtain if she were perfectly informed and optimizing (choosing her list price) knowing the exact value of λ. This means that having better information about λ is not necessarily beneficial for the seller, as the value of additional information may indeed be negative. This seems somewhat perplexing, as one should not be able to do better than a perfectly optimizing agent who knows exactly what demand she should expect, and thus, what her true value v^s should be. The explanation lies in the estimation of her reservation value by the seller. Indeed, even if the overoptimistic seller is not optimizing correctly with respect to the true λ (she sets a too high p^L and stays too long on the market by 'screening' too much and 'over-rejecting' some buyers offers), she genuinely *overestimates her reservation value* $v^s(\alpha, \beta)$. This gives her a better 'bargaining position' (the threshold at which she leaves the bargaining game) in the bargaining game, which allows her to obtain a higher sale price than the one a PI seller would obtain if she was trading with the same buyers. One can directly see this in the bargaining rule: ceteris paribus, if v^s is higher, $v^s + 0.5(v^b_{(n)} - v^s) = 0.5v^s + 0.5v^b_{(n)}$ is also higher. There is some level of overconfidence where the overconfident seller obtains a sale price sufficiently high to offset the longer time she spends on the market, yielding a better outcome overall.

At some point (not visible in Figure 2.6.2), the gains in the bargaining position are offset by a too large 'expectation error'. By being 'too overoptimistic' the agents spend too long time on the market, which offset the stronger bargaining position and yields a lower value than being perfectly informed.[12]

Notice that the highest values in case of overoptimism are only explained by this improved bargaining position. To check this, I can recompute a variant of the model where the seller is infinitely more patient than the buyers in the bargaining game. This way she has all the bargaining power and can always extract the highest buyer valuation ($v^b_{(n)}$) without splitting the pie. Thus, she already has

[12]It does not appear with our estimated set of parameters, but with other parameters, for very small λ draws, the black curve will again be higher than the red one. There is an 'optimal level' of 'overconfidence' (not too large or too small) for which the misinformed agent performs better than her perfectly informed counterpart. You can see this by zooming in on Figure 2.6.2: the positive gap between II and PI values (when λ is low) has an inverted U shape.

full bargaining power and cannot have a stronger bargaining position due to her overconfidence. In this case, the information paradox disappears: the perfectly informed agent always performs better (not displayed here).[13]

This explains that while the imperfectly informed with a high draw of λ is, on the other hand, considerably worse off than her perfectly informed counterpart, she does not only commit optimization errors (setting a too low list price, resulting in less screening of buyers, lower sale price and shorter time spent on market) due to her imperfect information set, but she also has a weaker bargaining position because she is too 'pessimistic' in her reservation value estimation v^s.

Thus, the cost of imperfect information is higher for people facing high demand. High demand sellers would prefer to be perfectly informed of the demand that they face, while on some levels, the low demand sellers can be better off by being overoptimistic and ignorant that they face a low demand.

2.7 Conclusion

Taking advantage of a new large dataset of real estate listings, I highlight some evidence that imperfect information and sellers' learning impact the selling outcomes on the housing market. I have developed a simple theoretical model with a new Bayesian learning application in order to explain some of the housing market stylized facts.

Learning about the demand by rational but imperfectly informed home-sellers is a key feature of the model used to explain these facts. In particular, the progressive downward adjustment/correction of the belief of individuals facing low demand is key to explaining the decreasing list price dynamics, and matches well the distribution of time spent on market. My work also highlights a paradox that the value of information is not necessarily positive. Indeed, by being imperfectly informed and overconfident about the demand, a seller can overestimate her reservation value and have a stronger bargaining position. This allows her to extract a higher sale price, which can compensate her mis-optimization and longer time spent on the market (with respect to her perfectly informed

[13]Technically, when the seller is infinitely more patient, her reservation value (v^s) vanishes from the bargaining equation determining the sale price. For a given buyer value (v^b), v^s now only impacts the decision to reject an offer, without impacting the sale price. Thus, it will only lead to a misguided rejection decision and the stronger bargaining position will not matter (because the seller already has all the power). To be clear, II seller will still get a higher average sale price (because she screens higher value buyers with her rejections). However, the longer time on the market will offset this advantage and the PI seller will always be better off.

counterpart).

This theoretical work could serve as the foundation for future applications relative to the home-selling problem. For example, it can be easily extended to study learning within complete neighborhoods, or to study dynamic entry/withdrawal decisions of listings by the sellers to match market stocks.

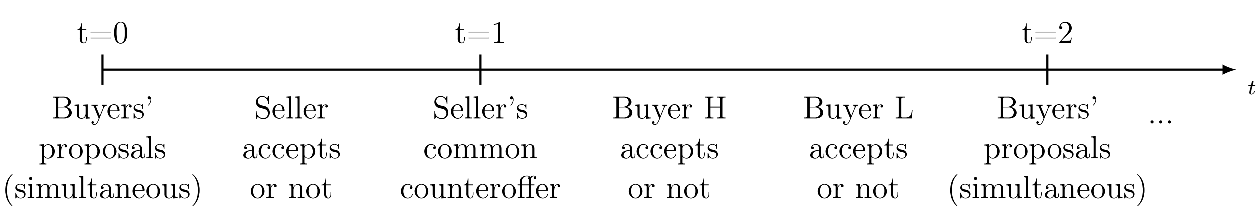
Appendix

2.A Bargaining rule

I study a simplified version of the problem with one seller and only two buyers and with normalized valuations (as described next). This generalizes easily to N buyers with ordered valuations: only the two buyers with the two highest valuations matter with complete information.
The proof is inspired from Shaked and Sutton (1984) and Binmore et al. (1989), i.e. bilateral bargaining with outside option (for the player playing second). The problem setup and thus the result is also close to the 'auctioning model' in Binmore et al. (1992).[14]

2.A.1 The Problem: 1 Seller - 2 Buyers

There is one seller of an indivisible good, with reservation value $v_s = 0$. The seller can sell it to one of two buyers H and L (high and low) with valuations $v_H = 1 > v_L = v$.[15] The three agents have a common discount factor ρ. The period length is τ. Denote for simplicity $\delta = \rho^\tau$. If the good is sold at price p after t periods, the seller's payoff is $p\delta^t$, the successful buyer's payoff is $(v_b - p)\delta^t$ and the losing buyer gets zero. Information is complete.
The *timeline* of the problem is as follow:



[14]Except for a different timeline (buyers start for me, which change the result), and some notation/normalization changes. However, no proof is provided in the book, and the references given to find the proofs are unavailable/unpublished or do not contained the proof at all.
[15]I normalize $v_H = 1$ so that the 'size of the cake' to share is one as usual. I denote v_L as v for simplicity.

In period $t = 0$, each buyer simultaneously makes a proposal to the seller. She may accept one of these offer or reject both. If the seller accept one of the offer, the two players trade and the game ends. For simplicity, if the seller wants to accept an offer that both buyers made (i.e., in the case of a tie), the tie breaking rule is that the seller will opt for buyer H.

If the seller reject both offers, there is a delay τ and they go to the next bargaining period. The seller then makes a common counteroffer to both buyers. The highest valuation buyer either accept or reject it. If he rejects it, the low valuation buyer can then choose whether or not to accept it. If both buyers reject the offer, they go to the next period (with delay τ) where they make simultaneous counteroffers, as in the first period, and the game repeats itself, etc.

2.A.2 Solution

This game always has a unique subgame-perfect equilibrium outcome where the good is sold immediately to buyer H (the seller accepts his offer directly) at a price:

- $p = \frac{\delta}{1+\delta}$ if $\frac{\delta}{1+\delta} \geq v$. Notice that this is the bilateral bargaining price of the seller with buyer H, in this case, it is as if the second buyer was absent (he represents a non-credible threat for buyer H / non-credible outside option for the seller).
- $p = v$ if $\frac{\delta}{1+\delta} < v$. In this case the presence of the buyer L matters, gives more 'power' to the seller, and forces buyer H to pay a higher price than if buyer L was absent.

2.A.3 Proof

As in the classic Shaked and Sutton (1984) proof, let m_b and M_b be the infimum and supremum of equilibrium payoffs to the buyer H in the game. Let m_s and M_s be the infimum and supremum of equilibrium payoffs to the seller in the companion game in which she would move first (i.e. starting from period $t = 1$, for example). Let us also assume for now that the buyer L always makes the same equilibrium offer denoted s (≤ 1 since $v_L \leq 1 = v_H$). From the point of view of bilateral bargaining between the seller and buyer H, it acts as an 'outside option' for the seller: if she accepts it, she obtains s and she 'leaves' buyer H with nothing.[16]

[16] Obviously since buyer H can give more than L (since $v_H = 1 \geq v_L$), he is always able to attract the seller to trade with himself, by offering something greater or equal than s if necessary.

As in the bilateral bargaining with outside option proof from Binmore et al. (1989), I have the following system of inequalities which hold:

$$m_b \geq v - max\{\delta M_s,\ s\} \tag{2.1}$$

$$v - M_b \geq max\{\delta m_s,\ s\} \tag{2.2}$$

$$m_s \geq v - \delta M_b \tag{2.3}$$

$$v - M_s \geq \delta m_b \tag{2.4}$$

Inequality (1) can be explained as follows: the seller must accept any opening offer greater than what she can get by making a counteroffer to buyer H in the next period, or by accepting the low buyer offer (thus no δ cost). As a consequence, the buyer H cannot get less than $v - max\{\delta M_s^H,\ s\}$, hence the first inequality. Inequality (2) follows from the fact that the seller must get at least either δm_s^H by making a counteroffer to buyer H, or s by accepting the low buyer offer. As a consequence, the buyer H can get at most $M_b \leq v - max\{\delta m_s^H,\ s\}$, hence the second inequality. Similarly, inequality (3) and (4) comes from the same reasoning but for buyer H and thus, there is no s involved (as if $s = 0$, i.e. no offer from another seller that he could accept for example/no 'outside option').

Now, to determine the equilibrium outcomes, distinguish three cases:

- If $s \leq \delta m_s$: *($\iff$ the offer from L is irrelevant)*
  Combining (1) and (4) yields:
  $\delta - \delta^2 M_s \leq \delta m_b \leq 1 - M_s$, thus $\delta - \delta^2 M_s \leq 1 - M_s$, which gives: $M_s \leq 1/(1+\delta)$.
  Combining (2) and (3) (rewritten) yields:
  $1 - m_s \leq \delta M_b \leq \delta - \delta^2 m_s$, thus $1 - m_s \leq \delta - \delta^2 m_s$, which gives: $1/(1+\delta) \leq m_s$.
  Thus:

  $$\frac{1}{1+\delta} \leq m_s \leq M_s \leq \frac{1}{1+\delta}$$

  Similarly, combining (2) and (3) for the upper bound, and (1) and (4) for the lower bound, yields:

$$\frac{1}{1+\delta} \leq m_b \leq M_b \leq \frac{1}{1+\delta}$$

Thus, $m_s = M_s = m_b = M_b = 1/(1+\delta)$ in this case (and buyer should offer $1/(1+\delta)$ to the seller, who will accept in this case). Thus, this case should happen when $\delta m_s = \delta/(1+\delta) \geq s \iff s \leq \delta/(1+\delta)$. [17]

- If $\delta m_s < s < \delta M_s$: (2) becomes: $1 - M_b \geq s > \delta m_s$, but we still have, as before: $1 - M_b \geq \delta m_s$, thus we will still find $\frac{1}{1+\delta} \leq m_s \leq M_s \leq \frac{1}{1+\delta}$, which is a contradiction.
  Thus, this case is not possible.

- If $\delta M_s \leq s$: *($\iff$ the offer from L is greater than what the seller could get with classical bilateral bargaining with H)*
  For the buyer, we directly have from (2) that: $M_b \leq 1 - s$, and from (1) $m_b \geq 1 - s$. Thus:

$$1 - s \leq m_b \leq M_b \leq 1 - s \iff m_b = M_b = 1 - s$$

Which means that the buyer should immediately make an offer of s to the seller, and he will not be able to obtain more.
We still need to compute the seller outcomes (to check when this case happen).
As before, from (1) and (4) we have: $M_s \leq 1 - \delta(1-s)$. From (2) and (3) we have: $1 - \delta(1-s) \leq m_s$.

$$1 - \delta(1-s) \leq m_s \leq M_s \leq 1 - \delta(1-s) \iff m_s = M_s = 1 - \delta(1-s)$$

[17] Intuitively, instead of resorting to inequalities, see the proof as in Shaked and Sutton (1984) in each case. Let's take an example with the proof for the upper bound of M_b.
We focus on the subgame starting from period $t = 2$. The game which starts at this point is the same as the initial game (its first repetition) but with a discounted sum of payoffs $= \delta^2$ (cannot get more than this). By definition, the buyer can get at most $\delta^2 M_b$ at this point. Now, consider the (companion) subgame starting in the preceding period $t = 1$. Any offer by the seller which gives the buyer more than the supremum of its payoffs ($\delta^2 M_b$) should be accepted. So there is no perfect equilibrium in which the buyer receives more than $\delta^2 M_b$, and thus it follows that the seller should get at least $\delta - \delta^2 M_b$ in this period (it is $\delta - \delta^2 M_b$ and not $1 - \delta^2 M_b$ since the discounted value of the total payoff at time $t = 1$ is δ and not 1). In other words, $m_s \geq \delta - \delta^2 M_b$. As a consequence, starting in period $t = 0$, the seller will not accept anything less than the infimum of what she will receive in the game beginning next period (which has present value $\delta - \delta^2 M_b$). Thus, the buyer can get, at most $M_b \geq 1 - \delta + \delta^2 M_b$. This finally gives the upper bound: $M_b \leq 1/(1+\delta)$.
Obtain that $m_b \geq 1/(1+\delta)$ and the results for the seller by similar reasoning.

Thus, this case should happen when $\delta M_s = 1 - \delta(1-s) \leq s \iff s \leq \delta/(1+\delta)$.

Therefore, if subgame-perfect equilibria exist, they generate a unique subgame-perfect equilibrium outcome. In addition, existence is trivial: each player always demands his equilibrium payoff when proposing, and accept his equilibrium payoff (or more) when responding.
Thus, we have the result that the buyer H makes an offer at p, which is immediately accepted by the seller with the offer from the low buyer s as her outside option. With p defined as:

$$p = \begin{cases} 1 - 1/(1+\delta) = \delta/(1+\delta), & \text{if } s \leq \delta/(1+\delta) \\ s, & \text{if } s \geq \delta/(1+\delta) \end{cases}$$

The result is quite intuitive: either the offer from buyer L is too low and is not taken into account by the buyer H and the seller (irrelevant outside option for the seller, they do classic bilateral bargaining), or it is high enough and allows the seller to gain a credible threat, which increases her payoff.

