Spring 2022

Essays in International Trade and Spatial Economics

Robert John Finlay
Yale University Graduate School of Arts and Sciences, robert.finlay@yale.edu

Abstract

Essays in International Trade and Spatial Economics

John Finlay

2022

The first chapter of this dissertation studies the relationship between credit constraints, exporting, and misallocation. When financial markets are imperfect, credit constraints hinder firm growth, distort the allocation of inputs, and lower aggregate productivity. Such constraints are particularly costly when they bind for the most productive firms. I focus on exporters, a group of firms that the international trade literature has identified as uniquely productive. Are exporters credit constrained? Do policies that target exporters which are ubiquitous, particularly in developing countries mitigate or worsen misallocation? I answer these questions by combining a natural experiment in India with a quantitative model.

I exploit a directed credit policy in India as a source of exogenous variation in credit supply. Eligibility was determined by a cutoff in physical capital, allowing me to estimate its effects with a regression discontinuity design. Exporters responded strongly to the relaxation of credit constraints caused by the policy: they borrowed more, hired more workers, and sold more output. By contrast, I find no effect on non-exporters. I conclude that credit constraints must be relatively more important for exporting firms.

Motivated by this finding, I build a model of heterogeneous entrepreneurs that links credit constraints and the decision to export. Two forces shape exporting: productivity and access to credit. Which of these dominates determines the relative

importance of credit constraints across exporting and non-exporting firms. I estimate the model using the natural experiment, and find that the decision to export is strongly driven by productivity. The result is that credit constraints bind for many exporters; in the model, 37% of exporters and 8% of non-exporters are constrained. Inputs are misallocated and exporters are inefficiently small. In counterfactual experiments, I find that directly relaxing the credit constraint of exporters raises aggregate productivity by 3.33%. However, I also show that subsidizing exporter employment worsens misallocation, because relatively unproductive, unconstrained exporters are the primary beneficiaries.

The second chapter considers quite a different topic: the relationship between income inequality and spatial sorting. Housing expenditure shares decline with income. A households skill level determines its income, and therefore its housing expenditure share, its sensitivity to housing costs and its preferences over different locations. The result is spatial sorting driven by differences in cost-of-living between skill groups. Increases in the aggregate skill premium amplify these differences and intensify sorting. To quantify this mechanism, I augment a standard quantitative spatial model with flexible nonhomothetic preferences, disciplining the strength of the housing demand channel using consumption microdata. I find that the rising skill premium caused 23% of the increase in spatial sorting by skill since 1980.

Essays in International Trade and Spatial Economics

A Dissertation

Presented to the Faculty of the Graduate School

Of

Yale University

In Candidacy for the Degree of

Doctor of Philosophy

By

John Finlay

Dissertation Director: Michael Peters

May 2022

Contents

Acknowledgements **iv**

1 Exporters, Credit Constraints, and Misallocation **1**

1.1 Introduction 1

1.2 Are Exporters Credit Constrained? 7

1.3 Model 19

1.4 Estimation 30

1.5 Does Targeting Exporters Raise Productivity? 44

1.6 Conclusion 55

2 Housing Demand, Inequality, and Spatial Sorting [1] **56**

2.1 Estimating the Income Elasticity of Housing Demand 60

2.2 Model 72

2.3 Calibration 83

2.4 The Skill Premium and Sorting 87

2.5 Real inequality 92

2.6 Conclusion 94

[1] This is joint work with Trevor Williams

Acknowledgments

First and foremost, I am grateful to my advisors Michael Peters, Samuel Kortum, and Lorenzo Caliendo, for their guidance and encouragement. More broadly, I thank all the faculty and students in Yale's economics department for their insightful comments across various seminars, workshops, and breakfasts: in particular, Hortense Fong, Ahyan Panjwani, Vitor Possebom, Jaehee Song, and Trevor Williams were all kind enough to read some or all of the chapters below. I also thank my friends and family for their support and kindness.

Chapter 1

Exporters, Credit Constraints, and Misallocation

1.1 Introduction

Exporters are different: they use more capital and labor, sell more output, and are more productive than other firms.[1] Policies that target exporters, through subsidies or via favorable access to credit, are widespread, particularly in developing countries.[2] Furthermore, trade liberalization allows exporting firms to expand at the expense of non-exporters (Melitz 2003). In both cases, the result is a reallocation of capital and labor towards exporters, and the desirability of such a reallocation hinges on whether it raises or lowers aggregate productivity. In turn, the effect of this reallocation on aggregate productivity depends crucially on the presence of frictions that distort the size of exporters relative to non-exporters.

In this paper, I combine a natural experiment in India with a quantitative model to show that credit constraints are just such a friction. Empirically, I find that ex-

[1] See Bernard and Jensen (1999) for the US, Aw, Chung, and Roberts (2000) for Taiwan, and Clerides, Lach, and Tybout (1998) for Colombia, Mexico and Morocco.

[2] For examples, see Itskhoki and Moll (2019).

porters respond to an exogenous increase in credit supply by borrowing more, hiring more workers, and selling more output. In contrast, similarly sized non-exporters do not respond to this shock. I use these results to estimate a dynamic model in which heterogeneous entrepreneurs produce output subject to credit constraints and decide to export. The estimated model implies that credit constraints bind for many exporters but few non-exporters, and that exporters are on average inefficiently small. Reallocating inputs towards them therefore has the potential to raise aggregate productivity. Finally, I use the estimated model to study the productivity effects of a range of policies that target exporters, distinguishing between their effects on misallocation between exporters and non-exporters, and their effects on misallocation within each of these sets of firms. I find that both dimensions of misallocation are quantitatively important in determining the success or failure of the policies I study. Below I discuss each aspect of the paper in detail.

I begin by using India's Priority Sector Lending (PSL) policy as a source of exogenous variation in firms' credit constraints. Banks were incentivized to lend to firms eligible for PSL, and a cutoff rule determined eligibility. Manufacturing firms with capital below 50 million rupees (roughly 1 million USD) were eligible, while firms with capital above this level were not. This policy allows me to explicitly compare the importance of constraints across exporters and non-exporters, precisely because it was not contingent on a firm's export status. Using a regression discontinuity design, I show that eligible exporters borrowed 33% more, hired 25% more workers, and sold 22% more output, while I find no effect of PSL on non-exporters. I further show that PSL did not cause any change in eligible firms' export choices on either the intensive or extensive margin. Instead, it allowed exporters to expand both their domestic and export sales symmetrically. This fact suggests that export sales *per se* are not uniquely distorted by credit constraints. Instead, the type of firm that chooses to export is particularly likely to find credit constraints binding.

Motivated by these findings, I build a model which connects credit constraints and exporting. Entrepreneurs differ in their productivity and in their fixed costs of exporting and accumulate physical capital and liquid assets over time. They must pay for the labor they use before production takes place, and do so using either liquid assets or by borrowing using physical capital as collateral. Some entrepreneurs — those with relatively high productivity, but relatively low levels of liquid assets and physical capital — will hit a binding credit constraint, where they wish to hire more workers but cannot borrow to do so. Finally, entrepreneurs choose whether or not to export.

The model highlights two factors in the decision to export. More productive entrepreneurs find it more worthwhile to overcome the fixed cost of exporting because their sales abroad will be large. All else equal, more productive entrepreneurs are also more likely to be constrained, and so this force tends to make exporters more constrained. However, entrepreneurs with better access to credit — determined by their stocks of liquid assets and physical capital — are also more likely to export, because they are more able to expand to take advantage of the larger market they can access by exporting. This force tends to make exporters less constrained. Therefore, which of these two forces dominates is crucial in determining the relative importance of credit constraints for exporters.

I estimate the model by targeting the results of the natural experiment. The natural experiment disciplines the two forces mentioned above: whether the decision to export is mainly driven by productivity or by access to credit. Since in the natural experiment exporters responded strongly to a change in their credit constraints, the estimation infers that many exporters are constrained, and therefore that productivity is the main driver of the decision to export. The estimated model implies that 37% of exporting firms are at a binding credit constraint, compared to only 8% of non-exporters. Exporters are inefficiently small and have high marginal products of capital and labor; for example, the marginal revenue product of labor is roughly 9% higher among exporters. This difference in marginal products implies that reallocating inputs towards exporters has the potential to raise aggregate productivity.

I use the estimated model to study two policies that encourage exporters to expand. The first policy directly relaxes the credit constraint of exporters, while the second subsidizes their employment. While these policies cause comparable amounts of reallocation towards exporters, I show that they have sharply different consequences for aggregate productivity. The credit policy allows constrained exporters, who have relatively high marginal products, to expand, thus lowering misallocation. In the long run, this policy raises aggregate productivity by 3.33%. In contrast, the employment subsidy primarily benefits unconstrained exporters with low marginal products because these are the firms most able to expand in response to the subsidy. Thus it worsens misallocation and lowers aggregate productivity. My results highlight that subsidies struggle to undo the misallocation created by credit constraints, even when targeted towards a group of firms (exporters) in which such constraints are prevalent. To be effective, subsidies must encourage the most productive firms to expand; but, almost by definition, constrained firms cannot do so. On the other hand, directly tackling the source of the distortion yields substantial gains.

Finally, I consider a third intervention that implicitly targets exporters: lowering trade costs. As with the employment subsidy above, I find that heterogeneity within the set of exporters largely offsets any productivity gains from reallocation towards exporters on average. I contrast my results with those obtained from a model in which misallocation is the result of exogenous wedges, as in Hsieh and Klenow (2009). I show that such a model substantially overstates the productivity enhancing effects of reductions in trade costs.

This paper relates to four broad literatures. First, I contribute to an empirical literature that measures the firm-level effects of credit constraints. Focusing on exporters, Amiti and Weinstein (2011) and Paravisini et al. (2015) show that shocks to bank health are transmitted to export sales, while Zia (2008) studies the removal of subsidized export credit in Pakistan. Relative to these papers, I study a policy that affected both exporters and similarly sized non-exporters, allowing me to compare its effects across these two groups. My paper is also connected to a literature that analyzes the effects of the Priority Sector Lending policy (Banerjee and Duflo 2014; Kapoor, Ranjan, and Raychaudhuri 2017). Particularly relevant is Rotemberg (2019). That paper studies the same policy, and develops an empirical methodology to estimate its indirect effects via general equilibrium. My focus is instead on heterogeneity in the policy's direct effects, in particular across exporting and non-exporting firms. I show that this heterogeneity is informative about the determinants of the decision to export in a model in which entrepreneurs differ in their productivity and assets, and use it to estimate the model's key parameters.

Second, this paper is related to a literature that incorporates financial frictions into models of international trade. Manova (2012) and Leibovici (2021) show that the pattern of aggregate trade flows across countries and sectors is consistent with models in which financial frictions inhibit trade. Chaney (2016) links financial constraints to exchange rate fluctuations. Kohn, Leibovici, and Szkup (2014) study how these frictions affect the dynamics of new exporters and Brooks and Dovis (2020) show that conclusions about how they interact with the gains from trade are sensitive to exactly how credit constraints are modeled. These papers point to a variety of ways credit constraints might interact with exporting; by distorting the extensive or intensive margins, or simply by limiting overall firm size. The results of my natural experiment support a model in which export sales *per se* are not uniquely constrained. Rather, credit constraints limit the ability of some firms to expand overall, and these constrained firms are likely to be exporters.

Third, I contribute to a literature that studies the gains from trade in the presence of misallocation (Berthou et al. 2019; Bai, Jin, and Lu 2019). In these papers, misallocation results from exogenous distortions in input markets, whereas in my model, misallocation is endogenously generated by credit constraints, as well as adjustment costs in physical capital. I show that this distinction matters. In my model, the firms most able to expand in response to falling trade costs are unconstrained exporters with relatively low marginal products. As a result, misallocation within the set of exporting firms rises, limiting the overall gains from trade. I show that this force vanishes when misallocation is the result of exogenous wedges, because, conditional on export status, a reduction in trade costs affects all firms symmetrically. Hence, my results highlight the importance of explicitly modeling the source of misallocation for understanding how it will interact with a given policy change.

Fourth, a large literature in macroeconomics links financial constraints and misallocation (Buera, Kaboski, and Shin 2011; Midrigan and Xu 2014; Moll 2014). I show empirically that such constraints distort the decisions of a particularly productive group of firms — exporters. Moreover, in linking a natural experiment to a model of financial constraints and misallocation, my paper is related to Kaboski and Townsend (2011) and Buera, Kaboski, and Shin (2021). While they study microfinance interventions that affect poor households and very small firms, I show that similar constraints are relevant for much larger firms.

The remainder of the paper proceeds as follows. Section 1.2 uses India's PSL policy to estimate the effects of credit constraints on exporting and non-exporting firms. Section 2.2 builds a model of credit constraints and selection into exporting, and Section 2.1 estimates this model by targeting the pattern of treatment effects found in Section 1.2. Section 1.5 explores the policy implications of the estimated model. Finally Section 2.6 concludes.