Now the question is to determine what is s, the equilibrium offer from the buyer L (if it exists). Let's assume that the buyer L cannot make an offer greater than his valuation v.[18]

- If $\delta/(1+\delta) > v$ then anyway L has no power to disturb bilateral bargaining between the

[18] Even if it seems natural, this is an important assumption to get rid of absurd equilibria where L would bid more than v in this setup. Because the buyers bids simultaneously with perfect information, it is close to the case of first price sealed bid auction with perfect information without restrictions on bids. It is well known that the set of Nash equilibria of this kind of auction is determined by three conditions: it is the set of profiles b of bids with $b_H \in [v_L, v_H] = [v, 1]$, $b_j \leq b_H \forall j \neq H$ and $b_j = b_H$ for some $j \neq 1$. Thus we could have any equilibrium offer $> v$ from the low buyer, and the proof may fall down.
On the other hand, if we impose that he cannot bid more than his valuation v - which makes sense in the context of bargaining where the set of possible offers is between the valuation of the seller and the one of the corresponding buyer - then the only Nash equilibrium is $b_H = b_L = v$ (and because of the tie breaking rule, H wins). Indeed, it is clear that any outcome with $b_H < v$ and $b_L < v$ is not an equilibrium since one of the two players would gain more by increasing his bid. Similarly: $b_H = v$ and $b_L < v$ is not an equilibrium either, since in this case H would be tempted to decrease slightly his offer in order to get more.
Otherwise, instead of making this assumption, one solution would simply be to make the buyers bid non simultaneously: H first, then L, in which case L would never have interest to bid more than H's offer if it's higher than v, and thus H would never bid more than v in a first price auction... One has to choose between the sequential bids or the simultaneous bids with the natural assumption that the buyers cannot bid more than their value. I prefer the former since it seems unnatural to make them bid sequentially with predefined order based on the valuations.

seller and H, s does not matter at all in the problem (irrelevant outside option). Buyer L could make any offer $s \leq v$ in equilibrium, it would not change the equilibrium outcomes, let's assume he bids $s = v$ in this case.

- If $\delta/(1+\delta) < v$ then L has some power (it is close to first price sealed bid auction with perfect information in this case).

  We find that $s = v$ in this case is the only equilibrium. Indeed: if $v > s$, it is possible that buyer H wins the auction by bidding b_H below v and above s which is not an equilibrium since in this case L would be better off by increasing his bid above b_H. At the same time, it is not possible that buyer L wins the auction in equilibrium (buyer H can always bid more). The only equilibrium offer from L is $s = v$.

Thus, as expected, in equilibrium the low buyer will bid his valuation $s = v$. This yields the final result that the only equilibrium payoffs is that buyer H makes an offer at p which is immediately accepted by the seller (who have an offer v from the low buyer as outside option). With p defined as:

$$p = \begin{cases} \delta/(1+\delta), & \text{if } v \leq \delta/(1+\delta) \\ v, & \text{if } v \geq \delta/(1+\delta) \end{cases}$$

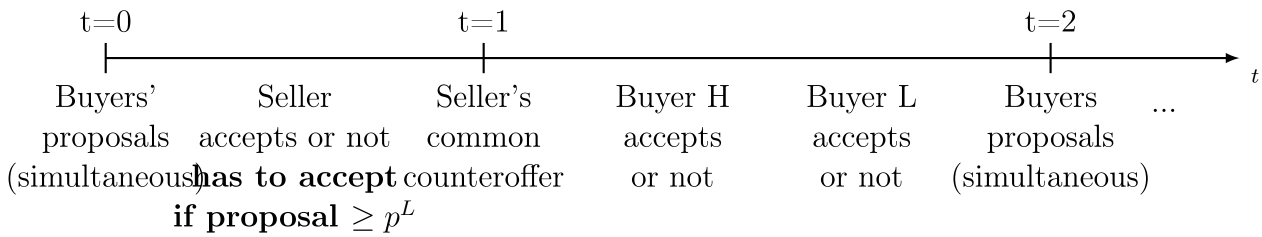
2.A.4 Extension to N buyers

One can easily generalize to more than two buyers. Indeed, we do not really care about additional buyers choices: as long as they are lower than v, they will not have any impact on the equilibrium outcomes. Only the two highest valuations matter, so simply consider that in the proof here; that buyer H and L are the two buyers with the highest valuations and the proof is already generalized for any number of buyers.

2.A.5 Extension: list price

I slightly modify the problem by adding an exogenous list price p^l. The list price serves as a commitment device; thus, if a buyer makes an offer greater or equal to the list price in the first period, the seller is obliged to accept to sell her good. If two buyers make offers greater than the list price, she obviously accepts the greatest offer (and in case of tie, she chooses buyer H). The list price only works in the first period; after which, the game is unchanged (in period $t = 2$ we go back to previous case where the seller can refuse an offer greater than p^L).

The problem is as follows:



Consequently, the list price only affects the buyer's choice (in the first period), not the seller's (who only endure it). From the point of view of the seller, either she receives an offer higher than p^L and then she cannot 'play', or, she can play the game as before.

From the point of view of the buyer H, his choice is just to choose between offering p^L (or more) or bargaining with the seller as usual (i.e. entering the classic game). Thus it will be quite simple: if he gets more by offering just p^L, he will do that, otherwise he won't. Buyer H simply has a choice to resort to an 'outside option' before the game even starts. The only trick is that buyer L is still present and can also offer more than p^L: thus we still have a competition between H and L, even if H wants to bid more than p^L. It means that we do not have a classic outside option $= p^L$ for buyer H, but instead an outside option $= max\{p^L, v\}$: because there is no equilibrium where H has to pay less than v, thus if $v \geq p^L$, then he will pay v instead of just p^L immediately (and obviously we have no concern about any bargaining since $v \geq p^L$ implies that the seller must accept immediately).

Basically we only have bargaining when $p^L > max\{\delta/(1+\delta), v\}$ now.

The equilibrium outcome is still that the seller accepts immediately (but sometimes because she *has to* accept) the initial offer from the buyer H at price p, where p is now defined as follows:

$$p = \begin{cases} \delta/(1+\delta), & \text{if } v \ \leq \ \delta/(1+\delta) \ \leq \ p^L \\ \\ v, & \text{if } \delta/(1+\delta) \ \leq \ v \ \leq \ p^L \\ & \text{if } p^L \ \leq \ v \\ \\ p^L, & \text{if } v \ \leq \ p^L \ \leq \ \delta/(1+\delta) \end{cases}$$

$$= max\left\{v, \ min\left(\delta/(1+\delta), \ p^L\right)\right\}$$

Now, if we assume that $\delta \to 1$ (either because the bargaining period $\tau \to 0$ or because both individual discount patience parameter $\rho \to 1$), we get an equal share of the pie $\delta/(1+\delta) = 0.5$. Moreover, if we rescale the problem to correspond to the one in the article, we get the same bargaining rule.

2.B Complete list of actual and simulated moments

Table 2.B.1: Actual and Simulated moments (complete table)

Moment	Actual	Simulated
Mean sale price	1.008	1.02
Mean ratio sale/final listing price	0.955	0.962
Mean initial list price	1.107	1.106
% of accepted offers equal to list price	0.15	0.235

% of accepted offers below list price	0.734	0.712
Mean week on the market (knowing that <52 weeks)	14.817	14.736
% unsold after 2 weeks	0.911	0.88
% unsold after 4 weeks	0.802	0.78
% unsold after 6 weeks	0.709	0.692
% unsold after 8 weeks	0.626	0.608
% unsold after 10 weeks	0.546	0.534
% unsold after 12 weeks	0.482	0.474
% unsold after 14 weeks	0.425	0.415
% unsold after 16 weeks	0.376	0.366
% unsold after 18 weeks	0.332	0.323
% unsold after 20 weeks	0.294	0.284
% unsold after 22 weeks	0.26	0.254
% unsold after 24 weeks	0.23	0.227
% unsold after 26 weeks	0.202	0.201
% unsold after 28 weeks	0.174	0.178
% unsold after 30 weeks	0.153	0.158
% unsold after 32 weeks	0.135	0.14
% unsold after 34 weeks	0.12	0.124
% unsold after 36 weeks	0.105	0.11
% unsold after 38 weeks	0.093	0.097
% unsold after 40 weeks	0.082	0.087
% unsold after 42 weeks	0.073	0.079
% unsold after 44 weeks	0.064	0.07
% unsold after 46 weeks	0.056	0.062
% unsold after 48 weeks	0.049	0.056
% unsold after 50 weeks	0.043	0.051

Mean list price in week 3	0.997	0.996
Mean list price in week 5	0.99	0.989
Mean list price in week 7	0.982	0.981
Mean list price in week 9	0.973	0.974
Mean list price in week 11	0.966	0.967
Mean list price in week 13	0.958	0.96
Mean list price in week 15	0.952	0.954
Mean list price in week 17	0.946	0.949
Mean list price in week 19	0.94	0.944
Mean list price in week 21	0.935	0.932
Mean list price in week 23	0.931	0.922
Mean list price in week 25	0.927	0.923
Mean list price in week 27	0.922	0.915
Mean list price in week 29	0.917	0.906
% sales/listing <0.70	0.014	0
% sales/listing <0.80	0.036	0
% sales/listing <0.85	0.064	0
% sales/listing <0.90	0.129	0.053
% sales/listing <0.92	0.186	0.213
% sales/listing <0.94	0.279	0.368
% sales/listing <0.95	0.344	0.436
% sales/listing <0.96	0.426	0.505
% sales/listing <0.97	0.522	0.564
% sales/listing <0.98	0.618	0.619
% sales/listing <0.99	0.697	0.669
% sales/listing <1.00	0.886	0.947
% sales/listing <1.02	0.949	0.959

% sales/listing <1.05	0.981	0.973
% sales/listing <1.10	0.992	0.988
% sales/listing <1.20	0.997	0.997

Chapter 3

Dynamics of Households' Consumption and Housing Decisions

Christophe Bruneel-Zupanc[1] and Thierry Magnac[2]

Abstract

We estimate a dynamic discrete and continuous choice model of households' decisions regarding their consumption, housing tenure and housing services over the life-cycle. We use non parametric identification arguments as in Bruneel-Zupanc (2021) to formulate an empirical strategy in two steps that (1) estimates discrete choice probabilities and continuous choice distribution summaries to be used in (2) Bellman and Euler equations that estimate the structural parameters. Specific modelling strategies are adopted because of unfrequent mobility due to housing transaction costs. Counterfactuals that can be evaluated are related to those transaction costs as well as of prudential policies such as downpayments.

Keywords: Dynamic models, discrete and continuous choices, non parametric identification, housing, policy evaluation.

JEL codes: C25, C61, D15, H31, R21

[1]E-mail address: christophe.bruneel@tse-fr.eu.
[2]E-mail address: thierry.magnac@tse-fr.eu

3.1 Introduction

Housing accounts for the bulk of assets of most households (61% out of an average wealth 276,000 € in France in 2018, Cazenave-Lacrouts et al., 2019). Housing is however a very specific asset since most of it also provides a non-monetary flow of housing services to its owner over time at the difference of many other assets (Li and Yao, 2007). Any shock on the value of housing may thus have strong impacts on consumption of other goods and consumer welfare (Ortalo-Magné and Rady, 2006). Moreover, housing is the target of sizeable public policies either in terms of household allowances or taxation. Transaction costs under the form of taxes on buying and selling houses are in France among the highest in the OECD and strongly affect the mobility of French households.[3]

In consequence, the evaluation of the impact of those transaction costs is high on the agenda of economists. For this, different types of dynamic models of housing are set up : partial equilibrium macro models such as Attanasio et al. (2012) or Li et al. (2016); general equilibrium macro models (Sommer and Sullivan, 2018; Bontemps et al., 2019); a sufficient statistics approach on the impact of prices (Berger et al., 2018; Etheridge, 2019). Those approaches nevertheless significantly restrict the heterogeneity between households. A few alternatives are offered by Bajari et al. (2013) and Khorunzhina and Miller (2019) using dynamic discrete choice models.

In this paper, we build upon these latter approaches and propose a dynamic continuous and discrete dynamic model whose non parametric identification has been recently analyzed by Bruneel-Zupanc (2021). We estimate a dynamic housing model in which the choice of housing tenure (ownership or renting), of housing services (e.g. house size) and consumption are the endogenous variables observed during several periods.

We follow Bajari et al. (2013) by having time-specific heterogeneity affecting preferences over consumption and housing services in a Constant Elasticity of Substitution (CES) set up. A second time-specific heterogeneity term among households governs the tenure decision between ownership and renting. In contrast, we follow a more standard Euler approach model by having consumption modeled with measurement errors only, which is typical of micro data in panels (Alan et al., 2009).

There are also various idiosyncrasies in housing data we take on board. First, the frequency of households moving at each period is rather small (between 5% for owners and 10% for renters) so that it creates selection that we deal with using information on consumption and restricting somewhat the dimension of heterogeneity. Second, when they move, some households are still

[3]https://www.globalpropertyguide.com/investment-analysis/Housing-transaction-costs-in-the-OECD

choosing a level of housing services which is very similar to the previous level. This should be hindered by the existence of transaction costs, so that to reconcile the data with the model, we introduce an exogenous shock to mobility. This could be caused by shocks to employment although those are not directly observed in the data.

Structural parameters are estimated in two steps as in Hotz and Miller (1993) although using the mix of discrete and continuous variables. In a first step, we estimate in a flexible way three static equations that concern housing services, consumption and housing tenure. At the second step, and using the approach suggested by non-parametric identification, we impose Euler and Belmann restrictions to recover structural parameters.

The data we use as the basis for our modelling strategy is the French extract of the European Survey of Income and Living Conditions (SILC) a 10-years rolling panel data set between 2004 and 2015. The advantages of this quite short panel is that it has reasonable good income and asset data, including house values, mortgages and variables related to labour earnings and benefits. Consumption however is to be reconstructed from asset and income data and this is also why we choose to model it with measurement errors.

As this paper remains preliminary, we present descriptive statistics, identification arguments and our empirical strategy. We do not report empirical results of the full estimation, nor the results of any counterfactual exercises. An interesting exercise would be to simulate the effects of the decrease in transaction costs on tenure, consumption and housing services as well as household welfare. Evaluating prudential policies such as down-payments, would also be an object of interest.

In the following, Section 3.2 briefly describes the data we intend and started to use. Section 3.3 sets up the theoretical framework. Identification arguments and the empirical strategy in two steps are presented in Section 3.4 and Section 3.5. Section 3.6 concludes.

Literature Review: The closest studies in a continuous and discrete dynamic set-up such as to ours are Bajari et al. (2013) and Li et al. (2016). They are both based on the now canonical dynamic model of housing in which housing tenure, housing services and consumption are modelled in a context in which credit constraints are important (e.g. Li and Yao, 2007; Attanasio et al., 2012). Bajari et al. (2013) uses the Panel Study of Income Dynamics data and a two-step method proposed in previous work by the first author (Bajari et al., 2007). It differs from our own two-step method in the second stage and consists in setting up moment conditions that use that some deviations from observed decisions are suboptimal. In contrast, we use all restrictions of the model in the second stage and the conditions on the way heterogeneity terms enter the model are clearer than theirs.

Li et al. (2016) estimates the parameters of their housing model by the simulated method of moments (SMM), the standard estimation method in the literature at the intersection of consumption studies and partial equilibrium macroeconomics. Among other results, these authors estimate the elasticity of substitution between non-durable consumption and housing services using a Constnat Elasticity of Substitution (CES) specification for preferences and find that this substitution elasticity is significantly lower than one. In this vein, Khorunzhina (2021) provides an interesting way of identifying this substitution elasticity by using maintenance expenses in the PSID. It allows an household specific price index to be constructed and used as exogenous variation affecting the ratio of consumption and housing services. This identifies this substitution elasticity, at least for owners staying in the same house.

Bruneel-Zupanc (2021) provides the main identification arguments used in our paper, as well as an empirical application to the canonical dynamic model of household consumption and labour market participation. He does not analyze a two-consumption good case as we do and does not consider measurement errors in consumption. Our procedure also relates to non linear GMM estimation of the Euler equation as in Alan et al. (2009).

In terms of economic results, our paper belongs to the strand of economic studies analyzing the mitigation of income risks by households through intertemporal smoothing of consumption (Blundell et al., 2008). We can indeed use our results in order to model the degree of insurance that households can achieve at various points over their life-cycle. Furthermore, we can also compare our results to recent analyses by Browning et al. (2013) on the impact of house prices on consumption. The authors find that the effect is mainly due the collateral value of housing and not to a wealth effect.