1.2 Are Exporters Credit Constrained?

In this section I exploit variation in eligibility for a directed credit policy, Priority Sector Lending (PSL), as a source of exogenous variation in the availability of credit. I find that eligible exporting firms borrowed more, hired more workers and sold more output, while eligible domestic firms did not respond in any economically or statistically significant way. My results suggest that credit constraints are more important for exporting firms than for similarly sized non-exporting firms. Subsection 1.2.1 discusses the details of the PSL policy and Subsection 2.1.1 introduces my data. In Subsection 1.2.3 I outline my estimation strategy, and 2.1.4 presents results.

1.2.1 Priority Sector Lending

Under India's Priority Sector Lending (PSL) policy, all banks are obliged to allocate at least 40% of net credit to the 'priority sector', which includes agriculture, transport, and small businesses (Banerjee and Duflo 2014). If a bank fails to reach this quota, it faces financial penalties.[3] Therefore, banks have a strong incentive to lend to firms in the priority sector, and priority sector firms enjoy favorable access to credit. Variation in PSL eligibility across firms thus has the potential to act as a source of exogenous variation in access to credit.

The priority sector includes manufacturing firms with plant and machinery (a subset of physical capital) below a certain cutoff. This cutoff has moved around over time. For example, Banerjee and Duflo (2014) studied the effects of an increase in the cutoff from 6.5 million rupees to 30 million rupees in 1998, as well as a subsequent decrease in 2000. I focus on a later change in the policy, when in 2007 the cutoff was raised from 10 to 50 million rupees, roughly 1.1 million USD. Thus in 2007, firms with plant and machinery below 50 million rupees became eligible for PSL, while

[3]Banerjee and Duflo (2014) show that for the bank they study, the share of lending to the priority sector is always close to 40%, suggesting that this constraint is binding.

firms with plant and machinery just above this level remained ineligible.

In principle, banks could have increased lending to the newly eligible firms in two ways. First, if firms were credit constrained, banks could have offered to raise their credit limits. Second, banks could have lowered the cost of borrowing. In Appendix A.1 I investigate the second possibility and find that PSL eligibility did not lower firms' borrowing costs. This is consistent with evidence presented by Banerjee and Duflo (2014), who find that interest rates did not fall for eligible firms. Therefore I interpret the effects of PSL eligibility reported below as evidence of the presence of credit constraints.

Since I will report separate results for exporting and non-exporting firms, it is important to note that PSL did not distinguish between these two types of firms. Credit extended to exporters was counted towards the quota if and only if the firm had plant and machinery below 50 million rupees. Thus, both types of firm were subject to the same policy; differences in treatment effects across these two groups therefore reflect differences between exporting and non-exporting firms, rather than differences in the application of the PSL policy.

Finally, firms eligible for PSL were also eligible for a number of other programs run by the Ministry for Micro, Small, and Medium Enterprises (MSME). In practice the vast majority (70%) of MSME's budget was devoted to credit guarantee and support schemes (Rotemberg 2019). We would expect these credit guarantees to have effects similar to those of PSL, and since my goal is not to measure the effects of PSL *per se*, but rather to use eligibility as a source of exogenous variation in credit supply, the presence of such credit guarantee schemes does not present a problem. MSME also provided entrepreneurs with access to to training programs, which would be expected to raise firm productivity. Rotemberg (2019) finds that eligibility had a negligible effect on firm productivity, suggesting such training programs were unimportant. I therefore follow Banerjee and Duflo (2014) and interpret eligibility as a shock to firms' access to credit.

1.2.2 Data

My main analysis relies on the Prowess dataset, compiled by the Centre for Monitoring the Indian Economy (CMIE). This is a panel of firms beginning in 1980 whose source is audited financial statements. I use information on the value of plant and machinery, total borrowing, wage bills, total sales, and export sales.[4] Table 1.1 reports summary statistics from Prowess for 2007. Prowess is not representative of the universe of Indian firms. Instead, it focuses on larger firms, and among these firms it has fairly complete coverage. For example, firms in Prowess account for $60-70\%$ of economic activity in the organized industrial sector and 75% of corporate taxes collected by the Government of India (De Loecker et al. 2016). Prowess does a good job of capturing firms affected by the PSL policy and is therefore ideal for this paper. Panel (a) of Figure 1.1 shows the density of plant and machinery across firms in 2007 alongside the cutoff for PSL eligibility. Roughly 35% of firms in 2007 had plant and machinery below 50m rupees and were therefore eligible for PSL. Panel (b) of Figure 1.1 shows that PSL affected both exporting and non-exporting firms; 18.7% of exporters and 48.6% of non-exporters in Prowess were eligible for PSL in 2007. See Appendix A.1 for more details.

1.2.3 Research Design and Estimation

The crucial feature of the policy described in Subsection 1.2.1 is that a firm's eligibility changed discretely as plant and machinery crossed the 50 million rupee

[4] For most firms, Prowess does not report employment separately from wage bills. I therefore assume that all firms face the same wages, so that a firm's employment is proportional to its wage bill.

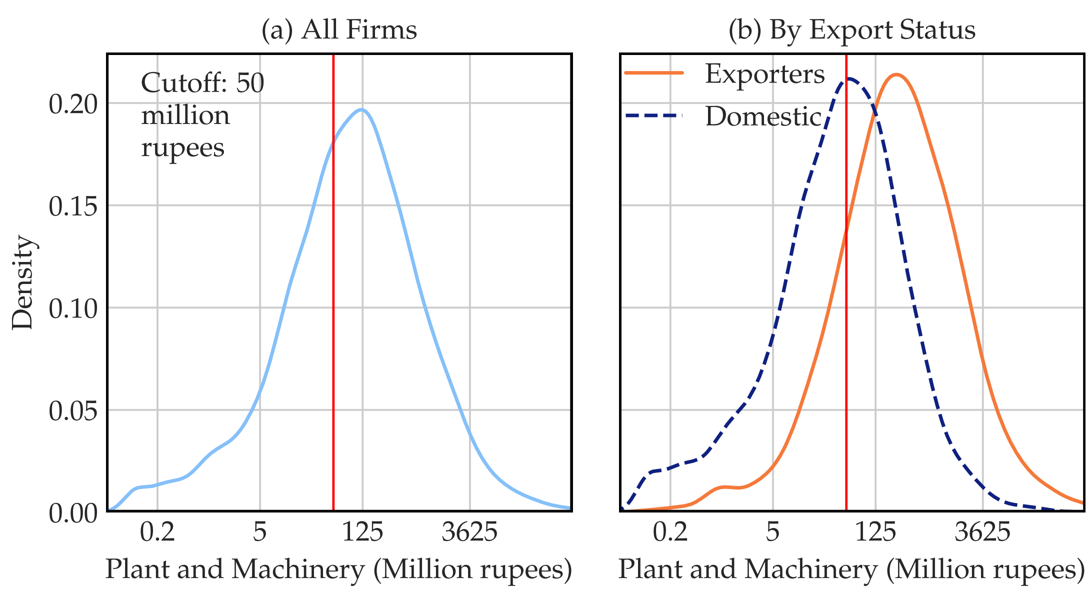
Table 1.1: Summary Statistics

	(1) *All Firms*	(2) *Exporters*	(3) *Non-exporters*
Number of Firms	7349	3449	3900
Median Sales	505	1257	205
Median Plant and Machinery	111	257	53
Plant and Machinery Below 50 Million, %	34.6	18.7	48.6

Source: Prowess dataset, manufacturing firms, 2007.

Note: Units for sales and plant and machinery are millions of rupees. 'Exporters' defined as firms with positive export sales in 2007.

Figure 1.1: Plant and Machinery Distribution



Source: Prowess dataset, manufacturing firms, 2007. *Notes*: Panel (a) shows the density of (log) plant and machinery in all manufacturing firms in 2007. The vertical line shows the 50 million rupee cutoff for PSL eligibility. Panel (b) shows the density (log) plant and machinery in exporting and non-exporting firms, where export status is defined using sales in 2007 threshold. I consider models of the form

$$\mathbb{E}\left[y_{it}|x_{i0}\right] = f_y(x_{i0}) + \beta_y \mathbb{I}\{x_{i0} \leq c\}$$

where i indexes firms and t indexes years. y_{it} is the outcome of interest — (log) loans, employment, and sales. x_{i0} is log plant and machinery in year 0, which I take to be 2007, and $c = \log(50)$ is the cutoff for PSL eligibility. β_y is the parameter of interest and measures the effect of PSL eligibility on the outcome y. Note that this is an average treatment effect for the set of firms with plant and machinery equal to 50 million rupees, rather than an average over all firms.

The effect of PSL eligibility, β_y, is identified under the assumption that the function $f_y(x_{i0})$ is continuous at $x_{i0} = c$. f_y represents the expected value of y_{it} in the absence of PSL. Assuming continuity of f_y is therefore equivalent to assuming that without PSL the outcomes y_{it} would have varied smoothly across the cutoff c. Any discontinuous jumps we observe can then be attributed to the effects of PSL.

As is standard in the regression discontinuity literature (Imbens and Lemieux 2008), I approximate the unknown function f_y using a local linear regression, so that estimating β_y reduces to estimating

$$y_{it} = \varphi_0 + \varphi_1(x_{i0} - c)\mathbb{I}\{x_{i0} \leq c\} + \varphi_2(x_{i0} - c)\mathbb{I}\{x_{i0} > c\} + \beta_y \mathbb{I}\{x_{i0} \leq c\}$$

by weighted least squares, with the weights determined by bandwidth and kernel choices. In choosing these values I follow Calonico, Cattaneo, and Titiunik (2014) — see Appendix A.3 for details. In some specifications I include controls — year and industry fixed effects, and lagged values of the outcomes y_{it}. When I do so, I follow the advice of Calonico et al. (2019) and include them additively, without interacting them with the cutoff dummy.

1.2.4 Results

Main Results

Table 1.2 shows my main results. Each column reports results for three different outcomes: loans, employment, and sales. I categorize firms into exporters and non-exporters based on their sales in 2007. All outcomes are measured in logs, so the point estimates can be interpreted as the percentage difference between firms who were just eligible for PSL based on their plant and machinery in 2007 and those who were just ineligible.

Columns (1) and (2) report results for exporters and non-exporters, respectively, with outcomes measured between 2008 and 2012. Column (1) shows that eligible

Table 1.2: Effects of Priority Sector Lending

Outcome	(1) *Exporters*	(2) *Domestic*	(3) *Exporters*	(4) *Domestic*
Loans	0.388**	0.022	0.328**	0.020
	(0.177)	(0.134)	(0.153)	(0.117)
Employment	0.233*	−0.027	0.246**	0.000
	(0.133)	(0.107)	(0.115)	(0.075)
Sales	0.121	0.033	0.224*	0.110
	(0.163)	(0.128)	(0.122)	(0.113)
Years	2008-12	2008-12	2008-12	2008-12
Pre-policy controls	No	No	Yes	Yes
N	13243	12757	11553	9382

Source: Prowess Dataset, all manufacturing firms, 2005-2012

Note: Columns show results for different specifications; rows show results for different outcomes. Each estimate reports the discontinuity in the outcome at plant and machinery equal to 50 million rupees. Plant and machinery measured in 2007; export status defined using sales in 2007. All outcomes are measured in logs and a positive number indicates a positive effect of being eligible for Priority Sector Lending. (1) and (2) show results with year, industry, and firm age fixed effects. (3) and (4) additionally control for pre-policy outcomes measured in 2005. Standard errors clustered at the firm level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