Berger et al. (2018) also tries to assess the impact of housing prices on consumption and finds a significant effect that can be summarized by a simple "sufficient statistics" which is the product of the household marginal propensity to consume and the value of their housing. In a close contribution, Etheridge (2019) uses the same decomposition exercise as Blundell et al. (2008) in the case of linear income and housing risks and shows that a positive common shock to house prices in the UK increases consumption inequality in cross-section. His empirical conclusions also hold in a more non-linear structural model. In particular, income and housing risk interactions are shown to be important to understand consumer behaviour because increases in house price alleviate borrowing constraints whereas decreases strengthen them. Paz-Pardo (2021) is another example of recent analyses on how households deal with risks arising either from their incomes or assets. Using the PSID, the author shows that changes in the dynamics of income account for a large part of the recent reduction of homeownership by young households and that these investments are not compensated by other assets in their portfolios.

Other key dimensions of the decision making of households are not modelled here although some are developed in the recent literature. Blundell et al. (2016b) analyzes how family labour supply adjusts to directly mitigate risks in income. Öst (2012) models the simultaneous housing and fertility decisions in a reduced form setting, and finds a positive correlation between homeownership and fertility in particular for younger households. In a more structural set-up, Khorunzhina and Miller (2019) models how households choose homeownership, fertility and labor supply. Interestingly, they use a two-step method in a dynamic discrete choice model (Hotz and Miller, 1993) and this is the inspiration for our own estimation method although it also uses continuous choices although in a setting where labour participation and hours are exogenous.

The importance of credit constraints in relation to housing tenure has been shown repetitively as in the early work of Ortalo-Magné and Rady (2006). Pizzinelli (2017) is a more recent example studying the interaction between prudential regulations on credit – imposing loan to value and loan to income ratios to household mortgages – and labour supply of households, and their impact on homeownership. This literature, as well as Berger et al. (2018), points out that the leverage position of households severely restricts their ability to smooth income risks. Iacoviello and Pavan (2013) also shows the importance of loan to value ratio, or downpayment constraints, as well as the importance of non-convex adjustments costs in housing models. The very large transaction costs in France when purchasing a house dampens liquidity and is a important explanatory factor of the low level of mobility across houses for owners.

Finally, our results could also be compared to those of studies using French micro-data and that evaluate housing policies. Housing allowances, zero-interest loans, housing tax credit, real-estate transaction cost and residence tax have been the focus of such policy evaluations in recent years. Grislain-Letremy and Trévien (2014) estimates the impact of housing allowances on prices of housing services in France between 1987 and 2012 and confirms the inflationary effect that Fack (2006) uncovered. Zero-interest access-to-ownership loans are evaluated by Gobillon and le Blanc

(2008). Housing tax credits and their inflationary effects are evaluated uisng difference-in-difference methods by Bono and Trannoy (2019) and Chapelle et al. (2018). Bérard and Trannoy (2018) analyses the impact on prices and quantities of an increase of the transaction tax in various local areas in 2014. Bontemps et al. (2019) uses a Sommer and Sullivan (2018) general equilibrium model to analyze counterfactuals such as the impact of a decrease in transaction costs and housing taxes. A reduced form analysis of dynamic housing and labour-market participation decisions is provided by Kamionka and Lacroix (2018) and uses the same data as in our study. Although they model income in a richer way than we do, they do not estimate the structural parameters of a housing model.

3.2 Data

We use the French extract of the European Survey of Income and Living Conditions (EU-SILC), a ten-years rolling panel dataset between 2004 and 2015. We select only couples that stayed together during the survey, in order to avoid modeling the merge or division of assets when couples are formed or when they divorce. We focus on individuals less than 60 years old as retirees may face very different housing market condition (borrowing in particular).

After cleaning the data, we are left with 7,108 unique couples, giving 22,625 household-year observations. The descriptive statistics of the data, by ownership status, are given in Table 3.2.1. 73% of the household own their properties. On average homeowners are older than renters, they consume more, live in larger houses and get less benefits. The housing mobility is quite low in the French data: overall, only 7% of the households change their residence during a year, i.e. we observe a total of 1,542 moves. It suggests that there is a very high cost of moving for the households.

	All		***Owner***		***Renter***	
	Mean	(SD)	Mean	(SD)	Mean	(SD)
Consumption (c)	39.18	(40.68)	40.56	(43.85)	35.46	(30.20)
House size in sqm (s)	107.98	(37.47)	117.13	(36.85)	83.17	(26.22)
Homeownership (d)	0.73	(0.44)	1.00	(0.00)	0.00	(0.00)
Past Homeownership	0.71	(0.45)	0.97	(0.18)	0.02	(0.13)
House price/rent (yearly) in K euros			228.91	(92.22)	6.08	(2.74)
Est. House value			221.89	(69.10)	6.10	(2.18)
Moved this year	0.07	(0.25)	0.05	(0.22)	0.12	(0.32)
Age	42.67	(9.36)	43.81	(9.07)	39.57	(9.42)
Financial Asset	24.11	(51.16)	27.89	(55.23)	13.85	(36.07)
Landlord Asset	32.54	(90.42)	38.29	(97.06)	16.99	(66.87)
Income	45.02	(20.67)	48.21	(21.06)	36.37	(16.75)
Benefits	4.18	(5.32)	3.58	(4.83)	5.79	(6.19)
Housing Benefits	0.41	(1.08)	0.17	(0.66)	1.07	(1.61)
Live in Social Housing	0.11	(0.31)	0.00	(0.00)	0.41	(0.49)
Number of children	1.35	(1.12)	1.38	(1.10)	1.26	(1.17)
Number of children below 3 y.o.	0.24	(0.48)	0.21	(0.46)	0.32	(0.53)
Number of children below 6 y.o.	0.49	(0.73)	0.45	(0.71)	0.59	(0.76)
Number of children below 18 y.o.	1.30	(1.13)	1.30	(1.10)	1.29	(1.20)
Age of the youngest child	7.55	(5.65)	8.09	(5.68)	5.99	(5.26)
Observations	22625		16525		6100	

Table 3.2.1: Descriptive Statistics of the complete sample and by homeownership status.

d_{-1} / d	0	1	Total
0	597	565	1162
	(51.4)	(48.6)	(100.0)
1	109	271	380
	(28.7)	(71.3)	(100.0)
Total	706	836	1542
	(45.8)	(54.2)	(100.0)

Table 3.2.2: Housing Tenure (d) change when moving

Renters are more mobile (12%) than homeowners (5%). As described in Table 3.2.2, 48.6% of renters who move are becoming homeowners, while homeowners are considerably less likely to become renters (71.3% become owner of a new house and only 28.7% become renters). This suggests some differential cost of switching tenures for households who were owners and household who were renters.

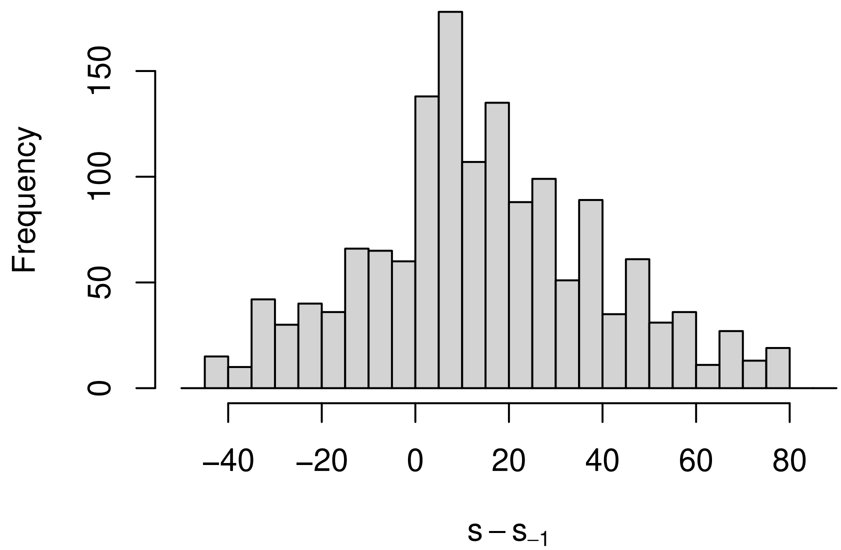


Figure 3.2.1: Histogram of $s - s_{-1}$ conditional on moving and **same tenure**

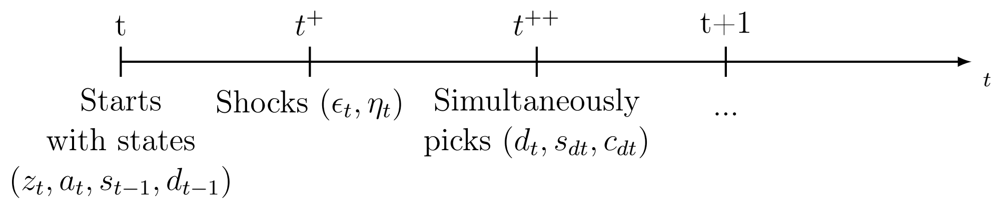
Finally, as shown in Figure 3.2.1, conditional on moving, people move to larger houses, with an average house size change ($s - s_{-1}$) of about 20 square meters.[4] Surprisingly though, we observe that some households also move to very similar properties. This is at odds with the high cost of moving, and it suggests that some households are probably moving for some non-housing reasons, e.g. labor mobility shocks.

3.3 Framework

We build a dynamic life-cycle model from 25 to 60 y.o. which incorporates all the empirical facts described in the previous section.

Each period, the timing of the problem is as follows:

[4] We also observe the same pattern when using the house value adjustment: on average households move to houses with higher estimated value.



Each period, the household decide of its housing tenure (d, $= 1$ for owners, $= 0$ for renters), housing services (s_d) and non-housing consumption (c).[5] They do so given three endogenous states: their previous housing tenure (d_{-1}) and housing services (s_{-1}), their asset (a). And also given some exogenous states z, including the household income and other demographics (x) which mainly include the age, the number of children and their age. They also know prices (and forecast them): p for rent, q for house prices and interest rate (r). Two iid shocks occur each period. They are unobserved by the econometrician, but known by the households. Additively separable shocks ϵ affect the housing-tenure choice. ϵ follows an extreme value type 1 distribution, as in the discrete choice literature (McFadden, 1980; Hotz and Miller, 1993). A non-separable shock η affects the substitution between housing services and non-housing consumption. η also affect the housing-tenure choice. As its distribution is not identified, we normalize $\eta \sim \mathcal{U}(0, 1)$.

Finally, there is a random *exogenous binary move shock* m. With probability p^m, $m = 1$ and the household must move for exogenous reasons (e.g. labor mobility shock).[6] When $m = 0$ (with probability $1 - p^m$), the household can choose freely to move or not, yielding endogenous moves. m is not observed by the econometrician. We include it to capture the fact that some households move to very similar houses, even though the fixed cost of moving should be high.

Period utility function:

[5]The data counterpart to housing services s can either be the housing size in square meters s, or other housing variables such as the housing value. The advantage of our identification method is that s_d can be a different variable for each d, e.g. for owners s_1 is the estimated housing prices and for renters s_0 is the rent.

[6]We can allow for a p^m that depend on our variables, especially the demographics and the income.

The utility function take a CES form (close to Li et al., 2016) and is given by

$$
\begin{aligned}
U(d, c, s, s_{-1}, d_{-1}, x, m, \eta, \epsilon) &= \gamma(x) \frac{\left(c^{1-1/\gamma_d} + \phi(\eta) s^{1-1/\gamma_d}\right)^{\frac{1-\sigma}{1-1/\gamma_d}}}{1-\sigma} \\
&+ (\underbrace{\mu_0 + \mu_1 d_{-1}}_{\text{fixed moving cost}}) (\mathbf{1}\{\{d \neq d_{-1}\} \text{ or } \{s \neq s_{-1}, d = d_{-1}\} \text{ or } \{m = 1\}\}) \\
&+ (\underbrace{\nu_0 + \nu_1 d_{-1}}_{\text{tenure swapping cost}}) \mathbf{1}\{d \neq d_{-1}\} + \epsilon_d
\end{aligned}
$$

$\gamma(x)$ is an equivalence scale as a function of the demographics x. c and s are non separable choices under this CES specification. The shock η mainly affects the substitution between c_d and s_d and $\phi(\eta)$ is simply a strictly increasing transformation with respect to η (e.g. $\phi(\eta) = a + b\eta$) to get more general effects of the shock. The CES parameters γ_d and σ are such that $\partial U(\cdot)/\partial s_d \partial \eta > 0$ and $\partial U(\cdot)/\partial c_d \partial \eta < 0$. It means that the optimal housing service consumption (s_d^*) is strictly increasing with η and the optimal non-housing consumption (c_d^*) is strictly decreasing with it: η governs the substitution between c and s.[7]

You also have some additively separable costs in the utility function. In the model, we observe that an household move if $d \neq d_{-1}$ or $s \neq s_{-1}$ and $d = d_{-1}$. In this case, the household endures a fixed moving cost $\mu_0 + \mu_1 d_{-1}$. With $\mu_0 < 0$ and $\mu_0 + \mu_1 < 0$ because it is costly to move. This cost depends on the past tenure, as suggested in the data where owners do not move as much as renters (suggesting $\mu_1 < 0$). This fixed utility cost will yield optimal choice where individuals will choose not to move in the model, and stay in their current house: $s = s_{-1}$ and $d = d_{-1}$.[8] Individuals are also forced to move when $m = 1$. In this case, they endure the fixed utility cost, which becomes irrelevant for their choices because they face it no matter what they do. Thus, these individuals are not 'constrained' to choose $s = s_{-1}$ by the fixed cost (and there is a zero likelihood that they will pick exactly $s = s_{-1}$ since s is a continuous choice). The exogenous move friction m only enters the household problem here.

Moreover, there is also a tenure swapping cost $\nu_0 + \nu_1 d_{-1}$. We expect $\nu_0 < 0$ and $\nu_0 + \nu_1 < 0$ because it is costly to swap tenure. And probably $\nu_1 < 0$ because it is even more costly for previous home-owners to become renters: as observed in the data, homeowners are less likely to become renters.

[7]Note that we can also have a specification with complementary goods that would both increase with respect to η. We only need both choices to be stricly monotone with respect to η, increasing or decreasing.

[8]Notice that we also include monetary transaction costs in the model, which also explains part of the staying behaviour.

Transitions:

The *budget constraint* is:

$$
\begin{aligned}
a_{+1} \quad \leq \quad & (1+r)a + \text{income} - c - \underbrace{T(x, d, s, d_{-1}, s_{-1})}_{\text{Tax}} \\
& -\underbrace{ps(1-d)}_{\text{Net rent}} - \underbrace{q(s-s_{-1})dd_{-1}}_{\text{Price of house change}} - \underbrace{qsd(1-d_{-1})}_{\text{Price of new house}} + \underbrace{qs(1-d)d_{-1}}_{\text{Gains from selling and renting}} .
\end{aligned}
$$

Notice that d_{-1} enters the budget constraint as previous owners ($d_{-1} = 1$) are potentially richer than previous renters ($d_{-1} = 0$) because they can sell their housing asset when they move. It also enters the tax schedule. The tax (benefits) schedule $T()$ can be modelled realistically from the French law to include income tax, payroll tax, residential tax, property tax, housing sale and purchase tax and housing benefits.

The *transitions* of the other variables: household income and family demographics are exogenous in the model. Households can exogenously have a new child. They cannot divorce in this model. And we focus on couples so there is no couple formation either. This simplifies the asset and housing transition in cases of divorce or couple formation. Transitions of income and demographics depends on current income and demographics, not on s_{-1} or d_{-1}.