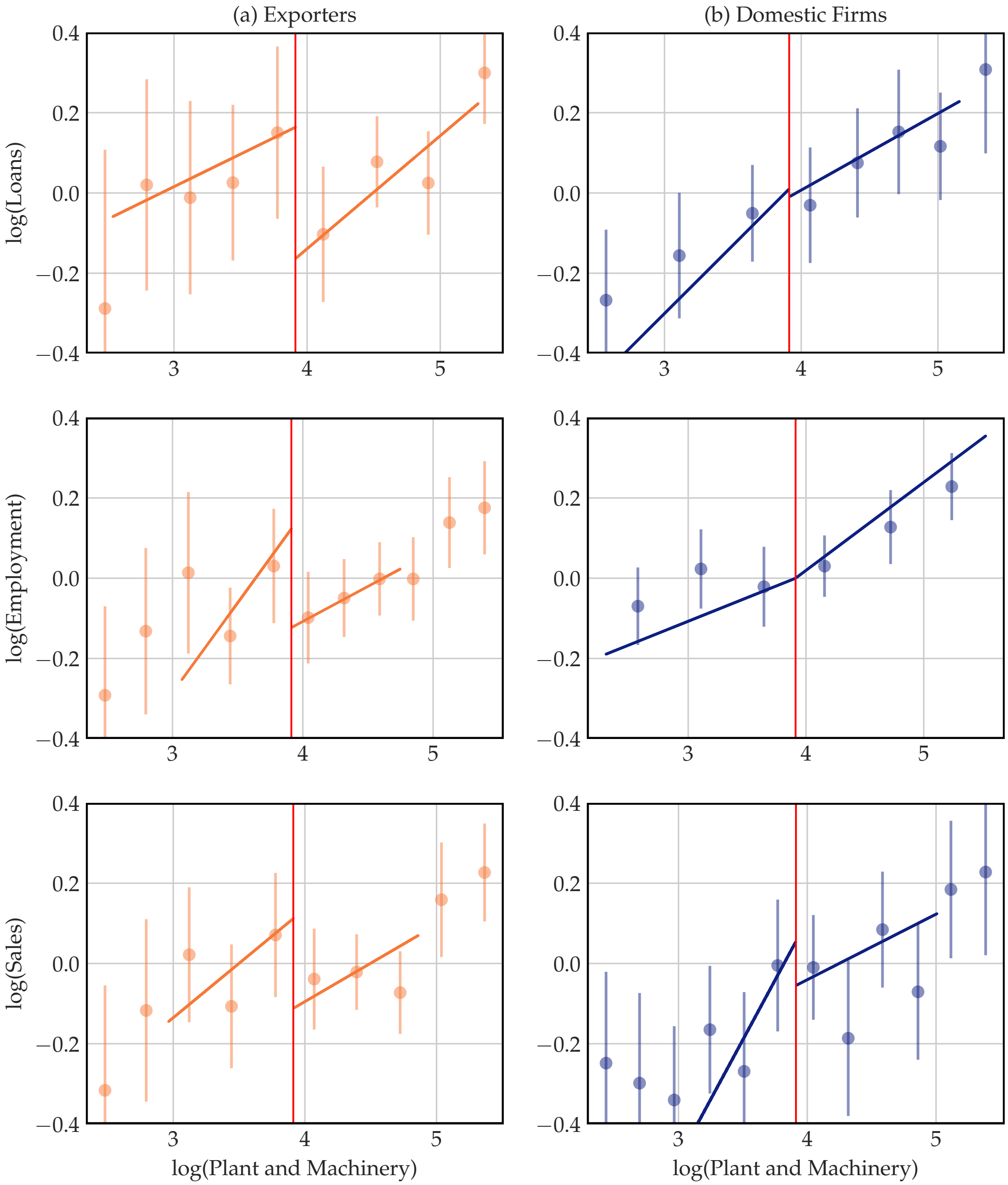
and ineligible exporters look very different. Eligible exporters borrowed 39% more and hired 23% more workers. They also sold more output, but this estimate is not statistically significant. By contrast, PSL eligibility did not have any effect on these outcomes for non-exporters. All the estimates in Column (2) are quantitatively small and statistically insignificant.

Next, in Columns (3) and (4) I control for outcomes in the pre-policy period, which I take to be 2005.[5] That is, each regression now includes log loans, employment and sales in 2005 as controls. Adding these controls has two benefits. First, by absorbing variation in the outcomes which existed prior to the policy, they allow me to estimate the effects of PSL eligibility more precisely. Second, these results provide a check on my identifying assumption. If the results with pre-policy controls differed sharply from those without, that would suggest that the results in Columns (1) and (2) reflected pre-existing differences between eligible and ineligible firms rather than the causal effect of PSL. Equally, finding similar results when these controls are included would suggest that the results in Columns (1) and (2) do indeed capture this causal effect. Column (3) continues to show that exporters responded strongly to PSL eligibility, by borrowing 33% more, hiring 25% more workers and selling 22% more output, although again this last outcome is more noisily measured. Consistent with Column (2), all the estimates for non-exporters in Column (4) are quantitatively small and statistically insignificant.

Figure 1.2 visualizes the results in (3) and (4) by showing binned scatterplots of the three outcomes in Table 1.2 against plant and machinery, with local linear regressions shown by the solid lines. The plots for exporters in Panel (a) show discontinuous jumps at 50 million rupees, whereby firms just below the cutoff borrowed more, hired more workers and sold more output. In contrast the plots for non-exporters in Panel (b) show small discontinuities with inconsistent signs. Columns (3) and (4) represent

[5] I exclude 2006 because the change in the PSL threshold was announced, but not implemented, in this year.

Figure 1.2: Regression Discontinuity Plots



Source: Prowess dataset, manufacturing firms, 2005-2012. *Notes*: Panel (a) shows results for exporters, Panel (b) for non-exporters. Each plot shows a binned scatterplot of an outcome (log loans, employment or sales), plotted against the log of 2007 plant and machinery, in a window around the cutoff for PSL eligibility. As in Columns (3) and (4) of Table 1.2, I control for year, industry, and firm age FE, as well as pre-policy outcomes. This cutoff is shown by the vertical line. The solid lines on either side of the cutoff plot local linear regressions fitted to the underlying data. Note that in each plot the outcome variable has been shifted by a constant so that the y-axis is centered on zero. The discontinuity in the solid line at the cutoff corresponds to (minus) the treatment effects reported in Columns (3) and (4) of Table 1.2.

my preferred specification, and will serve as targets for the model I estimate in Section 2.1.

Threats to Identification

Above I assumed that the function $f_y(x_{i0})$ was continuous at $x_{i0} = c$ in order to identify β_y, the causal effect of PSL eligibility. In my setting, the leading threat to identification is the manipulation of plant and machinery close to the cutoff. If firms can perfectly choose their 2007 plant and machinery,[6] and if the firms which choose to become eligible for PSL are systematically different than those which do not, then this kind of sorting could bias my results. Below I present two pieces of evidence that suggest that such sorting is not a problem in this case.

First, I check whether firms 'bunched' to the left of the 50 million rupee cutoff, which would suggest manipulation of plant and machinery (McCrary 2008). Figure A.1 in Appendix A.1 shows histograms of plant and machinery close to the cutoff, separated by export status. These histograms show no evidence of bunching. In Appendix A.1 I also report the results of formal statistical tests which do not reject the null of no bunching. If anything, these tests find there are slightly too few firms to the left of the cutoff, just the opposite of what we would expect if firms were manipulating their plant and machinery to become eligible for PSL.

Second, I perform a placebo test. In Table 1.3, I repeat the specification in Columns (1) and (2) of Table 1.2, continuing to use 2007 plant and machinery as the running variable, but now I use outcomes measured before the policy was implemented. In particular, I use outcomes from 2005. The idea here is that if firms that became eligible in 2007 differ from those that did not become eligible only because

[6] Note the qualifier 'perfectly.' Lee (2008) considers a setting in which agents can influence their assignment into treatment, and shows that if this is imperfect, i.e., if eligibility is at least partly determined by some random component, then a regression discontinuity design continues to identify the causal effect of treatment. In my setting, this random component might come from some randomness in the rate at which capital depreciates, for example.

of the causal effect of PSL, then we should detect no effect of PSL eligibility on any outcome before the policy was implemented. Columns (1) and (2) report the results of this exercise for exporters and non-exporters, respectively. All of the estimates are statistically insignificant, and for exporters the sign of the estimates varies across the different outcomes. I conclude that eligible and ineligible firms were not significantly different prior to the introduction of PSL, and only diverged after the policy was implemented. Together with the bunching check reported above, this placebo check suggests that sorting around the cutoff is not driving my results. Instead, they represent the causal effect of becoming eligible for PSL.

Rotemberg (2019) points out that PSL also had indirect effects that operated through changes in equilibrium prices, and that these indirect effects likely varied across sectors. Such equilibrium effects do not pose a threat to identification for my regression discontinuity design, even if they vary across sectors or across exporters and non-exporters. To see this, note that while exposure to equilibrium effects might be correlated with plant and machinery (because, for example, plant and machinery varies systematically across sectors), we would not expect this correlation to jump discontinuously at the 50 million rupee cutoff for PSL eligibility. Therefore, indirect effects do not violate the assumption that the potential outcomes $f_y(x_{i0})$ are continuous functions of plant and machinery at the cutoff.

Robustness Checks

In Appendix A.1 I investigate the robustness of the results in Columns (3) and (4) of Table 1.2. I show that they are not sensitive to the choice of bandwidth; to dropping observations very close to the cutoff for PSL eligibility; or to the years I use to measure the outcome variables. I obtain qualitatively similar point estimates when I drop all controls, although these estimates are noisy.

Are Export Sales Especially Constrained?

Table 1.2 showed that PSL eligibility had a significant effect on borrowing, employment, and sales in exporting firms. This might reflect the fact that exporting is a uniquely finance-intensive activity. For example, Manova (2012) points out that export sales may be particularly dependent on access to external financing because they involve large upfront costs and long lags between production and payment. Table 1.4 investigates this possibility by estimating the effect of PSL eligibility on extensive and intensive margin export decisions. Columns (1) studies the extensive margin of exporting: were eligible firms more likely to enter the export market? The outcome here is a dummy equal to 1 if the firm has positive export sales. As in Columns (3) and (4) of Table 1.2, I control for values of the outcome in the pre-policy period. The

Table 1.3: Effects of Priority Sector Lending: Placebo

	(1)	(2)
Outcome	*Exporters*	*Domestic*
Loans	−0.136	−0.007
	(0.171)	(0.133)
Employment	0.113	−0.075
	(0.143)	(0.120)
Sales	−0.006	0.112
	(0.164)	(0.180)
Years	2005	2005
N	2731	2694

Source: Prowess Dataset, all manufacturing firms, 2005-2012

Note: Columns show results for different specifications; rows show results for different outcomes. Each estimate reports the discontinuity in the outcome at plant and machinery equal to 50 million rupees. Plant and machinery measured in 2007; export status defined using sales in 2007; outcome measured in 2005. All outcomes are measured in logs and a positive number indicates a positive effect of being eligible for Priority Sector Lending. All regressions include year, industry and firm age fixed effects.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table 1.4: Effect of Priority Sector Lending on Exporting

Outcome	(1) *Extensive margin*	(2) *Intensive margin*
PSL Effect	−0.021	0.016
	(0.024)	(0.029)
Years	2008-12	2008-12
N	26000	11790

Source: Prowess Dataset, all manufacturing firms, 2005-2012

Note: Each estimate reports the discontinuity in the outcome at plant and machinery equal to 50 million rupees. Plant and machinery measured in 2007; export status defined using sales in 2007. In (1), the outcome is a dummy equal to 1 if the firm has positive export sales, and the sample is all firms. In (2), the outcome is the share of sales exported, and the sample is all firms with positive export sales. All regressions include year, industry and firm age fixed effects. All regressions also control for values of the outcome in the pre-policy period, i.e 2005.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

result in Column (1) is small and not statistically significant, indicating that PSL eligibility did not affect the extensive margin of exporting. Column (2) shows that the same is true for the intensive margin, i.e., the share of output a firm sells abroad, conditional on being an exporter. Eligible firms increase the share of their sales made abroad by about 1.6 percentage points, but this effect is not statistically significant. Together, the results in Columns (1) and (2) do not show any significant effect of PSL eligibility on either the extensive or intensive margins of exporting.

Summary

PSL eligibility relaxed firms' credit constraints and caused exporters to borrow more, hire more workers and sell more output. In contrast, non-exporters did not respond to this change in their credit constraints. I infer that credit constraints must be binding for a significant fraction of exporters close to the 50 million rupee cutoff, while they are less important for similarly sized non-exporters. PSL eligibility did not differentially affect the export sales of exporting firms; instead, it caused exporters to expand their foreign and domestic sales symmetrically. I therefore conclude that my results are not driven by export sales *per se*. Rather, the kind of firm which chooses to export must be particularly likely to find credit constraints binding.

1.3 Model

I now develop a model of credit constraints and selection into exporting, with two objectives in mind. First, I aim to make a tight connection between the model and the natural experiment analyzed in Section 1.2, allowing me to exploit my empirical results to quantify the aggregate importance of credit constraints for exporters and non-exporters. Second, I will use the estimated model to explore the effect of different policies that target exporters in Section 1.5.

1.3.1 Environment

Time is discrete. Entrepreneurs are the key agents of the model, and begin each period with a state $\omega \equiv (z, f, k, a)$. z denotes productivity, which evolves exogenously over time according to

$$\log z' = \rho_z \log z + \sigma_z \epsilon, \quad \epsilon \sim \mathcal{N}(0, 1). \tag{1.1}$$

Entrepreneurs also differ in a shock to the fixed cost of exporting f, which is exogenous and time-invariant. They endogenously accumulate physical capital k and liquid assets a over time.

In each period the entrepreneur takes the wage w as given, demands labor ℓ and produces according to

$$y = z k^{\alpha} \ell^{1-\alpha}. \tag{1.2}$$

Entrepreneurs face a working capital constraint, in that they must pay for the labor they hire before production takes place. These payments may be made by directly using liquid assets a, or by borrowing using physical capital k as collateral. Formally, the entrepreneur faces the constraint

$$w\ell \leq a + b$$

where

$$0 \leq b \leq \lambda k.$$

b is the total borrowing of the entrepreneur, which is limited by the amount of physical capital they can offer as collateral. Borrowing is costly; if the entrepreneur borrows b, they incur interest $r_b b$.