Dynamic life-cycle problem:

Let's introduce the dynamic optimization problem of the households. Slightly change the notation to add the index t to each variable. Here the households choose (d_t, s_{dt}, c_{dt}) in order to sequentially maximize their discounted sum of payoffs, with discount factor β. Let's define $V_t(z_t, d_{t-1}, s_{t-1})$ the ex ante value function of this discounted sum of future payoffs at the beginning of period t, just before the shocks (ϵ_t, η_t) are revealed and conditional on behaving according to the optimal decision rule afterwards.

$$
V_t(z_t, a_t, d_{t-1}, s_{t-1}) = \mathbb{E}\Big[\sum_{\tau=t}^{T} \beta^{\tau-t} \underset{d_\tau, s_{d\tau}, c_{d\tau}}{max} U(d_\tau, c_{d\tau}, s_{d\tau}, s_{\tau-1}, d_{\tau-1}, x_\tau, m_\tau, \eta_\tau, \epsilon_\tau)\Big]
$$

$$
\begin{aligned}
\text{s.t.} \quad & a_{t+1} \leq (1+r_t)a_t + \text{income}_t - c_t - T(x_t, d_t, s_t, d_{t-1}, s_{t-1}) \\
& \quad - p_t s_t(1-d_t) - q_t(s_t - s_{t-1})d_t d_{t-1} - q_t s_t d_t(1-d_{t-1}) + q_t s_t(1-d_t)d_{t-1}
\end{aligned}
$$

Thus, each period, the household chooses d, s_d and c_d to maximize their expected sum of payoffs:

$$\underset{d_t,s_{dt},c_{dt}}{max}\ U_t(d_t, c_{dt}, s_{dt}, s_{t-1}, d_{t-1}, x_t, m_t, \eta_t, \epsilon_t) + \beta \mathbb{E}_{z_{t+1}}\left[V_{t+1}(z_{t+1}, a_{t+1}, d_t, s_t)\middle| z_t, a_t, s_{dt}, c_{dt}, d_t, d_{t-1}, s_{t-1}\right].$$

Remember that z_t include the exogenous states, i.e. the demographics x_t, but also the income and the prices.

Retirement:

At 60 years old, we assume the household retire, and live off of their pension and wealth (housing and non-housing) for 15 more years.

Measurement errors:

Moreover, the non-housing consumption is observed with *measurement error* ζ by the econometrician, i.e.

$$c^{\text{obs}} = c^* + \zeta.$$

But ζ does not enter the household problem. The measurement error is independent from every other variables, and is iid every periods.

3.4 Identification

At each period, we observe data on the variables $(d, s_d, c_d, s_{-1}, d_{-1}, a, z)$. We only observe s_0 and c_0 if $d = 0$ and s_1 and c_1 if $d = 1$. In other words:

$$s_d = s_0(1-d) + s_1 d$$

$$c_d = c_0(1-d) + c_1 d.$$

Shocks m, η and ϵ are observed by the agents but not observed by the econometrician. Moreover, a noisy measure of c_d is observed because of *measurement errors*, i.e. we observe $c_d + \zeta$.

We first study identification of the following objects: the optimal *Conditional Continuous Choices*

(CCCs) $s^*_d(\eta, s_{-1}, d_{-1}, a, z, m)$ and $c^*_d(\eta, s_{-1}, d_{-1}, a, z, m)$, as well as the optimal *Conditional Choice Probabilities* (CCPs) $Pr(d|\eta, s_{-1}, d_{-1}, a, z, m)$. Then, following Bruneel-Zupanc (2021) we will use the identified optimal choices to identify the structural parameters of the model.

The identification proof will use arguments of Bruneel-Zupanc (2021). However, this paper cannot be applied directly as the setup here is more complicated for three main reasons. (i) d_{-1} violates the *exclusion restriction* of Bruneel-Zupanc (2021), and is not directly an instrument. Indeed, it affects the budget constraint and thus the future value. (ii) Because of fixed costs, the optimal choice s will not be strictly monotone. In a related note, the presence of unobserved m binary shock implies there are additional objects to identify. (iii) Consumption is an additional choice and this was not dealt with in the previous paper.[9] We show how to identify it if it is measured with errors.

To show how the identification works here, we proceed stepwise. First we show how the optimal choices would be identified if everyone was moving freely, i.e. if the fixed cost was irrelevant (everyone pays it even if they do not move). Then we show how the optimal choices are identified with a fixed moving cost in the model that makes some households stay in their previous house. Finally, we show how the consumption choice is also identified despite the measurement errors and how it can be used to identify the housing tenure probability choice even for stayers.
Without loss of generality, we proceed conditional on any given exogenous state z in this section, and we omit z in what follows.

3.4.1 Identification when everyone moves

Assume everyone moves. In other words, everyone pays the fixed cost, even if they do not move, such that households will no longer choose $s = s_{-1}$ to avoid paying the fixed cost. The utility function is then written as:

$$\begin{aligned} U(d, c, s, s_{-1}, d_{-1}, x, m, \eta, \epsilon) \quad &= \quad \gamma(x) \frac{\left(c^{1-1/\gamma_d} + \phi(\eta) s^{1-1/\gamma_d}\right)^{\frac{1-\sigma}{1-1/\gamma_d}}}{1-\sigma} \\ &+ \quad (\mu_0 + \mu_1 d_{-1}) \ + \ (\nu_0 + \nu_1 d_{-1}) \mathbf{1}\{d \neq d_{-1}\} \ + \ \epsilon_d \end{aligned}$$

[9]Notice that s_d in this paper would be the direct counterpart to c_d in Bruneel-Zupanc (2021). Because c_d is measured with additional errors, so it cannot be used directly to recover η.

Very importantly, notice that this is *equivalent to the general setup when* $m = 1$, i.e. where households are forced to move. A model without any fixed cost ($\mu_0 + \mu_1 d_{-1}$) term would also be equivalent in terms of optimal choices. Indeed, the fixed cost does not differentially affect the discrete alternatives so the optimal choice d^* does not depend on it. Furthermore, by additive separability, s_d^* and c_d^* are also independent from it.

Even when everyone moves, we cannot apply the strategy employed by Bruneel-Zupanc (2021) directly. Indeed, d_{-1} is not excluded from the budget constraint. So the optimal continuous choices s_d^* and c_d^* will depend on d_{-1} since, everything else equal, a previously homeowner is richer than a previously renter.

We adopt the following solution in this specific setup. One can show that d_{-1} only affect housing wealth in the budget constraint when everyone moves. Denote $\tilde{s}_d^*$ the optimal choices in this setup where everyone moves, or the *optimal housing service choice conditional on moving*. We can also call it the *unconstrained optimal choice*, in the sense that it is *not constrained by the fixed cost*. One can show that for all η, d and s'_{-1}:

$$\tilde{s}_d^*(\eta, d_{-1} = 1, s_{-1}, a) = \tilde{s}_d^*(\eta, d_{-1} = 0, s'_{-1} = \cdot, a' = a + q\frac{s_{-1}}{1+r}). \tag{3.1}$$

In other words, conditional on moving, ex-homeowners with financial asset a and house of size s_{-1} will have a *total wealth* equal to $(1+r)a + qs_{-1}$ (plus their income and taxes, that we abstract from here). They will make the same choice as ex-renters with the same total wealth, i.e. with financial asset $a' = a + qs_{-1}/(1+r)$. Because ex-renters ($d_{-1} = 0$) do not own their house, their previous house size s'_{-1} do not affect their wealth, so the property holds for all s'_{-1}. The property holds because s_{-1} only matters in terms of wealth for the previous homeowners without fixed cost of moving here. Apart from the housing wealth it provides them when they move, previous homeowners are exactly equivalent to previous renters here. So, *conditional on* ***total wealth***, $(1+r)a + qs_{-1}d_{-1}$, the optimal housing service choice $\tilde{s}_d^*$ is independent from d_{-1}. It is true because d_{-1} only matters in determining the total wealth, it does not play a role anywhere else in the other variables transition.

Notice also that d_{-1} is (strictly) *relevant* for the choice to be a homeowner or a renter at a given total wealth. Indeed, because of the *tenure swapping cost* $(\nu_0 + \nu_1 d_{-1})\mathbb{1}\{d \neq d_{-1}\}$, we have that:

$$Pr(D = 1|\eta, d_{-1} = 1, s_{-1}, a) > Pr(D = 1|\eta, d_{-1} = 0, s'_{-1} = \cdot, a' = a + q\frac{s_{-1}}{1+r}), \tag{3.2}$$

in which the strict inequality comes from the fact that $\nu_0 + \nu_1 d_{-1} < 0$ for all d_{-1}, i.e. it is costly to swap tenure, no matter the previous tenure. In other words, at equal total wealth, ex-homeowners are strictly more likely to be homeowners today than ex-renters.

Therefore, even if we cannot count on an exclusion restriction as defined in Bruneel-Zupanc (2021), we can apply refined version of the identification proof in the paper using property (3.1). In short, the idea of the proof is to match ex-homeowners ($d_{-1} = 1$) with endogenous states (s_{-1}, a) to ex-renters ($d_{-1} = 0$) with endogenous states yielding the same total wealth, i.e. $a' = a + qs_{-1}/(1+r)$ (and any s'_{-1}), instead of matching them with ex-renters with the same covariates ($a' = a$ and $s'_{-1} = s_{-1}$). For these pairs, we have the 'exclusion' and the 'relevance' of d_{-1}, conditional on the total wealth, and we can apply the same reasoning as in Bruneel-Zupanc (2021). We show how it works in what follows.

Sketch of the proof:

When $d_{-1} = 1$, *conditional on moving*, we have:

$$\begin{aligned}
h = \quad & Pr(\eta \leq h) \\
= \quad & Pr(\eta \leq h | s_{-1}, a, d_{-1} = 1) \\
= \quad & Pr(\eta \leq h | D = 0, s_{-1}, a, d_{-1} = 1) Pr(D = 0 | s_{-1}, a, d_{-1} = 1) \\
& + Pr(\eta \leq h | D = 1, s_{-1}, a, d_{-1} = 1) Pr(D = 1 | s_{-1}, a, d_{-1} = 1) \\
= \quad & Pr\Big(s \leq \tilde{s}_0^*(h, s_{-1}, a, d_{-1} = 1) | D = 0, s_{-1}, a, d_{-1} = 1\Big) \; Pr(D = 0 | s_{-1}, a, d_{-1} = 1) \\
& + Pr\Big(s \leq \tilde{s}_1^*(h, s_{-1}, a, d_{-1} = 1) | D = 1, s_{-1}, a, d_{-1} = 1\Big) \; Pr(D = 1 | s_{-1}, a, d_{-1} = 1) \\
= \quad & F_{S_0 | D=0, s_{-1}, a, d_{-1}=1}(\tilde{s}_0^*(h, s_{-1}, a, d_{-1} = 1)) \; Pr(D = 0 | s_{-1}, a, d_{-1} = 1) \\
& + F_{S_1 | D=1, s_{-1}, a, d_{-1}=1}(\tilde{s}_1^*(h, s_{-1}, a, d_{-1} = 1)) \; Pr(D = 1 | s_{-1}, a, d_{-1} = 1),
\end{aligned} \tag{3.3}$$

where the first equality comes from $\eta \sim \mathcal{U}(0, 1)$ by normalization. The second is because $\eta \perp (d_{-1}, a, s_{-1})$. The third equality comes from the law of total probability. The fourth equality comes from the strict monotonicity of $\tilde{s}_d^*(\eta, \cdot)$ with respect to η. The fifth equality is just a notation change for the conditional distribution functions.

Similarly, when $d_{-1} = 0$, conditional on moving, we have:

$$\begin{aligned}
h = \quad & Pr(\eta \leq h) \\
= \quad & Pr(\eta \leq h | s'_{-1}, a', d_{-1} = 0) \\
= \quad & Pr(\eta \leq h | D = 0, s'_{-1}, a', d_{-1} = 0) Pr(D = 0 | s'_{-1}, a', d_{-1} = 0) \\
& + Pr(\eta \leq h | D = 1, s'_{-1}, a', d_{-1} = 0) Pr(D = 1 | s'_{-1}, a', d_{-1} = 0) \\
= \quad & Pr\Big(s \leq \tilde{s}^*_0(h, s'_{-1}, a', d_{-1} = 0) | D = 0, s'_{-1}, a', d_{-1} = 0\Big)\ Pr(D = 0 | s'_{-1}, a', d_{-1} = 0) \\
& + Pr\Big(s \leq \tilde{s}^*_1(h, s'_{-1}, a', d_{-1} = 0) | D = 1, s'_{-1}, a', d_{-1} = 0\Big)\ Pr(D = 1 | s'_{-1}, a', d_{-1} = 0) \\
= \quad & F_{S_0 | D = 0, s'_{-1}, a', d_{-1} = 0}(\tilde{s}^*_0(h, s'_{-1}, a', d_{-1} = 0))\ Pr(D = 0 | s'_{-1}, a', d_{-1} = 0) \\
& + F_{S_1 | D = 1, s'_{-1}, a', d_{-1} = 0}(\tilde{s}^*_1(h, s'_{-1}, a', d_{-1} = 0))\ Pr(D = 1 | s'_{-1}, a', d_{-1} = 0),
\end{aligned}$$

for the same reasons as what we have for $d_{-1} = 1$.[10]

Now, recall property (3.1) that, for all η, d and s'_{-1}:

$$\tilde{s}^*_d(\eta, d_{-1} = 1, s_{-1}, a) = \tilde{s}^*_d(\eta, d_{-1} = 0, s'_{-1} = \cdot, a' = a + q\frac{s_{-1}}{1 + r}). \tag{3.4}$$

Thus, *conditional on moving*, with $a' = a + q s_{-1}/(1 + r)$, we have:

$$\begin{aligned}
h = \quad & F_{S_0 | D = 0, s_{-1}, a, d_{-1} = 1}(\tilde{s}^*_0(h, s_{-1}, a, d_{-1} = 1))\ Pr(D = 0 | s_{-1}, a, d_{-1} = 1) \\
& + F_{S_1 | D = 1, s_{-1}, a, d_{-1} = 1}(\tilde{s}^*_1(h, s_{-1}, a, d_{-1} = 1))\ Pr(D = 1 | s_{-1}, a, d_{-1} = 1) \\
= \quad & F_{S_0 | D = 0, s'_{-1}, a', D_{-1} = 0}(\tilde{s}^*_0(h, s'_{-1}, a', d_{-1} = 0))\ Pr(D = 0 | s'_{-1}, a', d_{-1} = 0) \\
& + F_{S_1 | D = 1, s'_{-1}, a', d_{-1} = 0}(\tilde{s}^*_1(h, s'_{-1}, a', d_{-1} = 0))\ Pr(D = 1 | s'_{-1}, a', d_{-1} = 0).
\end{aligned}$$

[10]Notice here that $\tilde{s}^*_d(\cdot, d_{-1} = 0) \perp s_{-1}$. We see later that s_{-1} only matters to determine the position of the mass point of the optimal s when the household does not move if $d_{-1} = 0$. So, for the conditional distribution function of s, we could remove s'_{-1}.

Therefore, we can proceed as in Bruneel-Zupanc (2021), and rewrite it as:

$$
\begin{aligned}
&\Bigg(F_{S_0|D=0,s_{-1},a,D_{-1}=1}(\tilde{s}_0^*(h,s_{-1},a,d_{-1}=1))\ Pr(D=0|s_{-1},a,D_{-1}=1)\\
&\quad - F_{S_0|D=0,s'_{-1},a',D_{-1}=0}(\tilde{s}_0^*(h,s'_{-1},a',d_{-1}=0))\ Pr(D=0|s'_{-1},a',D_{-1}=0)\Bigg)\\
= -\quad &\Bigg(F_{S_1|D=1,s_{-1},a,D_{-1}=1}(\tilde{s}_1^*(h,s_{-1},a,d_{-1}=1))\ Pr(D=1|s_{-1},a,D_{-1}=1)\\
&\quad - F_{S_1|D=1,s'_{-1},a',D_{-1}=0}(\tilde{s}_1^*(h,s'_{-1},a',d_{-1}=0))\ Pr(D=1|s'_{-1},a',D_{-1}=0)\Bigg).
\end{aligned}
$$

Under property (3.1), we have a mapping between the policy functions with different previous housing wealth but similar total wealth. Thus we have only two policy functions to identify (instead of four, two for each d), with four different conditional distributions thanks to the relevance of d_{-1}. It yields

$$
\begin{aligned}
&\Bigg(F_{S_0|D=0,s_{-1},a,D_{-1}=1}(\tilde{s}_0^*(h,s_{-1},a,d_{-1}=1))\ Pr(D=0|s_{-1},a,D_{-1}=1)\\
&\quad - F_{S_0|D=0,s'_{-1},a',D_{-1}=0}(\tilde{s}_0^*(h,s_{-1},a,d_{-1}=1))\ Pr(D=0|s'_{-1},a',D_{-1}=0)\Bigg)\\
= -\quad &\Bigg(F_{S_1|D=1,s_{-1},a,D_{-1}=1}(\tilde{s}_1^*(h,s_{-1},a,d_{-1}=1))\ Pr(D=1|s_{-1},a,D_{-1}=1)\\
&\quad - F_{S_1|D=1,s'_{-1},a',D_{-1}=0}(\tilde{s}_1^*(h,s_{-1},a,d_{-1}=1))\ Pr(D=1|s'_{-1},a',D_{-1}=0)\Bigg)\\
\iff \quad & \Delta F_{S_0}(\tilde{s}_0^*(h,s_{-1},a,d_{-1}=1)) = \Delta F_{S_1}(\tilde{s}_1^*(h,s_{-1},a,d_{-1}=1)). \qquad (3.5)
\end{aligned}
$$

In the model when everyone moves, the functions ΔF_{S_d} can be *directly estimated*, from the data (as conditional distributions). It only remains to identify policies: $\tilde{s}_d^*(h,s_{-1},a,d_{-1}=1)$. One can directly apply Bruneel-Zupanc (2021) to show that functions $\tilde{s}_d^*(h,s_{-1},a,d_{-1}=1)$ are identified as the unique solution to this system of equation (3.5).

In fact here, because we have *strict* relevance (3.2), one can show that ΔF_{S_0} and ΔF_{S_1} are invertible. Thus, the mapping between $\tilde{s}_0^*$ and $\tilde{s}_1^*$ is identified directly. Indeed, since $\tilde{s}_d^*$ are strictly monotone with respect to η, we can rewrite (3.5) as:

$$\Delta F_{S_0}(\tilde{s}_0^*(s_1, s_{-1}, a, d_{-1} = 1)) = \Delta F_{S_1}(s_1).$$

And there is a unique mapping between s_1 and s_0, obtain by taking the inverse:

$$\tilde{s}_0^*(s_1, s_{-1}, a, d_{-1} = 1) = \Delta F_{S_0}^{-1}\Big(\Delta F_{S_1}(s_1)\Big)$$

Now, to obtain $\tilde{s}_d^*$ as a function of η for both d, simply plug these into (3.3).
Therefore, in the model with only movers ($m = 1$), the optimal housing service $\tilde{s}_d^*(h, s_{-1}, a, d_{-1} = 1)$ are *identified* for all d, η, s_{-1} and a.

Moreover, these policies are strictly increasing, so one can invert them to recover the unobserved η for each observation of s_d in the data.

$$\eta = (\tilde{s}_d^*)^{-1}(s_d^{obs}, s_{-1}, a, d_{-1})$$

From there, it is as if $\eta = h$ were observed. We can use it to compute the conditional choice probabilities (CCPs)

$$Pr(D = 1|\eta, d_{-1}, s_{-1}, a).$$

Thus, the optimal discrete choice probabilities are also identified.

As for the consumption choice $\tilde{c}_d^*(h, s_{-1}, a, d_{-1})$, we can also use the inversion of $\tilde{s}_d^*(\eta, \cdot)$ to recover η and proceed as if it was observed. Thus, one just need to run the non parametric regression

$$c_d^{obs} = \tilde{c}_d^*(\eta, s_{-1}, a, d_{-1}) + \zeta,$$

which directly identifies the optimal consumption policies and the distribution of the measurement errors ζ.

Therefore, the three policy choices are identified without fixed moving costs in a model where everyone moves.

3.4.2 Identification with fixed costs, when some households do not move

In a model with fixed costs, when not everyone is moving, the identification will be more complicated because some households will choose not to move, and select $s_d^* = s_{-1}, d^* = d_{-1}$.

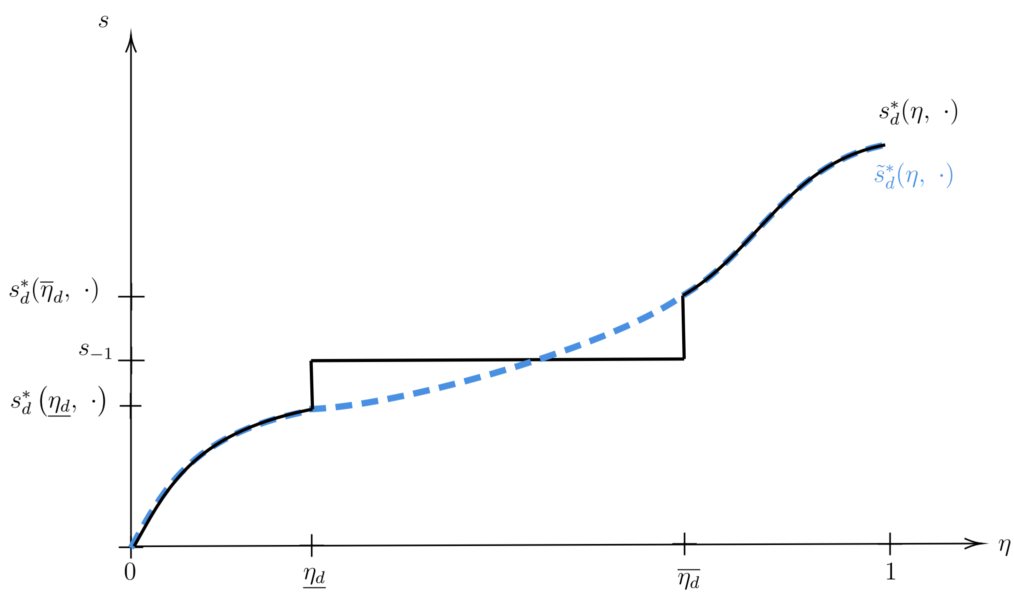


Figure 3.4.1: Optimal housing services function

The presence of the fixed cost transforms the problem as shown in Figure 3.4.1. Between two values of η, $\underline{\eta}_d(a, s_{-1}, d_{-1})$ and $\bar{\eta}_d(a, s_{-1}, d_{-1})$, with $0 \leq \underline{\eta} \leq \bar{\eta} \leq 1$, households will choose to stay and $s_d^*(\eta, a, s_{-1}, d_{-1}) = s_{-1}$. Outside the interval $[\underline{\eta}_d(a, s_{-1}, d_{-1}), \bar{\eta}_d(a, s_{-1}, d_{-1})]$, households will move, and select the optimal housing service defined in the previous section: $\tilde{s}_d^*(\eta, \cdot)$. Notice that it is possible that $\underline{\eta}_d = 0$ or $\bar{\eta}_d = 1$ for some values of the covariates (in particular s_{-1}). In fact we can even have $\underline{\eta}_d = \bar{\eta}_d = 0$ or $\underline{\eta}_d = \bar{\eta}_d = 1$, in which case no one moves given these state variables. If $d \neq d_{-1}$, $s_d^*(\eta, a, s_{-1}, d_{-1} \neq d) = \tilde{s}_d^*(\eta, a, s_{-1}, d_{-1} \neq d)$ because if they swap tenure these households are already moving by construction. So the boundaries only matter when $d = d_{-1}$.

Another property is that the mass point is shifting with the value of s_{-1}. In other words, if $s_{-1}^a < s_{-1}^b$, then $\underline{\eta}_d(a, s_{-1}^a, d_{-1}) \leq \underline{\eta}_d(a, s_{-1}^b, d_{-1})$ and $\bar{\eta}_d(a, s_{-1}^a, d_{-1}) \leq \bar{\eta}_d(a, s_{-1}^b, d_{-1})$. This ordering also translate in the s_d.

Moreover, households who endure the exogenous move shock $m = 1$ are necessarily moving and their optimal housing service choice $s_d^*(\eta, \cdot, m = 1) = \tilde{s}_d^*(\eta, \cdot)$ for all η. So if we observe that an household does not move in the data, it means that $m = 0$. Households with $m = 0$ will never pick an optimal s_d different from s_{-1} between $[s_d^*(\underline{\eta}_d, \cdot), s_d^*(\bar{\eta}_d, \cdot)]$ if they keep the same tenure $d = d_{-1}$.

Which means that households who move without changing tenure with an optimal s_d^* choice in $[s_d^*(\underline{\eta}_d, \cdot), s_d^*(\bar{\eta}_d, \cdot)]$ endured the exogenous move shock $m = 1$. In other words, moves to houses similar to the previous one (close to s_{-1}) are due to the exogenous move shock in the model. Because of the fixed cost, households would never pick such a choice if they were not forced to move.

Identification with fixed costs is more complicated as the optimal s_d^* are not strictly monotone with respect to η anymore, and there are more objects to identify (the bounds). We proceed piece by piece for the identification. First, we show how to identify the boundaries in terms of s_d, i.e. how to identify $s_d^*(\underline{\eta}_d, s_{-1}, a, d_{-1})$ and $s_d^*(\bar{\eta}_d, s_{-1}, a, d_{-1})$ for all (s_{-1}, a, d_{-1}). Second, we show how to identify the probability of receiving and exogenous moving shock, p^m. Third, we show how to identify s_d^* outside boundaries $\underline{s}_d, \bar{s}_d$. Fourth, we show how to identify $s_d^*(\eta, \cdot, m = 1)$ between boundaries. Finally, we show how to recover $Pr(D = 1|\eta, \cdot, m = 1)$, as well as the consumption choice of movers and the measurement error distribution.

<u>Step 1:</u> *identification of $s_d^*(\underline{\eta}_d)$ and $s_d^*(\bar{\eta}_d)$*

We drop the dependence of $\underline{\eta}_d, \bar{\eta}_d, s_d(\underline{\eta}_d), s_d(\bar{\eta}_d)$ on (a, s_{-1}, d_{-1}) in the notation here to simplify the exposition. As already mentioned, $\underline{\eta}_d$ and $\bar{\eta}_d$ only matters when $d_{-1} = d$.

First, notice that if there was no moving shock, i.e. if $m = 0$ for all household, then the boundaries are straightforward to identify. Indeed, they correspond to the highest value of s_d before $s_d^* = s_{-1}$ and the lowest value of s_d such that $s_d^* = s_{-1}$ (conditional on the covariates).

The presence of exogenous moving shocks prevents us from identifying the boundaries as easily because household with $m = 1$ will move freely and could pick s_d close to s_{-1} without changing tenure. The idea for the identification of the boundaries rely on the fact that, above $\underline{s}_d$ and below $\bar{s}_d$, we should observe jumps (proportional to p^m) in the conditional density of s_d^* because only the population with $m = 1$ will remain, while the households with $m = 0$ will all choose to stay at $s_d^* = s_{-1}$.

Formally, since m is an exogenous shock, independent from everything else, for all s we have

$$F_{S_d|D=d,D_{-1}=d,s_{-1},a}(s_d) = p^m \ F_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=1}}(s_d) + (1 - p^m) \ F_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=0}}(s_d),$$

and for all $s \neq s_{-1}$ (because at s_{-1} the density has a mass point):

$$\frac{dF_{S_d|D=d,D_{-1}=d,s_{-1},a}(s_d)}{ds_d} = p^m \, \frac{dF_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=1}}(s_d)}{ds_d} + (1-p^m) \, \frac{dF_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=0}}(s_d)}{ds_d}$$

$$\iff f_{S_d|D=d,D_{-1}=d,s_{-1},a}(s_d) = p^m \, f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=1}}(s_d) + (1-p^m) \, f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=0}}(s_d),$$

where we define the density $f_{S_d|D=d,D_{-1}=d,s_{-1},a}(s_d) = dF_{S_d|D=d,D_{-1}=d,s_{-1},a}(s_d)/ds_d$, and where $p^m = Pr(m=1)$.

Now notice that, for all $s \notin [s_d^*(\underline{\eta}_d), s_d^*(\bar{\eta}_d)]$:

$$F_{S_d|D=d,D_{-1}=d,s_{-1},a}(s_d) = \quad F_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=1}}(s_d) = F_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=0}}(s_d)$$

and

$$f_{S_d|D=d,D_{-1}=d,s_{-1},a}(s_d) = \quad f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=1}}(s_d) = f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=0}}(s_d).$$

While, for all $s \in [s_d^*(\underline{\eta}_d), s_d^*(\bar{\eta}_d)] \setminus \{s_{-1}\}$:

$$f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=0}}(s_d) = 0.$$

Then, for all $s \in [s_d^*(\underline{\eta}_d), s_d^*(\bar{\eta}_d)] \setminus \{s_{-1}\}$:

$$f_{S_d|D=d,D_{-1}=d,s_{-1},a}(s_d) = p^m \, f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=1}}(s_d)$$

Thus, we can *identify* $s_d^*(\underline{\eta}_d)$ and $s_d^*(\bar{\eta}_d)$ by observing *discontinuities in the density*. Indeed, in a close neighborhood to $s_d^*(\underline{\eta}_d)$,

$$f_{S_d|D=d,D_{-1}=d,s_{-1},a}(s_d)\Big|_{s_d=\tilde{s}_d^*(\underline{\eta}_d)} = \begin{cases} p^m \, f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=1}}(s_d) & \\ \quad + \quad (1-p^m) \, f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=0}}(s_d) & \text{if } s_d < s_d^*(\underline{\eta}_d) \\ p^m \, f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=1}}(s_d) & \text{if } s_d \geq s_d^*(\underline{\eta}_d) \end{cases}$$

Thus, if $\underline{\eta}_d > 0$, we identify $\tilde{s}_d^*(\underline{\eta}_d)$ as the only $s_d < s_{-1}$ such that there is a jump in the density.And if there is no such point, it means that $\underline{\eta}_d = 0$, meaning that the fixed cost covers all the optimal choices $\tilde{s}_d$ below s_{-1}.

Similarly, in a close neighborhood to $\tilde{s}^*_d(\bar{\eta}_d)$,

$$f_{S_d|D=d,D_{-1}=d,s_{-1},a}(s_d)\Big|_{s_d=\tilde{s}^*_d(\bar{\eta}_d)} = \begin{cases} p^m \; f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=1}}(s_d) & \text{if } s_d < s^*_d(\bar{\eta}_d) \\ p^m \; f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=1}}(s_d) & \\ \quad + \quad (1-p^m) \; f_{S_d|D=d,D_{-1}=d,s_{-1},a,\mathbf{m=0}}(s_d) & \text{if } s_d \geq s^*_d(\bar{\eta}_d) \end{cases}$$

Thus, if $\bar{\eta}_d < 1$, we identify $\tilde{s}^*_d(\bar{\eta}_d)$ as the only $s_d > s_{-1}$ such that there is a jump in the density. And if we observe no such point, it means that $\bar{\eta}_d = 1$.

Notice that the knowledge of p^m is not required for the identification of the boundaries: one just need to observe a jump in the density, which will occur as long as $p^m \in (0,1)$. In the extreme case where $p^m = 1$, everyone moves and the result of the previous section holds. In the other extreme where $p^m = 0$, there is no exogenous move and so we identify the boundaries as stated previously: as the lowest value s_d below s_{-1} and the highest value greater than s_{-1}.