The entrepreneur can potentially sell in two markets, domestic and foreign. Let y_d denote the amount the entrepreneur sells domestically and y_x the amount sold abroad. The entrepreneur chooses these quantities subject to

$$y_d + \tau y_x = y$$

where $\tau \geq 1$ is an iceberg trade cost. Each market is monopolistically competitive, and the entrepreneur sets prices subject to CES demand with elasticity σ,

$$p_d y_d = \left(\frac{p_d}{P}\right)^{1-\sigma} D, \quad p_x y_x = \left(\frac{p_x}{P^*}\right)^{1-\sigma} D^* \tag{1.3}$$

where p_d and p_x are the prices charged in each market. P is the domestic price level, D is total domestic demand and P^* and D^* are their foreign analogues.

In addition to the iceberg cost, exporters also incur a fixed cost F every period. I model this fixed cost as

$$F(\omega) = f z^{\vartheta} \tag{1.4}$$

where ω indexes the entrepreneur's state and f is a time-invariant shock to the entrepreneur's fixed cost. I assume

$$\log f \sim \mathcal{N}\left(\mu_f, \sigma_f^2\right). \tag{1.5}$$

Notice that (1.4) allows the cost of exporting to depend directly on productivity z. This relationship may be positive or negative depending on the sign of ϑ. If $\vartheta < 0$, for example, then more productive entrepreneurs face lower fixed costs. Intuitively, this might capture the idea that an entrepreneur who is skilled at producing goods is also skilled at overcoming the logistical and regulatory hurdles involved in selling internationally. As we will see in Subsection 1.3.5, ϑ is a key parameter in determining the drivers of selection into exporting.

1.3.2 Static Problem

First I focus on an entrepreneur who chooses to sell only domestically. Rearranging (1.3) and substituting, this entrepreneur's profit function may be written

$$\begin{aligned} \pi_d(\omega) = \max_{\ell, b} \quad & \tilde{p}_d \left(z k^{\alpha} \ell^{1-\alpha}\right)^{\frac{\sigma-1}{\sigma}} - w\ell - r_b b \\ \text{s.t} \quad & w\ell \leq a + b, \quad 0 \leq b \leq \lambda k, \end{aligned} \tag{1.6}$$

where

$$\tilde{p}_d = \left(P^{\frac{\sigma-1}{\sigma}} D^{\frac{1}{\sigma}}\right). \tag{1.7}$$

Notice that CES demand creates a source of decreasing returns to scale at the firm level, with returns to scale determined by $\left(\frac{\sigma-1}{\sigma}\right)$.

The profit maximization problem of an entrepreneur who chooses to export is similar. First, given total output y, the optimal choice of y_d and y_x satisfies

$$\frac{y_d}{y_x} = \tau^{\sigma} \left(\frac{P}{P^*}\right)^{\sigma-1} \left(\frac{D}{D^*}\right).$$

Given this allocation across markets, the exporter then solves a problem analogous to (1.6).

$$\pi_x(\omega) = \max_{\ell,b} \quad \tilde{p}_x \left(zk^{\alpha}\ell^{1-\alpha}\right)^{\frac{\sigma-1}{\sigma}} - w\ell - r_b b \tag{1.8}$$

$$\text{s.t} \quad w\ell \leq a + b, \quad 0 \leq b \leq \lambda k,$$

where

$$\tilde{p}_x = \left(\left(P^{\frac{\sigma-1}{\sigma}} D^{\frac{1}{\sigma}}\right)^{\sigma} + \tau^{-\sigma}\left(P^{*\frac{\sigma-1}{\sigma}} D^{*\frac{1}{\sigma}}\right)^{\sigma}\right)^{\frac{1}{\sigma}}. \tag{1.9}$$

Comparing (1.7) and (1.9), we can see that the only difference between exporting and non-exporting firms is that exporters effectively face a higher output price, i.e $\tilde{p}_x > \tilde{p}_d$. An entrepreneur exports if doing so is sufficiently profitable to justify paying the fixed cost, i.e if $\pi_x(\omega) - F(\omega) \geq \pi_d(\omega)$. Given this decision, the overall profits of an entrepreneur with state ω are

$$\Pi(\omega) = \max\{\pi_d(\omega), \pi_x(\omega) - F(\omega)\}.$$

1.3.3 Dynamic Problem

Given a state ω, the entrepreneur solves the static problem above to obtain profits $\Pi(\omega)$. The entrepreneur must then choose a', next period's stock of liquid assets, and i, investment in physical capital, to solve an infinite horizon dynamic programming problem. I assume that physical capital is subject to a fixed adjustment cost. Letting

$V(\omega)$ denote their value function, the entrepreneur solves

$$
\begin{aligned}
V(\omega) = \max_{a',i} \quad & \log(c) + \beta \mathbb{E}\left[V(\omega')|\omega\right], \qquad (1.10)\\
\text{where} \quad & \omega' = \left(z', f, k', a'\right), \\
& a' + i + c = \Pi(\omega)\left(1 - \phi \mathbb{1}\{i \neq 0\}\right) + (1 + r_a)a, \\
& a' \geq 0, \quad c \geq 0, \quad k' = (1 - \delta)k + i \geq 0.
\end{aligned}
$$

c is the entrepreneur's choice of consumption, β is their discount factor, δ is the rate of depreciation of physical capital, and r_a is the interest the entrepreneur earns on their savings of liquid assets. (1.10) defines the value function $V(\omega)$ and also the entrepreneur's policy function $g(\omega)$, which describes the optimal choice of a' and i, and therefore k', given an initial state ω. Three features of (1.10) are worth noting:

(i) The entrepreneur faces a fixed cost of adjusting their physical capital stock. In particular, I follow Cooper and Haltiwanger (2006) and model this cost as a fraction ϕ of profits.

(ii) The entrepreneur must choose k' before learning next period's productivity z'. This 'time to build' creates an additional source of friction in the entrepreneur's problem.

(iii) The entrepreneur is constrained to hold $a' \geq 0$, implying they cannot issue *intertemporal* debt to finance investment or consumption. Thus the only borrowing in my model is *intratemporal* debt, issued to fund the hiring of workers.

Together (i)-(iii) impede the entrepreneur's ability to adjust their capital stock and thus imply that conditional on capital there will be some dispersion in productivity z. This dispersion is important in relating the model to the natural experiment. There, by construction, eligible and ineligible firms had very similar capital stocks, but differed widely in their responses to changes in their credit constraints. In the model, dispersion in z conditional on k will be an important driver of these differing responses.

1.3.4 Aggregation and Equilibrium

So far, I have taken interest rates, wages, and prices as given and studied the decisions of individual entrepreneurs. This is all that is needed for the estimation in Section 2.1. However, when I study policy interventions in Section 1.5 it will be necessary to specify how markets clear and how prices are determined.

I assume that the model represents a small manufacturing sector within a larger economy. Entrepreneurs in this sector can purchase physical capital at a fixed price, normalized to 1. They can also save at a fixed interest rate r_a and borrow to hire labor at a fixed interest rate r_b. Aggregate expenditure on manufactured goods is fixed at an exogenous level D. The supply of labor to the manufacturing sector is fixed at L, and the wage w adjusts to clear this market. Finally, for simplicity, I assume that the foreign and domestic economies are symmetric, so that $P = P^*$, $D = D^*$ and so on.

The export fixed cost, $F(\omega)$, is paid using an 'entry good'. One unit of this good is produced using one unit of labor. Let L_e denote the total labor used in the production of the entry good, and L_p labor used in the production of goods. Then

$$L = L_e + L_p.$$

The assumptions above imply that there are two endogenous prices to be determined, the nominal wage w and the CES price index P. These prices adjust to satisfy the labor and goods market clearing conditions

$$\int \ell(\omega) d\mathcal{G}(\omega) = L_p, \tag{1.11}$$

$$\int R(\omega) d\mathcal{G}(\omega) = D \tag{1.12}$$

where $\ell(\omega)$ is the labor demand and $R(\omega)$ the revenue of a firm with state ω, and $\mathcal{G}$ is the joint distribution over states. Having stated these market clearing conditions, I now define a static equilibrium:

Definition 1 (Static Equilibrium)**.** Given a joint distribution $\mathcal{G}$ over states ω, a *static equilibrium* consists of nominal wages w and a CES price index P, such that when entrepreneurs solve the static problem in Subsection 1.3.2 the labor and goods market clearing conditions (1.11) and (1.12) are satisfied.

Any change in the environment will also change the joint distribution of states $\mathcal{G}$ in the long run. In my counterfactual experiments, I will focus on equilibria in which $\mathcal{G}$ has converged to its new steady state.

Definition 2 (Steady State)**.** A *steady state* is a distribution of states $\mathcal{G}_{ss}$, a policy function g, a nominal wage w and a CES price index P, such that (i) given G_{ss}, w and P are a static equilibrium, and (ii) when z evolve according to its exogenous law of motion (1.1) and k and a evolve according to the policy function g, the resulting joint distribution over states is $\mathcal{G}_{ss}$.

1.3.5 Exporting and Credit Constraints

The model developed above is rich but not analytically tractable. In this subsection, I make the following simplifying assumptions to illustrate the fundamental forces in the model:

(i) I abstract from physical capital and assume $\alpha = 0$. Then an entrepreneur's ability to hire labor is entirely determined by liquid assets a.

(ii) I assume that $\sigma_f = 0$, so that there is no exogenous heterogeneity in the fixed cost of exporting. Each entrepreneur then faces a cost $F(\omega) = \exp(\mu_f) z^{\vartheta}$.

(iii) I assume that $\rho_z = 0$, so that z is identically and independently distributed over time. This implies that the distribution of assets is independent of productivity. I further assume that the asset distribution is exogenous and constant.

First let us suppose $\vartheta = 0$. An entrepreneur exports if the extra profits from doing so exceed the fixed cost of exporting. Formally, an entrepreneur exports if

$$\Delta\pi(z, a) \geq \exp(\mu_f) \tag{1.13}$$

where $\Delta\pi(z, a) = \pi_x(z, a) - \pi(z, a)$ denotes the extra profits the entrepreneur earns by exporting. Integrating (1.13) over productivity z defines the probability that a firm exports conditional on assets a. Likewise, integrating over a defines the probability a firm exports conditional on z. The solid lines in Figure 1.3 plot these conditional probabilities.

The solid line in Panel (a) of Figure 1.3 shows that the probability a firm exports increases with z. More productive firms are larger, and this makes overcoming a given fixed cost more worthwhile. This is the usual driver of selection into exporting in models with heterogeneous firms and fixed costs (Melitz 2003).

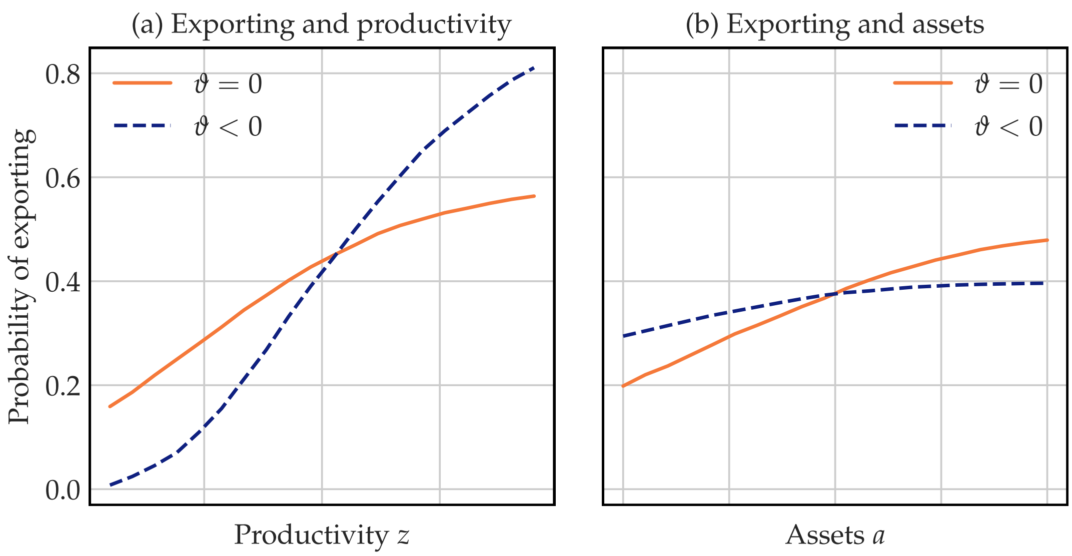
But Panel (b) shows that there is a second force at work in my model: the probability a firm exports is also an increasing function of its liquid assets a. A firm with low liquid assets cannot hire many workers and so produces at a small scale. Therefore such a firm doesn't find it worthwhile to start exporting. Moreover, even given its size, a firm with low a will not be able to expand when it does enter the export market, making $\Delta\pi(z, a)$ small for such a firm. Thus, the probability an entrepreneur exports increases in both z and a.