Step 2: identification of p^m

To identify p^m now, we will exploit the knowledge of the probability of not moving $s_d = s_{-1}$.

Notice that, by construction, since $m \perp \eta, \epsilon$

$$\begin{aligned} & Pr\Big(S_d \in [\tilde{s}^*_d(\underline{\eta}_d), \tilde{s}^*_d(\bar{\eta}_d)]\Big| D=d, D_{-1}=d, s_{-1}, a\Big) \\ = \; & Pr\Big(S_d \in [\tilde{s}^*_d(\underline{\eta}_d), \tilde{s}^*_d(\bar{\eta}_d)]\Big| D=d, D_{-1}=d, s_{-1}, a, m=0\Big) \\ \Bigg(\textit{and also} \; = \; & Pr\Big(S_d \in [\tilde{s}^*_d(\underline{\eta}_d), \tilde{s}^*_d(\bar{\eta}_d)]\Big| D=d, D_{-1}=d, s_{-1}, a, m=1\Big)\Bigg) \end{aligned}$$

Moreover, we know that:

$$Pr\Big(S_d \in [\tilde{s}^*_d(\underline{\eta}_d), \tilde{s}^*_d(\bar{\eta}_d)]\Big| D=d, D_{-1}=d, s_{-1}, a, m=0\Big) = Pr\Big(S_d = s_{-1}\Big| D=d, D_{-1}=d, s_{-1}, a, m=0\Big)$$

Now, $Pr(S_d = s_{-1}\Big| D=d, D_{-1}=d, s_{-1}, a, m=0)$ is not directly estimable as we do not observe m.

But we observe $Pr(S_d = s_{-1} \Big| D = d, D_{-1} = d, s_{-1}, a)$. And we know that:

$$\begin{aligned} Pr\Big(S_d = s_{-1}\Big|D = d, D_{-1} = d, s_{-1}, a\Big) = &\quad (1 - p^m)\ Pr\Big(S_d = s_{-1}\Big|D = d, D_{-1} = d, s_{-1}, a, m = 0\Big) \\ &+ p^m\ \underbrace{Pr\Big(S_d = s_{-1}\Big|D = d, D_{-1} = d, s_{-1}, a, m = 1\Big)}_{=0 \text{ by continuity of } \tilde{s}_d^*(\cdot)} \\ = &\quad (1 - p^m)\ Pr\Big(S_d = s_{-1}\Big|D = d, D_{-1} = d, s_{-1}, a, m = 0\Big). \end{aligned}$$

We obtain this because, when $m = 1$, the household adopts the optimal choice $\tilde{s}_d^*(\cdot)$. Thus $Pr(S_d = s_{-1}\Big|D = d, D_{-1} = d, s_{-1}, a, m = 1) = 0$ because $\tilde{s}_d^*(\cdot)$ is continuous, so the likelihood of choosing $\tilde{s}_d^*$ exactly equal to s_{-1} is zero.

By mixing the three previous properties, we obtain

$$\begin{aligned} &Pr\Big(S_d \in [\tilde{s}_d^*(\underline{\eta}_d), \tilde{s}_d^*(\bar{\eta}_d)]\Big|D = d, D_{-1} = d, s_{-1}, a\Big) \\ &\qquad = Pr\Big(S_d \in [\tilde{s}_d^*(\underline{\eta}_d), \tilde{s}_d^*(\bar{\eta}_d)]\Big|D = d, D_{-1} = d, s_{-1}, a, m = 0\Big) \\ &\qquad = Pr\Big(S_d = s_{-1}\Big|D = d, D_{-1} = d, s_{-1}, a, m = 0\Big) \\ &\qquad = \frac{Pr\Big(S_d = s_{-1}\Big|D = d, D_{-1} = d, s_{-1}, a\Big)}{1 - p^m}, \end{aligned}$$

where only p^m is unknown since we already identified $\tilde{s}_d^*(\underline{\eta}_d), \tilde{s}_d^*(\bar{\eta}_d)$. It gives that

$$p^m = \frac{Pr\Big(S_d \in [\tilde{s}_d^*(\underline{\eta}_d), \tilde{s}_d^*(\bar{\eta}_d)]\Big|D = d, D_{-1} = d, s_{-1}, a\Big) - Pr\Big(S_d = s_{-1}\Big|D = d, D_{-1} = d, s_{-1}, a\Big)}{Pr\Big(S_d \in [\tilde{s}_d^*(\underline{\eta}_d), \tilde{s}_d^*(\bar{\eta}_d)]\Big|D = d, D_{-1} = d, s_{-1}, a\Big)},$$

which *identifies* p^m.[11]

<u>*Step 3:*</u> *identification of $s_d^*(\eta, \cdot)$ outside the boundary points $\underline{\eta}_d, \bar{\eta}_d$*

As displayed in Figure 3.4.1, outside of the boundaries $(\underline{\eta}_d(a, s_{-1}, d_{-1}), \bar{\eta}_d(a, s_{-1}, d_{-1}))$, households move and $s_d^*(\eta, a, s_{-1}, d_{-1}) = \tilde{s}_d^*(\eta, a, s_{-1}, d_{-1})$ for all (a, s_{-1}, d_{-1}). In other words, outside of the boundaries, $s_d^*(\eta, a, s_{-1}, d_{-1}, m = 0) = s_d^*(\eta, a, s_{-1}, d_{-1}, m = 1) = \tilde{s}_d^*(\eta, a, s_{-1}, d_{-1})$, and the presence

[11] Notice that, since we are studying identification at given exogenous covariates, it means that we could allow for a p^m that depend on these covariates z.

of the moving shock is irrelevant: households endogenously choose to move and make the same choice as if they were facing a moving shock.

We build upon Section 3.4.1, using property (3.1) for the identification of the optimal s^*_d outside the boundaries, i.e. conditional on moving. i.e. we use:

$$\tilde{s}^*_d(\eta, d_{-1} = 1, s_{-1}, a) = \tilde{s}^*_d(\eta, d_{-1} = 0, s'_{-1} = \cdot, a' = a + q\frac{s_{-1}}{1+r}),$$

where s'_{-1} is the previous housing service of the previously renting household ($d_{-1} = 0$), which does not matter on their choice conditional on moving since ex renters have no housing wealth.
The bounds introduce some specificities though. Depending on what a, s_{-1} and s'_{-1} are, and using $a' = a + qs_{-1}/(1+r)$, we will have four boundaries. For $d = 1$, we will have household staying when $d_{-1} = 1$ and the boundaries will depend on a and s_{-1}, i.e. $\underline{\eta}_1(a, s_{-1}, d_{-1} = 1)$, $\bar{\eta}_1(a, s_{-1}, d_{-1} = 1)$. Similarly, previous renters ($d_{-1} = 0$) only stays in their home if they remain renters ($d = 0$) and the boundaries in this case will depend on a' and s'_{-1}, i.e. $\underline{\eta}_0(a', s'_{-1}, d_{-1} = 0)$, $\bar{\eta}_0(a', s'_{-1}, d_{-1} = 0)$. So, in fact for any given (a, s_{-1}, s'_{-1}) and $a' = a + qs_{-1}/(1+r)$, we will have identification of the optimal choices s^*_d for

$$\eta \notin [\underline{\eta}_0(a', s'_{-1}, d_{-1} = 0), \bar{\eta}_0(a', s'_{-1}, d_{-1} = 0)] \cup [\underline{\eta}_1(a, s_{-1}, d_{-1} = 1), \bar{\eta}_1(a, s_{-1}, d_{-1} = 1)].$$

Now, notice that property (3.1) holds for any s'_{-1}. So, provided that we have *enough variation in the observed support of* s_{-1}, $[\underline{\eta}_0, \bar{\eta}_0]$ are not important. One can identify s^*_1 and s^*_0 for all $\eta \notin [\underline{\eta}_1(a, s_{-1}, d_{-1} = 1), \bar{\eta}_1(a, s_{-1}, d_{-1} = 1)]$. Only the boundaries for a, s_{-1} when $d = 1$ matters. Indeed, we will proceed to the identification by taking several value of s'_{-1} in order to move the $\underline{\eta}_0, \bar{\eta}_0$ such that they 'do not matter'. Recall that $\underline{\eta}_0, \bar{\eta}_0$ increase with s'_{-1}. So typically, to identify $s^*_1(\eta, a, s_{-1}, d_{-1} = 1)$ and $s^*_0(\eta, a, s_{-1}, d_{-1} = 1)$ on $[0, \underline{\eta}_1(a, s_{-1}, d_{-1} = 1)]$, pick a sufficiently high s'_{-1} such that $\underline{\eta}_0(a', s'_{-1}, d_{-1} = 0) > \underline{\eta}_1(a, s_{-1}, d_{-1} = 1)$. If we observe enough variation in s'_{-1}, such an s'_{-1} should always exists.[12] Similarly, to identify $s^*_1(\eta, a, s_{-1}, d_{-1} = 1)$ and $s^*_0(\eta, a, s_{-1}, d_{-1} = 1)$ on $[\underline{\eta}_1(a, s_{-1}, d_{-1} = 1), 1]$, pick a sufficiently low s'_{-1} such that $\bar{\eta}_0(a', s'_{-1}, d_{-1} = 0) > \bar{\eta}_1(a, s_{-1}, d_{-1} = 1)$.

Therefore, for all (a, s_{-1}), $s^*_0(\eta, a, s_{-1}, d_{-1} = 1)$ and $s^*_1(\eta, a, s_{-1}, d_{-1} = 1)$ are identified for

[12] Notice that in the first step we only identify the bounds in terms of s and not the value of $\underline{\eta}, \bar{\eta}$ directly. So this is a guessing game: take s'_{-1}, it yields $\underline{s}_0 = s^*_0(\underline{\eta}_0(a', s'_{-1}, d_{-1} = 0), d_{-1} = 0, a', s'_{-1})$. Now run the identification procedure in step 1. If we notice that $s^*_1(\underline{s}_0) < \underline{s}_1 \equiv s^*_1(\underline{\eta}_1(a, s_{-1}, d_{-1} = 1), d_{-1} = 1, a, s_{-1})$, it means that s'_{-1} was not high enough (because $\underline{\eta}_0 < \underline{\eta}_1$), and one need to run it again with a higher s'_{-1}.

all $\eta \notin [\underline{\eta}_1(a, s_{-1}, d_{-1} = 1), \bar{\eta}_1(a, s_{-1}, d_{-1} = 1)]$.[13] In other words, we fully identify functions $\tilde{s}^*_d(\eta, a, s_{-1}, d_{-1} = 1)$ and $s^*_d(\eta, a, s_{-1}, d_{-1} = 1, m = 1)$ on this subset, i.e. conditional on endogenously moving.

Step 4: identification of $\tilde{s}^*_d(\eta, \cdot)$ for $\eta \in [\underline{\eta}_1, \bar{\eta}_1]$

It remains to identify $s^*_d(\eta, a, s_{-1}, d_{-1} = 1, \mathbf{m} = \mathbf{1})$ for both d, on the subset of $\eta \in [\underline{\eta}_1, \bar{\eta}_1]$. Notice that for $m = 0$, we do not need to identify $s^*_1(\eta, a, s_{-1}, d_{-1} = 1, \mathbf{m} = \mathbf{0})$ because by definition it is equal to s_{-1} on this subsample. So what remains to be identified are the policies for movers $\tilde{s}_d$, or equivalently the policies conditional on $m = 1$.

The idea for the identification is close to what we did in the previous step: we will once again shift s_{-1} in order to shift the mass point. Notice that

$$\begin{aligned} \tilde{s}_d(\eta, d_{-1} = 1, s_{-1}, a) = & \tilde{s}_d(\eta, d_{-1} = 1, s'_{-1}, a') \\ & \forall\ s'_{-1}, a' \text{ s.t. } \frac{qs'_{-1}}{1+r} + a' = \frac{qs_{-1}}{1+r} + a. \end{aligned} \tag{3.6}$$

In words, property (3.6) is similar to the idea in property 3.1. We have that conditional on moving, only the total wealth matters for households. So, even two ex-owners ($d_{-1} = 1$) with the same total wealth will make the same housing service choice when they move. How their wealth is composed of financial or housing assets does not matter once they move.

Therefore, conditional on having enough variation in s_{-1}, we can vary s'_{-1} at fixed total wealth ($= qs_{-1}/(1 + r) + a$) in order to identify $s_d(\eta, d_{-1} = 1, s_{-1}, a)$ for all η. For example, one can proceed to the identification procedure described in the previous step with two different values of s_{-1}: s_{-1} and s'_{-1}. With s'_{-1}, a' such that

$$s^*_1(\underline{\eta}_1(a', s'_{-1}, d_{-1} = 1), d_{-1} = 1, a', s'_{-1}) > s^*_1(\bar{\eta}_1(a, s_{-1}, d_{-1} = 1), d_{-1} = 1, a, s_{-1}).$$

In which case the optimal choice for the movers $\tilde{s}^*_1$ and $\tilde{s}^*_0$ are identified for all $\eta \in [0, 1]$.

Step 5: identification of η for movers and of $Pr(D = 1|\eta, d_{-1}, s_{-1}, a, m = 1)$ for all $\eta \in [0, 1]$

We proceed exactly as in steps 3 and 4 by shifting s_{-1}. While we identify $s^*_d(\cdot, m = 1)$, we can directly

[13] And the policies when $d_{-1} = 0$ are identified using the ones with $d_{-1} = 1$ through property (3.1), as in Section 3.4.1.

invert them, to recover $\eta = (\tilde{s}_d^*)^{-1}(s_d^{obs}, s_{-1}, a, d_{-1})$. In other words, *for movers, η is identified*, as if it was an observed variable. Then we use it to compute the conditional choice probabilities (CCPs) of movers:

$$Pr(D = 1|\eta, d_{-1}, s_{-1}, a, m = 1) \text{ for all } \eta \in [0, 1]$$

Remark, on the subsample of moving households, $Pr(D = 1|\eta, d_{-1}, s_{-1}, a, m = 1) = Pr(D = 1|\eta, d_{-1}, s_{-1}, a, m = 0) = Pr(D = 1|\eta, d_{-1}, s_{-1}, a)$. So the only case where we do not directly identify $Pr(D = 1|\eta, d_{-1}, s_{-1}, a, m = 0)$ is when $s_d^* = s_{-1}$. Because in this case, we cannot invert s_d^* to recover the corresponding value of the unobserved shock η.

<u>Step 6:</u> *identification of $c_d^*(\eta, d_{-1}, a, m = 1)$ for all $\eta \in [0, 1]$ and the measurement error ζ*
Similarly, we invert $s_d^*(\cdot, m = 1)$ to recover η from observed s_d for households who move exogenously. Then we use this η as if it was observed to run the non parametric regression

$$c_d^{obs} = c_d^*(\eta, s_{-1}, a, d_{-1}, m = 1) + \zeta$$

This regression directly identifies the optimal consumption policies of movers and the *distribution of measurement errors ζ, denoted F_ζ*. As for $Pr(D = 1|\eta, \cdot)$, for movers, $c_d^*(\eta, s_{-1}, a, d_{-1}, m = 1) = c_d^*(\eta, s_{-1}, a, d_{-1}, m = 0)$. So we also identify $c_d^*(\cdot, m = 0)$ for movers. It remains to identify the optimal consumption choice of stayers.