Importantly, z and a have opposing effects on the probability an entrepreneur is credit constrained. Holding a fixed, a higher productivity z raises labor demand and makes the firm more likely to hit a binding constraint. Holding z fixed, a higher

a relaxes this constraint. Therefore the fact that exporters are selected on both of these dimensions makes it theoretically ambiguous whether exporters are more or less likely than non-exporters to be constrained.[7] In particular, if the decision to export is driven mainly by an entrepreneur's assets, few exporters and many non-exporters will be constrained. An immediate implication is that exporters will be inefficiently large, and policies that encourage them to expand will worsen misallocation.

Now let us suppose $\vartheta < 0$, so that more productive entrepreneurs are better at paying the fixed cost of exporting; this will turn out to be the empirically relevant case in Section 2.1. The dashed lines in Figure 1.3 show how $\vartheta < 0$ changes the decision to export. As ϑ becomes negative, the probability a firm exports becomes

Figure 1.3: The Determinants of Exporting



Notes: Panel (a) shows the probability a firm exports as a function of productivity z, in the simple model (i.e imposing Assumptions (i) - (iii) above). The solid line assumes ϑ, the productivity-fixed cost elasticity, is 0. The dashed line assumes $\vartheta < 0$. Panel (b) shows the same probability as a function of liquid assets a.

[7] To make this concrete, consider the following (trivial) special case: suppose $\sigma_z = 0$, so that entrepreneurs differ only in their assets a. The decision to export then depends only on a, and, if the fixed cost μ_f is sufficiently high, it is possible to show that no constrained entrepreneur will ever choose to export, because they cannot expand enough to recoup the fixed cost. This special case is conceptually similar to one analyzed by Bai, Jin, and Lu (2019), who show that if selection into exporting is entirely driven by distortions then the gains from trade must be negative.

more sensitive to productivity z. In addition to the scale motive that was present when $\vartheta = 0$, now a higher productivity directly lowers the export fixed cost. This is captured by the relatively steep dashed line in Panel (a). As z becomes a more important driver of exporting, assets a necessarily become less important, as shown by the relatively shallow dashed line in Panel (b).

Figure 1.4 shows how this change in the drivers of the decision to export translates into a change in the characteristics of exporters and non-exporters. The solid line in Panel (a) plots the share of constrained firms among exporters, and the dashed line plots this share among non-exporters, as ϑ varies between -0.50 and 0.50. As ϑ rises, selection into exporting is less and less driven by productivity, so exporters become less likely to be constrained. Notice that when ϑ is sufficiently positive, constrained firms are more common among non-exporters. The following theorem makes the intuition above precise.

Theorem 1. *Suppose ϑ, the elasticity of fixed costs with respect to productivity, falls (i.e., becomes more negative). Suppose also that the average fixed cost μ_f varies so that the share of exporters remains constant. Then the share of constrained exporters rises, and the share of constrained non-exporters falls.*

See Appendix B.3 for a proof.

Panel (b) of Figure 1.4 relates these constraints to misallocation. Here I follow Hsieh and Klenow (2009) and use the marginal revenue product of labor ($MRPL$) as a measure of misallocation. $MRPL$ is defined as

$$MRPL \equiv \frac{dR}{d\ell}$$

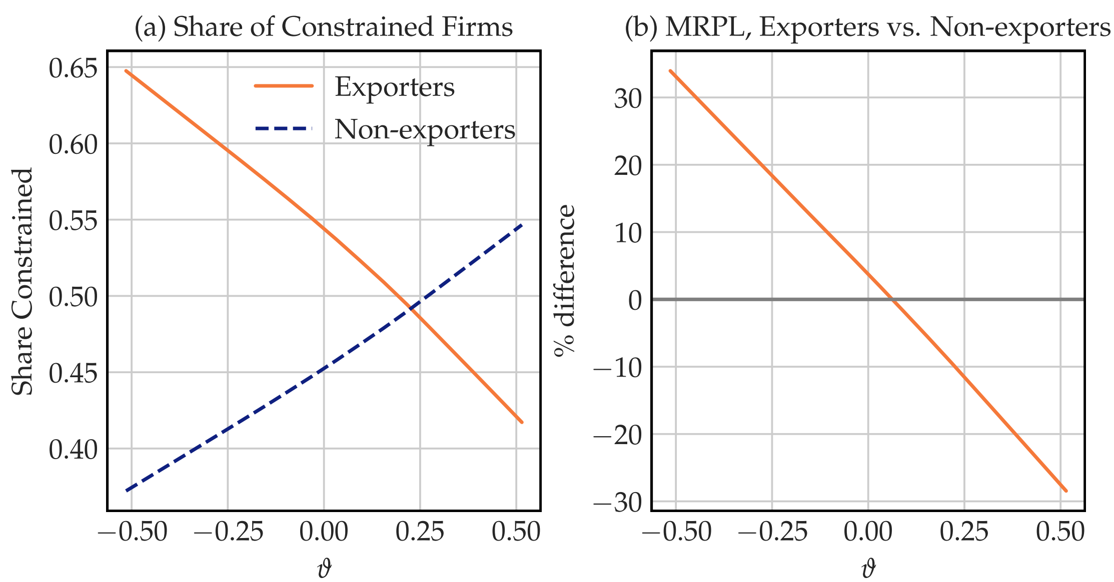
where R denotes a firm's revenue and ℓ its labor input. When $MRPL$ differs across firms, it is possible to raise aggregate output by reallocating labor towards high $MRPL$ firms. In my model, credit constraints create variation in $MRPL$. Unconstrained firms hire workers until $MRPL$ is equal to the wage w, while constrained firms have $MRPL > w$. Panel (b) shows that when ϑ is low, and many exporters are constrained, average exporter $MRPL$ is about 30% higher than non-exporter $MRPL$, i.e., exporters are inefficiently small. As ϑ rises, this gap changes sign and exporters become inefficiently large.

Summary

In this simple model, two forces — productivity z and assets a — shape an entrepreneur's decision to export. Which of these two forces dominates determines how many exporters are constrained relative to non-exporters, and in turn whether exporters are inefficiently large or small. The elasticity of fixed costs with respect to productivity, ϑ, governs the relative strength of these two forces.

The same two forces appear in the full model, i.e., without assumptions (i)-(iii)

Figure 1.4: Exporter Characteristics and ϑ



Notes: Panel (a) shows how the share of constrained firms among exporters (solid line) and non-exporters (dashed line) varies with ϑ, the productivity-fixed cost elasticity. Panel (b) shows the percentage difference in average marginal revenue products of labor between exporters as non-exporters. Consistent with Theorem 1, as ϑ varies I vary μ_f so that the share of exporters remains 0.43, its value in the Prowess dataset in 2007.

above. When z is not iid, productivity and assets will likely be positively correlated. As long as they are not perfectly correlated, however, both will play a role in the decision to export. Physical capital k, acting as collateral, behaves similarly to liquid assets by increasing an entrepreneur's ability to hire labor.

1.4 Estimation

I now estimate the parameters of the model developed in Section 2.2. I begin by externally calibrating several parameters. I follow Buera, Kaboski, and Shin (2021) and set the discount factor β equal to 0.85. I set the depreciation rate δ to 0.06 and the capital share α to 0.33. I set the elasticity of demand σ equal to 6.67, so that firm level returns to scale are $\left(\frac{\sigma-1}{\sigma}\right) = 0.85$ as in Midrigan and Xu (2014). I choose the iceberg trade cost τ so that in the model exporters sell 25% of their output abroad, the average in Prowess in 2007. Interest rates in the model are exogenous, so I follow Buera, Kaboski, and Shin (2021) and set the real interest rate on savings, r_a, to 0. I set the real interest rate on borrowing, r_b, to 5%, based on average (real) borrowing costs of firms in Prowess in 2007.[8]

Above I focused on a special case of the model with no dispersion in the export fixed cost shock (i.e., $\sigma_f = 0$) and showed that ϑ, the elasticity of fixed costs with respect to productivity, is a key parameter in determining the relative importance of credit constraints across exporters and non-exporters. The same intuition continues to apply once we allow $\sigma_f > 0$, but what matters now is the magnitude of ϑ relative to σ_f. For this reason, it is useful to note that at each point in time productivity z and the overall fixed cost of exporting F are log normally distributed with correlation

[8] The model also requires values for total demand D and for total employment L. I set these to 1 without loss of generality.

coefficient θ, where

$$\theta = \frac{\vartheta \sigma_z}{\sqrt{\vartheta^2 \sigma_z^2 + \sigma_f^2}}.$$

Note that θ is a monotonically increasing function of ϑ, and always has the same sign as ϑ. Since it is natural to think in terms of this correlation, I report my results in terms of θ rather than ϑ. Therefore, the parameters to be estimated are

(i) θ — the correlation between productivity and export fixed costs.

(ii) λ — the collateralizability of physical capital,

(iii) ρ_z and σ_z — the autocorrelation and standard deviation of the productivity process,

(iv) μ_f and σ_f — the mean and standard deviation of the export fixed cost shock,

(v) ϕ — the fixed cost of capital adjustment.

I estimate these parameters, plus a parameter that scales the size of the PSL policy (introduced below), by targeting 11 moments. These are the six treatment effects in Columns (3) and (4) of Table 1.2 and five descriptive statistics: the standard deviation of log sales growth, the autocorrelation of log sales, the fraction of firms that export, the average difference in log sales between exporters and non-exporters, and the frequency of investment 'spikes', defined as changes in firm-level capital stock above 20% in absolute value. These moments are summarized in Column (1) of Table 1.5. Note that the number of target moments exceeds the number of parameters to be estimated, i.e., the model is overidentified.

Mapping Model to Natural Experiment

The first part of my estimation strategy asks the model to match the effects of PSL in the data. I implement the PSL policy within the model by defining a cutoff

Table 1.5: Estimation

	(1) Data	(2) Model	(3) Model ($\theta = 0$)
(a) Moments			
Treatment Effects			
Non-exporter Loans	0.020	0.030	0.098
Non-exporter Employment	-0.000	0.027	0.083
Non-exporter Sales	0.110	0.015	0.049
Exporter Loans	0.328	0.335	0.215
Exporter Employment	0.246	0.284	0.181
Exporter Sales	0.224	0.162	0.109
Descriptive Statistics			
Standard deviation of log sales growth	0.526	0.526	0.523
Autocorrelation of log sales	0.971	0.970	0.971
Fraction of exporters	0.434	0.434	0.434
Log sales difference, exporters vs. domestic	2.403	2.410	2.410
Fraction of changes in capital above 20%	0.227	0.228	0.228
(b) Parameters			
Externally Calibrated			
β — Discount factor		0.850	0.850
δ — Deprecation rate		0.060	0.060
σ — Demand elasticity		6.667	6.667
α — Capital share		0.330	0.330
Estimated			
θ — Productivity - fixed cost correlation		-0.261	0.000
λ — Collateralizability		0.730	0.716
T — PSL scale		1.100	1.060
ρ_z — Productivity persistence		0.898	0.898
σ_z — Productivity standard deviation		0.055	0.055
μ_f — Export cost shock, mean		0.485	0.976
σ_f — Export cost shock, standard deviation		53.34	2.706
ϕ — Adjustment costs		0.230	0.228

Note: Column (1) of Panel (a) shows moments calculated from the Prowess dataset in 2007 — see Appendix A.3 for details. Columns (2) and (3) show the same moments generated by the estimated model, with (3) imposing the restriction $\theta = 0$. Panel (b) shows values of externally calibrated and estimated parameters.

c at the 35th percentile of the capital distribution.[9] Firms below the cutoff become more able to borrow, while those above experience no change. Formally I suppose the parameter λ, which determines the collateralizability of physical capital, now depends on k

$$\log \lambda(k) = \begin{cases} \log \lambda + T & \text{if} \quad k \leq c \\ \log \lambda & \text{if} \quad k > c \end{cases}$$

where $T \geq 0$ is a parameter to be estimated. As in Section 1.2 I measure the effects of this policy on log loans, employment and sales using a regression discontinuity design, and separate firms into exporters and non-exporters. Two points on timing are worth mentioning. First, in the model, I assume firms choose whether to export or not and then learn of the policy. This implies that the only immediate effect of the policy is to enable some firms to expand employment. Second, I measure the effects of the policy in the period in which it was implemented. This is different than in the data, where, motivated by the possibility that this policy might have been implemented with a lag, I estimated its effects over five years. To the extent that these estimates pick up long-run effects that differ significantly from short-run effects, my model's targets are off. However, in robustness checks in Appendix A.1 I show that the estimated effects (in the data) are not very sensitive to the choice of years, suggesting this is not a serious problem.

To understand what parameters the natural experiment identifies,[10] consider taking a first order approximation around $T = 0$. Let C be an indicator equal to 1 if a firm is constrained and let E be an indicator equal to 1 if a firm exports. The average

[9] I choose the 35th percentile because in 2007, 50 million rupees, the cutoff for PSL eligibility, was at this point in the plant and machinery distribution.

[10] Formally, all the targeted moments jointly identify all the parameters, so this discussion is purely heuristic.

effect of PSL eligibility on (log) borrowing is then

$$\beta^b_{E=1} = [\mathbb{P}(C = 1|E = 1, k = c)]\, T,$$
$$\beta^b_{E=0} = [\mathbb{P}(C = 1|E = 0, k = c)]\, T.$$

$\beta^b_{E=1}$ is the effect on exporters and $\beta^b_{E=0}$ on non-exporters. Note that the policy only changes borrowing for constrained firms, and so its effects are scaled by the shares of constrained firms $\mathbb{P}(C = 1|E = 1, k = c)$ and $\mathbb{P}(C = 1|E = 0, k = c)$. Notice also that these shares are conditional on the firm having physical capital k equal to c, the cutoff for eligibility.

Recall from Subsection 1.3.5 that θ, the correlation between the export fixed cost and productivity, plays a key role in determining the relative importance of constraints for exporters and non-exporters, i.e., in determining $\mathbb{P}(C = 1|E = 1, k = c)$ and $\mathbb{P}(C = 1|E = 0, k = c)$. When $\theta < 0$, for example, exporters are strongly selected on productivity and many are constrained. Therefore $\beta^b_{E=1}$ and $\beta^b_{E=0}$ help identify T, the overall size of the treatment, and θ, which controls the difference in treatment effects between exporters and non-exporters.

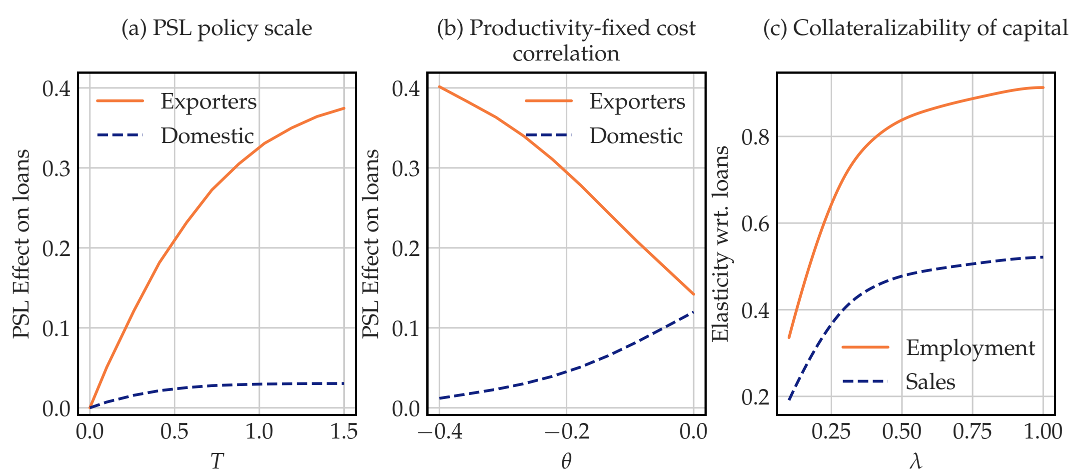
Panels (a) and (b) of Figure 1.5 illustrate this argument. Here I fix all the other parameters at the values estimated in Table 1.5. Panel (a) varies T between 0 and 1.50 and plots the resulting treatment effect on borrowing for exporters and non-exporters. Both treatment effects increase monotonically as the scale of the PSL policy rises. Notice, however, that both treatment effects are concave functions of T because as the credit constraint is relaxed, many firms become unconstrained and stop borrowing more. Panel (b) shows the same treatment effects varying θ between -0.40 and 0.0. The exporter treatment effect falls as θ rises, while the non-exporter treatment effect rises. Thus, the difference between these two treatment effects is informative about this parameter.

Turning to employment, and continuing to use a first order approximation, these treatment effects are

$$\beta^{\ell}_{E=1} = \left[\int \left(\frac{\lambda c}{\lambda c + a}\right) d\mathcal{G}_{ss}(a|C=1, E=1, k=c)\right] \beta^{b}_{E=1},$$
$$\beta^{\ell}_{E=0} = \left[\int \left(\frac{\lambda c}{\lambda c + a}\right) d\mathcal{G}_{ss}(a|C=1, E=0, k=c)\right] \beta^{b}_{E=0},$$

where $\mathcal{G}_{ss}$ is the steady state joint distribution of states. $\beta^{\ell}_{E=1}$ is the treatment effect on employment for exporters and $\beta^{\ell}_{E=0}$ is defined analogously. The key object here is $\left(\frac{\lambda c}{\lambda c+a}\right)$, which is the elasticity of employment with respect to borrowing for a constrained firm at the cutoff with assets equal to a. This elasticity differs across firms depending on a, hence the integrals above. When λ is large, firms rely primarily on borrowing, captured by λc, to meet their financing needs, and this elasticity will be large. When λ is small, they instead rely on retained earnings, captured by a, and this elasticity will be small. Therefore the size of the treatment effects on employment helps identify λ. Of course, this argument ignores the fact that changes in λ will

Figure 1.5: Parameters Identified by the Natural Experiment



Notes: Each plot is constructed by varying the parameter on the x-axis while holding all other parameters constant at their values in Column (2) of Table 1.5. The y-axis plots the moment(s) identifying each parameter. (a) shows the scale of the PSL policy T; (b) shows the productivity-fixed cost correlation θ; and (c) shows the collateralizability of physical capital λ.

also change the steady state distributions of capital and assets. The solid line in Panel (c) of Figure 1.5 plots the employment elasticity as a function of λ, allowing these distributions to vary. The employment elasticity is a monotonically increasing function of λ, in line with the simple intuition above.

Finally, the sales treatment effects are

$$\beta^s_{E=1} = \left[(1-\alpha)\left(\frac{\sigma-1}{\sigma}\right)\right]\beta^\ell_{E=1},$$
$$\beta^s_{E=0} = \left[(1-\alpha)\left(\frac{\sigma-1}{\sigma}\right)\right]\beta^\ell_{E=0}.$$

The sales treatment effects are mechanically related to the employment treatment effects — $(1-\alpha)\left(\frac{\sigma-1}{\sigma}\right)$ is just the elasticity of sales with respect to employment. Including the sales treatment effects therefore does not add anything to identification. Instead, they effectively provide an additional observation on the employment treatment effects and therefore help me estimate λ more precisely. The dashed line in Panel (c) of Figure 1.5 illustrates this point. The sales elasticity rises with λ and is proportional to the employment elasticity.

Additional Descriptive Statistics

The second part of my estimation strategy pins down the remaining parameters using the five descriptive statistics in Panel (a) of Table 1.5. I calculate these descriptive statistics using the Prowess dataset in 2007 — see Appendix A.3 for details. The parameters ρ_z and σ_z have a direct relationship with the persistence of log sales and the standard deviation of log sales growth. μ_f controls the level of the fixed cost of exporting and is closely related to the share of firms choosing to enter the export market. Given μ_f, σ_f then determines the intensity of selection into exporting. When σ_f is large, the decision to export is almost random, and the sales of exporters will be only slightly larger than those of non-exporters. When $\sigma_f = 0$, by contrast, the sales distributions of exporters and non-exporters will be entirely disjoint, and average sales among exporters will be much higher than among non-exporters. The difference in average sales between the two groups therefore identifies σ_f. Finally, the frequency of investment 'spikes' pins down the capital adjustment cost ϕ, because a large ϕ incentivizes high investment rates conditional on adjustment (Cooper and Haltiwanger 2006; Asker, Collard-Wexler, and Loecker 2014).

Estimation Procedure

I estimate the parameters in (i) - (v) above, plus the scale of the PSL policy T, by the Simulated Method of Moments. For ease of notation I collect the 8 parameters to be estimated into a vector Ψ. I calculate the 11 moments in Panel (a) of Table 1.5 in the data and collect them into a vector $\widehat{M}$. Given a guess on Ψ, I solve for steady state distribution of states $\mathcal{G}_{ss}$ from Definition 2 in Section 2.2. I draw a sample of firms from $\mathcal{G}_{ss}$ and calculate the model analogues of the moments in $\widehat{M}$, and collect these model moments into a vector $M(\Psi)$. I define a loss function

$$L(\Psi) = \left(\widehat{M} - M(\Psi)\right)' W \left(\widehat{M} - M(\Psi)\right)$$

where W is a weight matrix — I use a diagonal matrix which weights each moment by the inverse of the square of its standard error. Finally I choose Ψ to minimize $L(\Psi)$. For more details, please see Appendix A.3.

1.4.1 Results

Parameter Estimates

Column (2) of Table 1.5 shows my main results. Panel (a) shows how the model fits the target moments and (b) shows the estimated parameters. I start by discussing the estimated parameters. The values of σ_z, ρ_z, and ϕ are similar to existing estimates

(Cooper and Haltiwanger 2006; Asker, Collard-Wexler, and Loecker 2014).

I estimate that $\theta < 0$, implying that more productive entrepreneurs typically face lower export fixed costs. I also estimate a large σ_f, implying that the fixed cost of exporting f is very dispersed. This is a result of finding that selection into exporting is strongly driven by productivity, i.e., $\theta < 0$, while still matching the difference in average sales between exporters and non-exporters (the fourth moment in Table 1.5). A large σ_f implies that conditional on productivity z, exporters are almost randomly selected, which keeps the difference in average sales between exporters and non-exporters in the model consistent with its value in the data.

I estimate $\lambda = 0.73$, implying entrepreneurs can collateralize a large fraction of their physical capital. As Panel (c) of Figure 1.5 shows, this is a consequence of the relatively large employment and sales treatment effects I estimated for exporters. Finally, I estimate that $T = 1.10$, implying that PSL had a large effect on the borrowing constraints of eligible firms. Note that T is much larger than the effect of PSL eligibility on borrowing, even among exporters, because the constraint is only binding for a fraction of firms and because the treatment effect on borrowing is a concave function of T, as shown in Panel (a) of Figure 1.5.

Targeted Moments

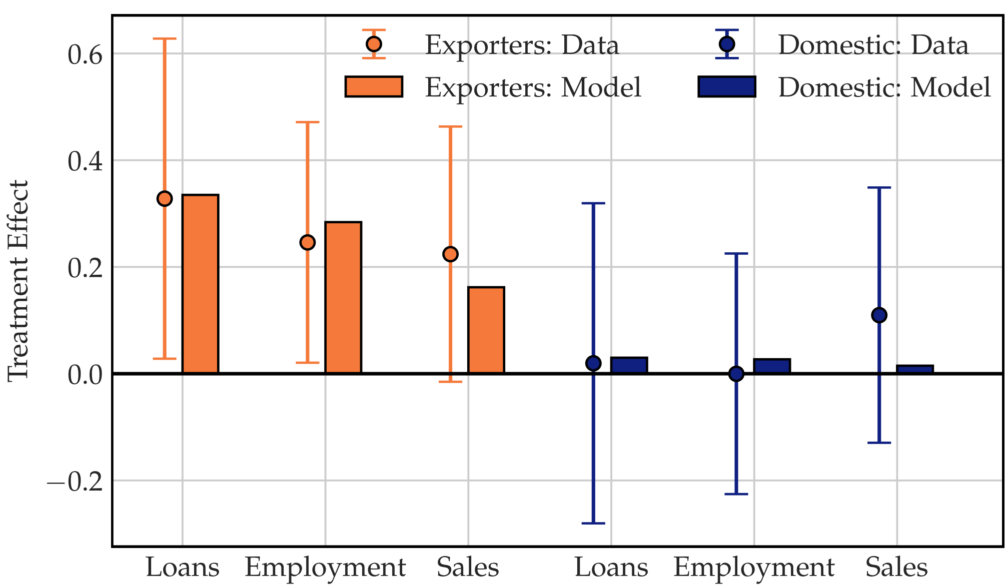
Panel (a) shows how successful the model is in matching the target moments. It fits the five descriptive statistics (almost) perfectly. Figure 1.6 therefore focuses on the model's ability to replicate the six treatment effects in Column (1) of Panel (a). The orange dots show the exporter treatment effects from the data, with 95% confidence intervals, while the blue dots show non-exporter treatment effects. The solid bars show the corresponding treatment effects implied by the estimated model. The model is quantitatively successful in capturing the pattern observed in Section 1.2: significant effects of PSL eligibility for exporters and negligible effects for non-exporters.

In Column (3) of Table 1.5, I explore the role of the productivity-fixed cost correlation θ by re-estimating the model imposing $\theta = 0$. While this model still fits the five descriptive statistics well, it is less successful in matching the pattern of treatment effects across exporters and non-exporters. In particular, it struggles to generate a large difference between exporters and non-exporters. For example, compared to Column (2), the difference between the exporter and non-exporter treatment effect on loans falls from 30.5 percentage points to 11.7 percentage points. I conclude that allowing for a rich pattern of selection into exporting by incorporating this correlation is crucial for enabling the model to match the estimated treatment effects.