3.4.3 Identification of c for stayers

Contrary to s_d, when $\eta \in [\underline{\eta}_d(a, s_{-1}, d_{-1} = d), \bar{\eta}_d(a, s_{-1}, d_{-1} = d)]$, c_d varies and is not fixed (as s_d is equal to s_{-1}). So it remains to identify $c_d^*(\eta, s_{-1}, a, d_{-1} = d, m = 0)$ for $\eta \in [\underline{\eta}_d(a, s_{-1}, d_{-1} = d), \bar{\eta}_d(a, s_{-1}, d_{-1} = d)]$. It is interesting that consumption provides some information on η (though the information is noisy because of measurement errors).

In the data, we only observe $c_d^{obs} = c_d^* + \zeta$ and the distribution of c_d^{obs}. However, ζ is an independent measurement error and we have identified its distribution F_ζ on the subset of movers in step 6 of the previous section. So, by *deconvolution* (e.g. Comte and Lacour, 2011), we can recover the distribution of the true $c_d^*(\eta, \cdot)$, denoted $F_{C_d|D=d,\cdot}(c_d)$, from any conditional distribution of observed c_d^{obs}.

Let us identify $c_d^*(\eta, s_{-1}, a, d_{-1} = d, m = 0)$ for all η. Recall that $c_d^*(\eta, s_{-1}, a, d_{-1} = d, m = 0)$ is already identify outside of the boundaries. So we only need to identify it for stayers, i.e. for $\eta \in [\underline{\eta}_d(a, s_{-1}, d_{-1} = d), \bar{\eta}_d(a, s_{-1}, d_{-1} = d)]$. Proceed by values of d. Start with $d = 1$. In the data, if households choose $s_d = s_{-1}$ and $d = d_{-1}$, it means that $m = 0$: they did not move so it means that they did not endure an exogenous shock forcing them to move. It is as if $m = 0$ was observed in this case. Using these observations, we recover $F_{C_d|D=d,\cdot,m=0}(c_d)$ (using deconvolution).
Do the following computation at any given (a, s_{-1}), from which we abstract in the following notation. For any $h \in [\underline{\eta}_d(a, s_{-1}, d_{-1} = d), \bar{\eta}_d(a, s_{-1}, d_{-1} = d)]$, we have

$$\begin{aligned}
Pr(\eta \leq h) &= Pr(\eta \leq h | D_{-1} = 1) \\
&= Pr(\eta \leq h | D = 1, D_{-1} = 1) Pr(D = 1 | D_{-1} = 1) \\
&\quad + Pr(\eta \leq h | D = 0, D_{-1} = 1) Pr(D = 0 | D_{-1} = 1) \\
&= Pr(\eta \leq h | D = 1, D_{-1} = 1, m = 0) Pr(D = 1 | D_{-1} = 1, m = 0)(1 - p^m) \\
&\quad + Pr(\eta \leq h | D = 1, D_{-1} = 1, m = 1) Pr(D = 1 | D_{-1} = 1, m = 1) p^m \\
&\quad + Pr(\eta \leq h | D = 0, D_{-1} = 1) Pr(D = 0 | D_{-1} = 1) \\
&= Pr(C_1^* \geq c_1^*(h, d_{-1} = 1, m = 0) | D_{-1} = 1, D = 1, m = 0) Pr(D = 1 | D_{-1} = 1, m = 0)(1 - p^m) \\
&\quad + Pr(\eta \leq h | D = 1, D_{-1} = 1, m = 1) Pr(D = 1 | D_{-1} = 1, m = 1) p^m \\
&\quad + Pr(\eta \leq h | D = 0, D_{-1} = 1) Pr(D = 0 | D_{-1} = 1). \qquad (3.7)
\end{aligned}$$

The first equality holds because $\eta \perp d_{-1}$. The second and third equalities come from the law of total probability, where $p^m = Pr(m = 1)$. We obtain the fourth equality because $c_1^*(\eta, \cdot)$ is a strictly monotone (decreasing) function of η. The only thing that is unknown in this equation (3.7) is the optimal policy. Now, when $d \neq d_{-1}$, the households move and we identified η by inverting the optimal s_d^* choice. So, $Pr(\eta \leq h | D = 0, D_{-1} = 1)$ is known. $Pr(D = 0 | D_{-1} = 1)$ can be directly estimated from the data too. Similarly, $Pr(\eta \leq h | D = 1, D_{-1} = 1, m = 1)$ was also already identified, because when $m = 1$ we observe moves in s_d, we can reverse it to identify η and its distribution. $Pr(D = 1 | D_{-1} = 1, m = 1)$ is also known as we identified $Pr(D = 1 | \eta, D_{-1} = 1, m = 1)$ for all η when $m = 1$. Finally, we can show that $Pr(D = 1 | \eta, D_{-1} = 1, m = 0)$ is also known.

Indeed, by the law of total probability,

$$Pr(D=1|D_{-1}=1) = Pr(D=1|D_{-1}=1,m=1)\ p^m + (1-p^m)\ Pr(D=1|D_{-1}=1,m=0).$$

Where p^m is identified, $Pr(D=1|D_{-1}=1)$ is observed from the data and $Pr(D=1|D_{-1}=1,m=1)$ has also been identified. So,

$$Pr(D=1|D_{-1}=1,m=0) = \frac{Pr(D=1|D_{-1}=1) - p^m\ Pr(D=1|D_{-1}=1,m=1)}{1-p^m}$$

and $Pr(D=1|D_{-1}=1,m=0)$ is known.

So, since the distribution $F_{C_1|D=1,D_{-1}=1,m=0}(c_d)$ has been identified by deconvolution, the only unknown in equation (3.7) is the optimal consumption choice $c_1^*(\eta, d_{-1}=1, m=0)$. Thus, it is identified by:

$$\begin{aligned}
1-&F_{C_1|D=1,D_{-1}=1,m=0}(c_1^*(h,d_{-1}=1,m=0))\\
&= \Big(Pr(\eta \leq h)\\
&\quad - Pr(\eta \leq h|D=1,D_{-1}=1,m=1)Pr(D=1|D_{-1}=1,m=1)p^m\\
&\quad - Pr(\eta \leq h|D=0,D_{-1}=1)Pr(D=0|D_{-1}=1)\Big)\ \frac{1}{Pr(D=1|D_{-1}=1,m=0)(1-p^m)}
\end{aligned}$$

Which means that:

$$\begin{aligned}
&c_1^*(h,d_{-1}=1,m=0)\\
&\quad = F^{-1}_{C_1|D=1,D_{-1}=1,m=0}\Bigg(1-\Big(Pr(\eta \leq h)\\
&\qquad - Pr(\eta \leq h|D=1,D_{-1}=1,m=1)Pr(D=1|D_{-1}=1,m=1)p^m\\
&\qquad - Pr(\eta \leq h|D=0,D_{-1}=1)Pr(D=0|D_{-1}=1)\Big)\ \frac{1}{Pr(D=1|D_{-1}=1,m=0)(1-p^m)}\Bigg),
\end{aligned}$$

and $c_1^*(h,d_{-1},m=0)$ is identified for all η in the boundaries. Since it was already identified outside the boundaries, it is identified for all $\eta \in [0,1]$.

We apply exactly the same reasoning for $d = 0$ to identify $c_0^*(\eta, d_{-1} = 0, m = 0)$ for all η in the boundaries $[\underline{\eta}_0(a, s_{-1}, d_{-1} = 0), \bar{\eta}_0(a, s_{-1}, d_{-1} = 0)]$, and thus for all $\eta \in [0, 1]$.

Information on $Pr(D = 1|\eta, d_{-1}, s_{-1}, a, m = 0)$ ***for*** η ***in*** $[\underline{\eta}_d, \bar{\eta}_d]$:
We can use the identified $c_d^*(\eta, s_{-1}, a, d_{-1} = d, m = 0)$ to identify $Pr(D = 1|\eta, d_{-1}, s_{-1}, a, m = 0)$ for $\eta \in [\underline{\eta}_d(a, s_{-1}, d_{-1} = d), \bar{\eta}_d(a, s_{-1}, d_{-1} = d)]$. Indeed, inside this set, $s_d^* = s_{-1}$. Thus we cannot invert the observation of s_d to recover the value of η. However, c_d^* is still varying, and we identify it. As a consequence, the *observed consumption contains information on* η. Because of the measurement errors we cannot directly invert the observed consumption to recover η as we are proceeding for the housing services. Yet c_d^{obs} still contains some information about η: the higher it is, the more likely it is that η is low for example (since c_d is decreasing with respect to η). More precisely, given that we observe $c_d, d, a, d_{-1}, s_{-1}$, we have

$$Pr(\eta = h|c_d^{obs}, d, s_{-1}, d_{-1} = d, a) = Pr\Big(\zeta = c_d^{obs} - c_d^*(h, s_{-1}, a, d_{-1} = d, m = 0)\Big),$$

where we drop the conditioning in the right hand side since ζ is independent from the rest.
Therefore, we can again compute the CCPs, $Pr(D = 1|\eta, d_{-1}, s_{-1}, a, m = 0)$ from the data but using the probabilistic likelihood of each η (instead of the knowledge of η which is identified for movers) for the observations where $s_d = s_{-1}$.

Therefore, we identify the optimal *Conditional Continuous Choices* (CCCs) $s_d^*(\eta, s_{-1}, d_{-1}, a, z, m)$ and $c_d^*(\eta, s_{-1}, d_{-1}, a, z, m)$ and the optimal *Conditional Choice Probabilities* (CCPs) $Pr(d|\eta, s_{-1}, d_{-1}, a, z, m)$, for all η, a, s_{-1}, d_{-1}. We also identify the measurement errors distribution F_ζ, and the probability to undergo an exogenous moving shock, p^m.

3.4.4 Identification of the structural parameters

We have proved the identification of the optimal choices in every period. Now we can use them and apply Bruneel-Zupanc (2021) directly to identify the structural parameters of our dynamic model.

3.5 Empirical Strategy

We build a two step estimation method of the parametric model. In a first stage, we estimate *parametric* optimal policies via maximum likelihood. More non-parametric alternative estimation methods can also be used, see Bruneel-Zupanc (2021). However here, we do not have many observations of moving individuals, so we prefer to use a more parametric estimation procedure for the optimal choices. In the second stage, we estimate the structural parameters of the model via forward simulation methods using the optimal policies estimated in the first stage, in the spirit of Hotz et al. (1994).

3.5.1 Optimal policies

We specify parametric functional forms for the optimal policies. Then we estimate the parameters by maximum likelihood, since the likelihood is known given the parametric functional form. We will specify functional forms for $s_d^*(\cdot), c_d^*(\cdot)$ and $Pr(D=1|\cdot)$. As well as functional forms for the boundaries $\underline{\eta}_d, \bar{\eta}_d$, the measurement errors and the probability of exogenous move shock p^m.

Housing service s:

First, we specify a functional form for the *housing service conditional on moving* $\tilde{s}_d^*$:

$$\begin{aligned} log(\tilde{s}^*) = \delta &+ \delta_1 d + \alpha_0(1-d)\psi(\eta) + \alpha_1 d\psi(\eta) \\ &+ \gamma_0(1-d)\Big[d_{-1}\big(\frac{q}{1+r}s_{-1}+a\big) + (1-d_{-1})a + \frac{\text{income}}{1+r}\Big] \\ &+ \gamma_1 d\Big[d_{-1}\big(\frac{q}{1+r}s_{-1}+a\big) + (1-d_{-1})a + \frac{\text{income}}{1+r}\Big] \\ &+ \lambda' x, \end{aligned}$$

where $d_{-1}(\frac{q}{1+r}s_{-1}+a)+(1-d_{-1})a+\frac{\text{income}}{1+r}$ represents the *total wealth.* $\psi(\eta)$ is a monotone transformation of η applied to account for nonlinear effect of η, e.g. $\psi(\eta)=F^{-1}(\eta)$ where F is a standard normal distribution. By *monotonicity*, we have the constraint on the parameters $\alpha_0>0$ and $\alpha_1>0$.

Now, recall that with fixed costs, the true optimal choice is:

$$s^* = \begin{cases} s_{-1} & \text{if } d = d_{-1} \text{ and } (\underline{\eta}_d \leq \eta \leq \bar{\eta}_d) \iff (\underline{s}_d \leq \tilde{s}^* \leq \bar{s}_d) \\ \tilde{s}^* & \text{otherwise.} \end{cases}$$

Thus, we need a parametric specification for $(\underline{s}_d, \bar{s}_d)$, which depend on s_{-1}. For now, we use a simple specification with two parameters for each d, $\underline{\kappa}_d$ and $\bar{\kappa}_d$:

$$\underline{s}_d(s_{-1}) = s_{-1} - \underline{\kappa}_d \quad \text{and } \bar{s}_d(s_{-1}) = s_{-1} + \bar{\kappa}_d.$$

We can include additional effect from the other variables (asset, income, demographics) on the boundaries but we keep it simple for now.

Housing tenure d:

For the housing tenure choice probability d, we use a logistic specification:

$$Pr(D = 1|\eta, d_{-1}, s_{-1}, x) = \frac{1}{1 + exp(-(\theta_0 d_{-1} + \theta_1 \psi(\eta) + \theta_2 \text{total wealth}))}.$$

Consumption c:

We use a specification similar to the specification of housing services for non-housing consumption. Except that we include s into it, such that for non movers, we will have a fixed s_{-1} but still have variation through η. And for the movers the variation will be captured through s, which is a strictly increasing function mapped to η anyway.

$$\begin{aligned} log(c^*) = \delta^c + \delta_1^c d + \alpha_0^c (1-d) s + \alpha_1^c d\ s \\ + \mathbb{1}\{s = s_{-1}\} \Big(\alpha_2^c (1-d) \psi(\eta) + \alpha_3^c d \psi(\eta) \Big) \\ + \gamma_0^c (1-d)(\text{total wealth}) + \gamma_1^c d (\text{total wealth}) + \lambda^{c\prime} x. \end{aligned}$$

And consumption is observed with measurement errors

$$c = c^* + \epsilon^c \quad \text{with} \;\; \epsilon^c \sim \mathcal{N}(0, \sigma^c),$$

where σ^c is another parameter to estimate.

Estimation via Maximum Likelihood:
Given the specification, including also a parameter p^m for the exogenous moving shock probability, we can compute the likelihood of each observation (s, d, c) in the data. Even when $s = s_{-1}$ we know the likelihood of observing it happening. So we can estimate all the parameters of the optimal choices laid down above via maximum likelihood.

3.5.2 Structural parameters

Once the parametric optimal choices are estimated, we use them in *forward simulations* of the model in order to estimate the structural parameters of the dynamic model, in the spirit of Hotz et al. (1994). In order to run these forward simulation, we also estimate the transition of the exogenous state variables (income and demographics) parametrically directly from the data beforehand.

As shown in Bruneel-Zupanc (2021), this two step estimation method has the advantage of being fast, as it avoids having to solve for the optimal choices in the model for each set of parameters.

3.6 Conclusion

This paper builds a dynamic model of household consumption and housing decisions. We provide identification conditions of this dynamic model and an estimation method built upon the identification proof. The planned estimation of the model will allow to estimate key parameters for this sample of French households, such as the substitution between housing and non-housing consumption and fixed housing switching costs. Once the model is estimated, we will be able to run many relevant policy counterfactuals, by modifying transaction costs for example.

Bibliography

Abbring, J. H. and y. Daljord (2020): "Identifying the discount factor in dynamic discrete choice models," *Quantitative Economics*, 11, 471–501.

Aguirregabiria, V. and P. Mira (2002): "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 70, 1519–1543.

——— (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1–53.

——— (2010): "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 156, 38 – 67, structural Models of Optimization Behavior in Labor, Aging, and Health.

Ahn, T., P. Arcidiacono, A. Hopson, and J. R. Thomas (2019): "Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors," Working Paper 26556, National Bureau of Economic Research.

Alan, S., O. Attanasio, and M. Browning (2009): "Estimating Euler equations with noisy data: two exact GMM estimators," *Journal of Applied Econometrics*, 24, 309–324.