Untargeted Moments

In addition to fitting the target moments well, the model also matches a number of untargeted moments, shown in Table 1.6. First, recall from Section 1.2 that in the

Figure 1.6: Treatment Effects: Model vs Data



Notes: 'Data' treatment effects from Columns (3) and (4) of Table 1.2. Error bars show 95% confidence intervals. 'Model' shows the treatment effects produced by the model at the parameters in Column (2) of Table 1.5.

data PSL eligibility had a negligible effect on the extensive margin of exporting, i.e., the probability a firm exports. The first row of Table 1.6 reproduces this figure and compares it to the same object in the model. Consistent with the data, in the model PSL eligibility has a tiny effect on the extensive margin of exporting — it raises the probability a firm exports by 0.00036. This is a natural consequence of the finding that the decision to export is largely driven by productivity. In Figure 1.3, we saw that when $\vartheta < 0$, so that productive entrepreneurs are more able to export, exporting is very insensitive to changes in an entrepreneur's ability to borrow.[11]

The next two moments relate to the the dynamics of exporting. In the data exporting is very persistent: 93% of firms that export in a given year continue to do so in the following year. The model generates a similar degree of persistence, with just over 90% of exporters in a given year exporting in the following year. In the data export entry is also accompanied by fast sales growth. To measure this, I regress log sales growth on an indicator equal to 1 if a firm starts exporting in that year. The third row of Table 1.6 reports results from the data and from the model. The model replicates the pattern of fast sales growth upon entry found in the data, but in fact produces too much. This is likely because, in the model, the fixed cost shock f is time-invariant and so moves in and out of exporting are largely driven by shocks to productivity z, which is highly correlated with sales.

The final two rows report differences in input demands across exporters and non-exporters. In each case I regress log input demands (i.e., employment and physical capital) on industry fixed effects and a dummy for export status. The results in Column (1) indicate that exporters in the data hire more labor and use more capital than non-exporters, and that the gap is smaller for capital. Column (2) shows that the model is qualitatively consistent with these facts, and also gets the relative size

[11] Note that in the simple model of Section 2.2, 'ability to borrow' is really a just an entrepreneur's liquid assets. But increasing an entrepreneur's assets has very similar effects to increasing its ability to collateralize physical capital, i.e to the PSL policy.

of the differences in capital and labor right. Relative to the data, however, exporters in the model use too much labor and too little capital.

1.4.2 Implications for Misallocation

The estimated model implies that credit constraints are concentrated among exporters. Among firms close to the PSL cutoff in the model 45% of exporters are at a binding credit constraint, compared to only 8% of non-exporters. Panel (a) of Figure 1.7 shows how these numbers vary across the capital distribution. Among both exporters and non-exporters, the share of constrained firms falls as capital rises. At every point in the capital distribution, there is a large difference between exporters and non-exporters. Overall 37% of exporters and 8% of non-exporters are constrained.

An immediate implication is that inputs are misallocated across exporters and non-exporters. To measure this misallocation in the model and relate it to the data, I use the marginal revenue products of labor and capital. Letting R denote a firm's revenue and ℓ and k its labor and capital inputs, these are defined as

$$MRPL \equiv \frac{dR}{d\ell} = (1-\alpha)\left(\frac{\sigma-1}{\sigma}\right)\frac{R}{\ell}, \quad MRPK \equiv \frac{dR}{dk} = \alpha\left(\frac{\sigma-1}{\sigma}\right)\frac{R}{k}.$$

Table 1.6: Untargeted Moments

Moment	(1) *Data*	(2) *Model*
Extensive margin effect of PSL	-0.021	0.000
Persistence of export status	0.933	0.905
Sales growth of new exporters	0.402	0.794
Log employment difference, exporters vs. non-exporters	2.066	2.309
Log capital difference, exporters vs. non-exporters	1.820	1.481

Note: 'Data' moments calculated using manufacturing firms in Prowess, 2006-07. See Appendix A.3 for details. 'Model' moments calculated using a panel of simulated firms from the estimated model, parameterized following Column (2) of Table 1.5.

where the second equality is a consequence of the Cobb-Douglas and CES demand assumptions. These objects are a natural measure of misallocation. If one firm has a higher $MRPL$ than another, reallocating labor towards that firm would raise aggregate revenue. Similarly, if $MRPL$ is typically high among exporters, then reallocating labor towards exporters will raise aggregate revenue.

I begin by asking how $MRPL$ and $MRPK$ differ between exporters and non-exporters in the model. Formally I simulate data from the estimated model and run the regressions

$$\log\left(\frac{R_i}{\ell_i}\right) = \tilde{\alpha}_\ell + \varphi_\ell e_i + \epsilon_i \tag{1.14}$$

$$\log\left(\frac{R_i}{k_i}\right) = \tilde{\alpha}_k + \varphi_k e_i + \eta_i \tag{1.15}$$

where e_i is a dummy equal to 1 if firm i exports. Note that $\left((1-\alpha)\left(\frac{\sigma-1}{\sigma}\right)\right)$ and $\left(\alpha\left(\frac{\sigma-1}{\sigma}\right)\right)$ get absorbed by the additive constants $\tilde{\alpha}_\ell$ and $\tilde{\alpha}_k$. The coefficients φ_ℓ and φ_k capture differences in average $MRPL$ and $MRPK$ between exporters and non-exporters. The results are reported in Column (1) of Table 1.7, and indicate that exporter $MRPL$ is 9.4% and higher and exporter $MRPK$ is roughly 93% higher.

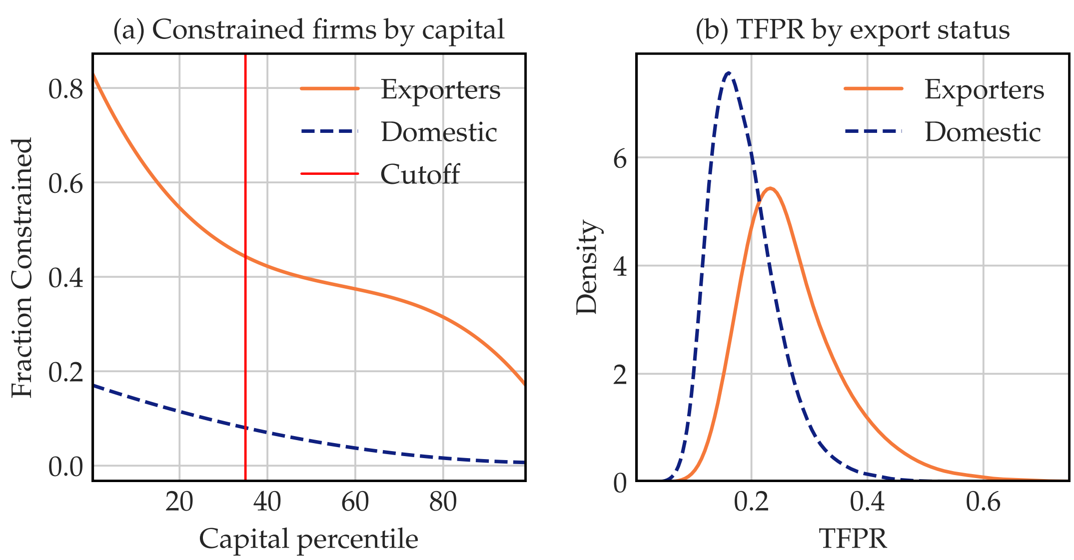
I also report results for revenue total factor productivity ($TFPR$), which aggregates $MRPK$ and $MRPL$ to give an overall measure of misallocation and is defined by

$$TFPR = MRPK^{\alpha} MRPL^{1-\alpha}. \tag{1.16}$$

The final row of Column (1) shows that exporter $TFPR$ is about 37% higher than non-exporter $TFPR$, and hence exporters are on average inefficiently small. Panel (b) of Figure 1.7 plots the distributions of $TFPR$ for both exporters and non-exporters. The distribution for exporters is clearly shifted to the right, but it is also true that there is substantial heterogeneity within each of these sets of firms. As we will see in Section 1.5, both dimensions of misallocation — between exporters and non-exporters, and within each group — will play an important role in determining the effects of any policy which targets exporters.

Given the Cobb-Douglas production and CES demand assumptions, I can also calculate $MRPL$ and $MRPK$ directly from the Prowess dataset and repeat these regressions.[12] Note an advantage of the log specification I use here is that variation in the capital and labor shares across industries can be absorbed by including industry fixed effects. I report the results in Column (2) of Table 1.7. Qualitatively the results are similar to those from the estimated model: exporter marginal revenue products are higher, and the gap is particularly large for capital. However, relative to the model, the data shows a larger gap in $MRPL$ and a substantially smaller one in

Figure 1.7: Exporting, Credit Constraints and Misallocation



Notes: Plots based on simulated data from estimated model, parameterized as in Column (2) of Table 1.5. Panel (a) plots the fraction of constrained firms in each percentile of the capital distribution; the dashed line shows domestic firms and the solid line shows exporters. Panel (b) plots the densities of revenue total factor productivity (TFPR) for exporters (solid) and non-exporters (dashed).

[12] Given the log specification, in principle these MRPL and MRPK numbers could have been directly inferred from the log differences in sales and inputs between exporters and non-exporters, i.e., the fourth moment in Table 1.5 and the fourth and fifth moments in Table 1.6. In practice these two procedures give slightly different results, because employment or capital data are missing for some firms.

Table 1.7: Marginal Revenue Products: Exporters vs. Non-exporters

	(1) Model	(2) Data
Labor (MRPL)	0.094	0.162***
	-	(0.030)
Capital (MRPK)	0.930	0.515***
	-	(0.038)
Total (TFPR)	0.370	0.278***
	-	(0.030)

Note: Column (1) reports the (log) difference in marginal revenue products between exporters and non-exporters in the model, estimated using the regressions (1.14) and (1.15). The final row aggregates these with weights $(1-\alpha)$ and α to form an estimate of the the difference in log TFPR between exporters and non-exporters. Column (2) reports the same difference using data from manufacturing firms in the Prowess dataset in 2007. Note that the regressions in (2) include industry fixed effects. The final row of (2) is produced in the same way as the final row of (1).

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

$MRPK$. Aggregating as in (1.16), in the data exporter $TFPR$ is 28% higher than non-exporter $TFPR$.

1.5 Does Targeting Exporters Raise Productivity?

The key findings from the estimation in Section 2.1 are: (i), credit constraints are binding for many exporters but few non-exporters; and (ii), exporters are inefficiently small, and reallocating labor and capital towards them would raise aggregate productivity. These findings have the potential to provide an efficiency rationale for the many real-world policies which target exporting firms. For example, surveying the literature on East Asia's 'miracle economies', Itskhoki and Moll (2019) point out that policies which subsidized the input purchases of exporters or provided them with favorable access to credit were widespread. As I show below, my results also imply a source of gains, or losses, from trade absent from models in which inputs are allocated efficiently.

Subsection 1.5.1 defines the outcome of interest — total factor productivity (TFP) — and shows theoretically how changes in the allocation of input across firms may affect it. Subsection 1.5.2 studies the effects of two specific policies on TFP. The first policy directly relaxes the credit constraint of exporting firms, while the second subsidizes their employment. These two policies cause comparable amounts of reallocation towards exporters. Surprisingly, however, I show that they have very different consequences for aggregate productivity. Finally, Subsection 1.5.3 uses the estimated model to quantify the effect of reductions in trade costs on TFP. I show that any gains from reallocation are modest, and contrast this finding with the results of a model in which misallocation is the result of exogenous wedges in input markets.

1.5.1 Reallocation and Total Factor Productivity

First, some notation. Let K denote total capital and L_p total labor used in production (recall the export fixed cost absorbs some labor). Recall from Section 2.2 that $\omega \equiv (z, f, k, a)$ indexes an entrepreneur's state and that $\mathcal{G}$ denotes the distribution over those states. Let Ω_x denote the set of states ω in which an entrepreneur chooses to export and let Ω_d be the states in which they choose to produce only for the domestic market. Let

$$S_{\ell x} = \left(\int_{\Omega_x} \ell d\mathcal{G} \right) L_p^{-1}, \quad S_{kx} = \left(\int_{\Omega_x} k d\mathcal{G} \right) K^{-1}$$

denote the shares of labor and capital held by exporters, and define $S_{\ell d}$ and S_{kd} analogously. Finally, let

$$s_\ell(\omega) = \begin{cases} \left(\frac{\ell(\omega)}{L_p}\right) S_{\ell x}^{-1} & \text{if} \quad \omega \in \Omega_x \\ \left(\frac{\ell(\omega)}{L_p}\right) S_{\ell d}^{-1} & \text{if} \quad \omega \in \Omega_d \end{cases}$$

be the labor demand of an entrepreneur with state ω, relative to total labor demand of the group — exporters or non-exporters — to which ω belongs. Define $s_k(\omega)$ analogously.