Altuğ, S. and R. A. Miller (1998): "The Effect of Work Experience on Female Wages and Labour Supply," *The Review of Economic Studies*, 65, 45–85.

Anenberg, E. (2016): "Information frictions and housing market dynamics," *International Economic Review*, 57.

Arcidiacono, P. (2004): "Ability sorting and the returns to college major," *Journal of Econometrics*, 121, 343 – 375, higher education (Annals issue).

Arcidiacono, P. and P. B. Ellickson (2011): "Practical Methods for Estimation of Dynamic Discrete Choice Models," *Annual Review of Economics*, 3, 363–394.

Arcidiacono, P. and R. A. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, 79, 1823–1867.

——— (2019): "Nonstationary dynamic models with finite dependence," *Quantitative Economics*, 10, 853–890.

——— (2020): "Identifying dynamic discrete choice models off short panels," *Journal of Econometrics*, 215, 473 – 485.

Arnold, M. (1999): "Search, Bargaining and Optimal Asking Prices," *Real Estate Economics*, 27, 453–481.

Attanasio, O. P., R. Bottazzi, H. W. Low, L. Nesheim, and M. Wakefield (2012): "Modelling the demand for housing over the life cycle," *Review of Economic Dynamics*, 15, 1–18.

Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75, 1331–1370.

Bajari, P., P. Chan, D. Krueger, and D. Miller (2013): "A Dynamic model of Housing Demand: Estimation and Policy implications," *International Economic Review*, 54, 409–442.

Bento, A. M., L. H. Goulder, M. R. Jacobsen, and R. H. von Haefen (2009): "Distributional and Efficiency Impacts of Increased US Gasoline Taxes," *American Economic Review*, 99, 667–99.

Berger, D., V. Guerrieri, G. Lorenzoni, and J. Vavra (2018): "House Prices and Consumer Spending," *Review of Economic Studies*, 85, 1502–1542.

Binmore, K., M. J. Osborne, and A. Rubinstein (1992): "Chapter 7 Noncooperative models of bargaining," Elsevier, vol. 1 of *Handbook of Game Theory with Economic Applications*, 179 – 225.

Binmore, K., A. Shaked, and J. Sutton (1989): "An Outside Option Experiment," *The Quarterly Journal of Economics*, 104, 753–770.

Blevins, J. R. (2014): "Nonparametric identification of dynamic decision processes with discrete and continuous choices," *Quantitative Economics*, 5, 531–554.

Blundell, R., M. Browning, and C. Meghir (1994): "Consumer Demand and the Life-Cycle Allocation of Household Expenditures," *The Review of Economic Studies*, 61, 57–80.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016a): "Female Labor Supply, Human Capital, and Welfare Reform," *Econometrica*, 84, 1705–1753.

Blundell, R. and T. Macurdy (1999): "Chapter 27 - Labor Supply: A Review of Alternative Approaches," Elsevier, vol. 3 of *Handbook of Labor Economics*, 1559 – 1695.

Blundell, R., T. Magnac, and C. Meghir (1997): "Savings and Labor-Market Transitions," *Journal of Business and Economic Statistics*, 15, 153–64.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98, 1887–1921.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016b): "Consumption Inequality and Family Labor Supply," *American Economic Review*, 106, 387–435.

Bono, P.-H. and A. Trannoy (2019): "The Impact of the 'Scellier' Income Tax Relief on Building Land Prices in France," *Economie et Statistique / Economics and Statistics*, 91–114.

Bontemps, C., F. Cherbonnier, and T. Magnac (2019): "Housing general equilibrium," *Working Paper*.

Bérard, G. and A. Trannoy (2018): "The impact of the 2014 increase in the real estate transfer taxes on the French housing market," *Economie et Statistique*.

Browning, M., M. Gørtz, and S. Leth-Petersen (2013): "Housing Wealth and Consumption: A Micro Panel Study," *The Economic Journal*, 123, 401–428.

Bruneel-Zupanc, C. (2021): "Discrete-Continuous Dynamic Choice Models: Identification and Conditional Choice Probability Estimation," *Working Paper*.

Carrillo, P. E. (2012): "An Empirical Stationary Equilibrium Search Model Of The Housing Market," *International Economic Review*, 53, 203–234.

——— (2013): "To Sell or Not to Sell: Measuring the Heat of the Housing Market," *Real Estate Economics*, 41, 310–346.

Carroll, C. D. (2006): “The method of endogenous gridpoints for solving dynamic stochastic optimization problems,” *Economics Letters*, 91, 312 – 320.

Case, K. E. and R. J. Shiller (1988): “The behavior of home buyers in boom and post-boom markets,” *New England Economic Review*, 29–46.

——— (2003): “Is There a Bubble in the Housing Market?” *Brookings Papers on Economic Activity*, 34, 299–362.

Cazenave-Lacrouts, M.-C., D. Guillas, G. Lebrault, and B. Mordier (2019): “10 patrimoine total,” *INSEE Focus*, 176.

Chapelle, G., B. Vignolles, and C. Wolf (2018): “Impact of a Housing Tax Credit on Local Housing Markets: Evidence from France,” *Annals of Economics and Statistics*, 101–131.

Chen, Y. and R. W. Rosenthal (1996a): “Asking Prices as Commitment Devices,” *International Economic Review*, 37, 129–155.

——— (1996b): “On the Use of Ceiling-Price Commitments by Monopolists,” *RAND Journal of Economics*, 27, 207–220.

Chernozhukov, V., I. Fernández-Val, W. Newey, S. Stouli, and F. Vella (2020): “Semiparametric estimation of structural functions in nonseparable triangular models,” *Quantitative Economics*, 11, 503–533.

Chernozhukov, V. and C. Hansen (2005): “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73, 245–261.

——— (2006): “Instrumental quantile regression inference for structural and treatment effect models,” *Journal of Econometrics*, 132, 491–525.

——— (2008): “Instrumental variable quantile regression: A robust inference approach,” *Journal of Econometrics*, 142, 379 – 398.

Chesher, A. (2003): “Identification in Nonseparable Models,” *Econometrica*, 71, 1405–1441.

Chiong, K. X., A. Galichon, and M. Shum (2016): “Duality in dynamic discrete-choice models,” *Quantitative Economics*, 7, 83–115.

Comte, F. and C. Lacour (2011): "Data-driven density estimation in the presence of additive noise with unknown distribution," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 601–627.

Crawford, G. S., O. Shcherbakov, and M. Shum (2019): "Quality Overprovision in Cable Television Markets," *American Economic Review*, 109, 956–95.

De Groote, O. and F. Verboven (2019): "Subsidies and Time Discounting in New Technology Adoption: Evidence from Solar Photovoltaic Systems," *American Economic Review*, 109, 2137–72.

de Wit, E. R. and B. van der Klaauw (2013): "Asymmetric information and list-price reductions in the housing market," *Regional Science and Urban Economics*, 43, 507–520.

D'Haultfœuille, X. and P. Février (2015): "Identification of Nonseparable Triangular Models With Discrete Instruments," *Econometrica*, 83, 1199–1210.

Dubin, J. and D. L. McFadden (1984): "An econometric analysis of residential electric appliance holdings and consumption," *Econometrica*, 52, 345–362.

Eckstein, Z. and K. I. Wolpin (1989): "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience," *The Review of Economic Studies*, 56, 375–390.

Escanciano, J. C., S. Hoderlein, A. Lewbel, O. Linton, and S. Srisuma (2015): "Nonparametric Euler equation identification and estimation," cemmap working paper CWP61/15, London.

Etheridge, B. (2019): "House Prices and Consumption inequality," *International Economic Review*, 60, 1781–1822.

Fack, G. (2006): "Are housing benefit an effective way to redistribute income? Evidence from a natural experiment in France," *Labour Economics*, 13, 747–771.

Feng, Q., Q. Vuong, and H. Xu (2020): "Estimation of Heterogeneous Individual Treatment Effects With Endogenous Treatments," *Journal of the American Statistical Association*, 115, 231–240.

GOBILLON, L. AND D. LE BLANC (2008): "Economic effects of upfront subsidies to ownership: The case of the Prêt à Taux Zéro in France," *Journal of Housing Economics*, 17, 1–33.

GRISLAIN-LETREMY, C. AND C. TRÉVIEN (2014): "The Impact of Housing Subsidies on the Rental Sector : The french Example," Working Papers 2014-24, Center for Research in Economics and Statistics.

HAN, L. AND W. C. STRANGE (2014): "Bidding Wars for Houses," *Real Estate Economics*, 42, 1–32.

——— (2015): *The Microstructure of Housing Markets*, Elsevier, vol. 5 of *Handbook of Regional and Urban Economics*, chap. 0, 813–886.

——— (2016): "What is the role of the asking price for a house?" *Journal of Urban Economics*, 93, 115–130.

HANEMANN, M. (1984): "Discrete-Continuous Models of Consumer Demand," *Econometrica*, 52, 541–61.

HANSEN, L. P. AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

HAURIN, D. (1988): "The Duration of Marketing Time of Residential Housing," *Real Estate Economics*, 16, 396–410.

HAYFIELD, T. AND J. S. RACINE (2008): "Nonparametric Econometrics: The np Package," *Journal of Statistical Software*, 27.

HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

HECKMAN, J. J., J. E. HUMPHRIES, AND G. VERAMENDI (2016): "Dynamic treatment effects," *Journal of Econometrics*, 191, 276 – 292, innovations in Measurement in Economics and Econometrics.

HECKMAN, J. J. AND T. E. MACURDY (1980): "A Life Cycle Model of Female Labour Supply," *The Review of Economic Studies*, 47, 47–74.

Heckman, J. J. and S. Navarro (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136, 341 – 396, the interface between econometrics and economic theory.

Heckman, J. J., J. Smith, and N. Clements (1997): "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *The Review of Economic Studies*, 64, 487–535.

Heckman, J. J. and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation1," *Econometrica*, 73, 669–738.

Hong, H. and M. Shum (2010): "Pairwise-Difference Estimation of a Dynamic Optimization Model," *The Review of Economic Studies*, 77, 273–304.

Horowitz, J. (1992): "The Role of the List Price in Housing Markets: Theory and an Econometric Model," *Journal of Applied Econometrics*, 7, 115–29.

Hotz, V. J. and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

Hotz, V. J., R. A. Miller, S. Sanders, and J. Smith (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, 61, 265–289.

Hu, Y. and M. Shum (2012): "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171, 32 – 44.

Iacoviello, M. and M. Pavan (2013): "Housing and debt over the life cycle and over the business cycle," *Journal of Monetary Economics*, 60, 221–238.

Imbens, G. W. and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica*, 77, 1481–1512.

Iskhakov, F., T. H. Jørgensen, J. Rust, and B. Schjerning (2017): "The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks," *Quantitative Economics*, 8, 317–365.

Kamionka, T. and G. Lacroix (2018): "Homeownership, Labour Market Transitions and Earnings," CIRANO Working Papers 2018s-35, CIRANO.

Kasahara, H. and K. Shimotsu (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77, 135–175.

Keane, M. P. and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.

Khorunzhina, N. (2021): "Intratemporal Nonseparability between Housing and Nondurable Consumption: Evidence from Reinvestment in Housing Stock," *Journal of Monetary Economics*, 117, 658–670, published online: 14 April 2020.

Khorunzhina, N. and R. A. Miller (2019): "American Dream Delayed: Shifting Determinants of Homeownership," *Working Paper*.

Knight, J. R. (2002): "list price, Time on Market, and Ultimate Selling Price: Causes and Effects of list price Changes," *Real Estate Economics*, 30, 213–237.

Li, W., H. Liu, F. Yang, and R. Yao (2016): "Housing over time and over the life cycle: a structural estimation," *International Economic Review*, 57, 1237–1260.

Li, W. and R. Yao (2007): "The Life-Cycle Effects of House Price Changes," *Journal of Money, Credit and Banking*, 39, 1375–1409.

Lipsitz, M., A. Belloni, V. Chernozhukov, and I. Fernandez-Val (2017): "Quantreg.nonpar: an R package for performing nonparametric series quantile regression," CeMMAP working papers CWP29/17, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Liu, C. H., A. Nowak, and S. Rosenthal (2014): "Bubbles, Post-Crash Dynamics, and the Housing Market," Working Papers 14-18, Department of Economics, West Virginia University.

Magnac, T. and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816.

Matzkin, R. L. (1993): "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 58, 137 – 168.

——— (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

——— (2007): "Nonparametric identification," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier, vol. 6 of *Handbook of Econometrics*, chap. 73.

——— (2008): "Identification in Nonparametric Simultaneous Equations Models," *Econometrica*, 76, 945–978.

McFadden, D. (1980): "Econometric Models for Probabilistic Choice Among Products," *The Journal of Business*, 53, 13–29.

Merlo, A. and F. Ortalo-Magne (2004): "Bargaining over residential real estate: evidence from England," *Journal of Urban Economics*, 56, 192–216.

Merlo, A., F. Ortalo-Magne, and J. Rust (2015): "The Home Selling Problem: Theory And Evidence," *International Economic Review*, 56, 457–484.

Miller, N. and M. Sklarz (1987): "Pricing Strategies and Residential Property Selling Prices," *Journal of Real Estate Research*, 2, 31–40.

Muggeo, V. (2018): "Using the R package quantregGrowth: some examples," .

Murphy, A. (2018): "A Dynamic Model of Housing Supply," *American Economic Journal: Economic Policy*, 10, 243–67.

Newey, W. K. (2007): "Nonparametric continuous/discrete choice models," *International Economic Review*, 48, 1429–1439.

Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Odean, T. (1998): "Volume, Volatility, Price, and Profit When All Traders Are Above Average," *The Journal of Finance*, 53, 1887–1934.

Ortalo-Magné, F. and S. Rady (2006): "Housing Market Dynamics: On the Contribution of Income Shocks and Credit Constraints*," *The Review of Economic Studies*, 73, 459–485.

Paz-Pardo, G. (2021): "Homeownership and portfolio choice over the generations," *Working Paper*.

Piazzesi, M. and M. Schneider (2009): "Momentum Traders in the Housing Market: Survey Evidence and a Search Model," *American Economic Review*, 99, 406–411.

Pizzinelli, C. (2017): "Housing, borrowing constraints, and labor supply over the life cycle," *Working Paper*.

Price, W. L. (1983): "Global optimization by controlled random search," *Journal of Optimization Theory and Applications*, 40, 333–348.

Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50, 97–109.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

——— (1994): "Chapter 51 Structural estimation of markov decision processes," Elsevier, vol. 4 of *Handbook of Econometrics*, 3081 – 3143.

Salant, S. (1991): "For Sale by Owner: When to Use a Broker and How to Price the House," *The Journal of Real Estate Finance and Economics*, 4, 157–73.

Shaked, A. and J. Sutton (1984): "Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model," *Econometrica*, 52, 1351–64.

Sommer, K. and P. Sullivan (2018): "Implications of US Tax Policy for House Prices, Rents, and Homeownership," *American Economic Review*, 108, 241–74.

Öst, C. E. (2012): "Housing and children: simultaneous decisions?— a cohort study of young adults' housing and family formation decision," *Journal of Population Economics*, 25, 349–366.

Timmins, C. (2002): "Measuring the Dynamic Efficiency Costs of Regulators' Preferences: Municipal Water Utilities in the Arid West," *Econometrica*, 70, 603–629.

Torgovitsky, A. (2015): "Identification of Nonseparable Models Using Instruments With Small Support," *Econometrica*, 83, 1185–1197.

Ulyssea, G. (2018): "Firms, Informality, and Development: Theory and Evidence from Brazil," *American Economic Review*, 108, 2015–47.

Vuong, Q. and H. Xu (2017): "Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity," *Quantitative Economics*, 8, 589–610.