Now we are in a position to define TFP, denoted by Z. Aggregating over firms, total real output in this economy can be written

$$Y = ZK^{\alpha}L_p^{1-\alpha}.$$

Because inputs are not necessarily allocated efficiently, Z depends on how inputs are allocated across firms. Since my focus is on reallocation between exporters and non-exporters, a helpful way of writing Z decomposes it into the allocation of inputs *between* these two sets of firms, and the allocation of inputs *within* each of these sets of firms. Formally

$$Z = \left(\left(S_{kd}^{\alpha} S_{\ell d}^{1-\alpha}\right)^{\frac{\sigma-1}{\sigma}} Z_d^{\frac{\sigma-1}{\sigma}} + \left(1+\tau^{1-\sigma}\right)^{\frac{1}{\sigma}} \left(S_{kx}^{\alpha} S_{\ell x}^{1-\alpha}\right)^{\frac{\sigma-1}{\sigma}} Z_x^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{1.17}$$

where

$$Z_d = \left(\int_{\Omega_d} \left(z s_k^{\alpha} s_\ell^{1-\alpha}\right)^{\frac{\sigma-1}{\sigma}} d\mathcal{G} \right)^{\frac{\sigma}{\sigma-1}}, \quad Z_x = \left(\int_{\Omega_x} \left(z s_k^{\alpha} s_\ell^{1-\alpha}\right)^{\frac{\sigma-1}{\sigma}} d\mathcal{G} \right)^{\frac{\sigma}{\sigma-1}}. \tag{1.18}$$

We can think of Z_x and Z_d as the productivity of a representative exporter and non-exporter, respectively.

Taking a first order approximation, any change in the allocation of capital and labor can also be decomposed into changes in between-group misallocation, and changes in within-group misallocation[13]

$$d \log Z = \underbrace{\left(\frac{\sigma}{\sigma-1}\right)\left(dS_{kx}\left(\overline{MRPK}_x - \overline{MRPK}_d\right) + dS_{\ell x}\left(\overline{MRPL}_x - \overline{MRPL}_d\right)\right)}_{Between} \tag{1.19}$$
$$+ \underbrace{(\rho_d d \log Z_d + \rho_x d \log Z_x)}_{Within},$$

where ρ_x is the share of exporters in total revenue and ρ_d the share of non-exporters. $\overline{MRPK}_x$ is the average (input-weighted) marginal revenue product of capital among exporters and $\overline{MRPK}_x$, $\overline{MRPL}_x$, and $\overline{MRPL}_d$ are defined analogously.

This expression shows the basic logic underlying any policy directed towards exporters. Exporters have high marginal revenue products relative to non-exporters, i.e., $\overline{MRPK}_x > \overline{MRPK}_d$ and $\overline{MRPL}_x > \overline{MRPL}_d$. Therefore reallocating inputs from non-exporters towards exporters should raise TFP. But it also highlights that this argument comes with an important caveat: such reallocation is beneficial provided it does not simultaneously worsen misallocation within the sets of exporting or non-exporting firms, as captured by $d \log Z_x$ and $d \log Z_d$.

1.5.2 Policy Interventions

Relaxing the Credit Constraint of Exporters

I suppose the policymaker can relax the credit constraint of exporting firms. Formally, the parameter λ now depends on a firm's export status and is denoted by λ_d for non-exporters and λ_x for exporters. I assume $\lambda_d = 0.73$, the value estimated in

[13]In this decomposition I hold the sets of exporting and non-exporting firms Ω_x and Ω_d fixed. Such extensive margin changes turn out to be quantitatively negligible for the counterfactuals I consider.

Section 2.1, but that for exporters it rises according to

$$\log \lambda_x = \log \lambda_d + \Delta_x.$$

Lacking a natural magnitude for Δ_x, I choose the size of the PSL policy estimated in Section 2.1 and set $\Delta_x = 1.10$. Starting from the steady state implied by the parameters estimated in Section 2.1, I shock the economy with this change in λ. I solve for the new equilibrium in the short run, holding the joint distribution of states (productivity, export fixed costs, capital, and liquid assets) constant; and in the long run, allowing this distribution to converge to its new steady state.

In the short run the share of employment in exporting firms rises by 2.30%, and in the long run this rises to 2.68%. Exporters also increase their share of the capital stock by 0.85%. Column (1) of Table 1.8 shows the effects on TFP. In keeping with (1.19), I decompose the overall effect into two components. First, I calculate a 'Between' component by holding the allocation of inputs within each set of firms constant and allowing the aggregate shares S_{kx}, etc., to vary. Second, I hold the

Table 1.8: Decomposing Policy Effects on TFP

	(1) Credit Policy		(2) Employment Subsidy	
	Short Run	*Long Run*	*Short Run*	*Long Run*
Between	0.09	0.07	0.09	0.24
Within	2.36	3.20	-0.04	-0.37
Overall	2.47	3.33	0.05	-0.12

Note: Each row shows the percentage increase in TFP caused by each component for each policy. 'Between' shows the effect of each policy, holding the allocation of inputs within the sets of exporting and non-exporting firms constant but allowing the aggregate shares of exporters and non-exporters to vary. 'Within' holds the aggregate shares constant but allows the allocation within the sets of exporting and non-exporting firms to vary. 'Overall' allows both to vary. Note that 'Within' and 'Between' need not sum to exactly 'Overall' because of second order terms.

aggregate shares fixed and calculate a 'Within' component by allowing the input share of each firm within the set of exporters and non-exporters to vary. Overall, TFP rises by 2.47% in the short run and by 3.33% in the long run. Both components make a positive contribution at every time horizon, but the largest source of gains is reduced misallocation within each set of firms. Constrained exporters drive this positive effect. As the credit constraint is relaxed, these firms, which have relatively high marginal products, expand and pull labor out of less productive firms.

Subsidizing Exporter Employment

The experiment above shows that relaxing the credit constraint of exporters can yield substantial TFP gains. It is not obvious, however, that this is a margin a policymaker can manipulate. I now ask whether a policymaker can achieve similar results using simpler instruments. In particular, I suppose a policymaker observes the high average marginal revenue product of labor among exporters and reasons that this misallocation could be resolved by an employment subsidy directed towards exporters. If workers receive a wage w, the wages facing exporters and non-exporters are

$$w_x = w(1 + t - t_x), \quad w_d = w(1 + t).$$

where t is a tax chosen so that the government budget balances.[14] I set the subsidy $t_x = 0.083$, chosen so that in the short run the percentage increase in exporter employment caused by this policy matches the short run effect of the credit policy above. Thus the two policies cause comparable amounts of reallocation between exporters and non-exporters.

The results of the employment subsidy policy are shown in Column (2) of Table 1.8. Again I show effects in both the short and long run and decompose these into

[14] Note that because labor is supplied inelastically, the symmetric tax t is not distortionary. All the effects of this policy are therefore the result of the fact that it targets exporters.

between and within components. In the short run, the employment subsidy causes a very small increase in TFP. In the long run, however, its effects change sign and TFP falls by 0.12%. Looking at the decomposition, we can see that the subsidy and credit policies have similar effects on reallocation between exporters and non-exporters (by construction). But the two policies have dramatically different effects on misallocation within each set of firms.

Understanding the Two Policies

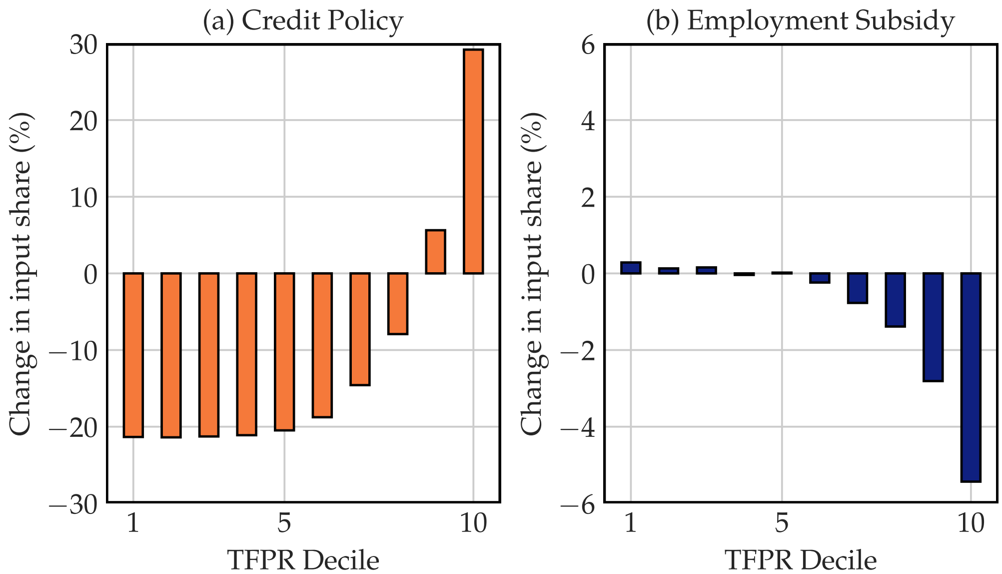
Figure 1.8 illustrates the difference between these two policies, focusing on exporters. I summarize the extent of misallocation in each firm using revenue total factor productivity (TFPR), as defined in (1.16). For each policy and each firm, I calculate the change in its share of capital, $\Delta \log s_k$, and the change in its share of labor, $\Delta \log s_\ell$. I aggregate these to form $\Delta \log s$, defined by

$$\Delta \log s = \alpha \Delta \log s_k + (1-\alpha) \Delta \log s_\ell .$$

Finally, I plot averages of $\Delta \log s$ by TFPR deciles for each policy. Figure 1.8 shows the results; a positive number indicates that firms in that TFPR bin on average grew as a result of a particular policy. Panel (a) shows the credit policy. The primary beneficiaries are constrained exporters with high TFPR. Since these firms were initially inefficiently small, aggregate TFP rises.

Panel (b) shows the employment subsidy. This policy causes less reallocation within exporters than the credit policy; notice that the scale in Panel (b) is smaller. More importantly, the reallocation that this policy does cause worsens misallocation. To see why this should be the case, compare the elasticity of labor demand with respect to the wage subsidy between constrained and unconstrained firms. For a

Figure 1.8: Reallocation Among Exporters Under Two Different Policies



Notes: Each plot bins exporters by TFPR and shows the average change in inputs within each bin. Panel (a) shows the effect of relaxing the exporter credit constraint; this policy has the largest effect on the exporter's with the highest TFPR. Panel (b) shows the effect of the exporter employment subsidy; this policy caused the highest TFPR exporters to contract relative to low TFPR exporters.

constrained firm this elasticity is

$$\frac{d \log \ell}{d \log t_x} = 1,$$

but for an unconstrained firm this elasticity is

$$\frac{d \log \ell}{d \log t_x} = \left(\frac{1}{1 - (1-\alpha)\left(\frac{\sigma-1}{\sigma}\right)} \right) \simeq 2.5.$$

Thus the subsidy causes unconstrained exporters to expand much faster than constrained ones. Since constrained firms typically have high marginal products, this explains the pattern in Panel (b) of Figure 1.8. Since low TFPR exporters expand at the expense of high TFPR exporters, aggregate TFP falls.

These results suggest that subsidies are not effective instruments in resolving the misallocation created by credit constraints. Even though in my counterfactual the subsidy was targeted to a group with a high share of constrained firms (i.e exporters), it backfired because it primarily benefitted unconstrained firms with low marginal products. The basic logic of this point goes beyond this specific example: subsidies are effective if the targeted firms expand in response, but almost by definition, constrained firms are unable to do this. Therefore subsidy policies will typically cause exactly the wrong firms to expand and worsen misallocation. On the other hand, an intervention that directly addresses the source of misallocation can yield a large improvement in TFP.

1.5.3 Trade, Reallocation and Credit Constraints

Since Melitz (2003) reallocation across firms has been central to accounts of the effects of trade liberalization, and two recent papers (Berthou et al. 2019; Bai, Jin, and Lu 2019) ask how this reallocation might affect aggregate productivity when inputs are misallocated across firms. I revisit this question using my model, in which misallocation is the result of specific distortions, i.e., credit constraints, as well as adjustment costs in physical capital. I start by extending the decomposition (1.19) to incorporate changes in the iceberg trade cost τ

$$d\log Z = \underbrace{-(1-x_d)d\log\tau}_{Direct} + \underbrace{(\rho_d d\log Z_d + \rho_x d\log Z_x)}_{Within} \\ + \underbrace{\left(\frac{\sigma}{\sigma-1}\right)\left(dS_{kx}\left(\overline{MRPK}_x - \overline{MRPK}_d\right) + dS_{\ell x}\left(\overline{MRPL}_x - \overline{MRPL}_d\right)\right)}_{Between} \tag{1.20}$$

where x_d is the share of expenditure devoted to domestically produced goods. The 'Direct' term here reflects the effect of falling trade costs holding the allocation of inputs across firms constant. Atkeson and Burstein (2010) show that in a large class of models this term captures all the gains from trade, but in my model this is not the case. Instead, the reallocation of inputs caused by falling trade costs may have a first order effect on productivity. These gains, or losses, from reallocation are captured by the final two terms in (1.20).

I shock the model with an exogenous decrease in the iceberg trade cost. I choose $d\log\tau = -0.10$, so that trade costs fall by roughly 10%.[15] I solve for the new equilibrium in both the short run and the long run, and decompose the change in TFP following (1.20). The results are shown in Column (1) of Table 1.9. Overall TFP rises by 0.428% in the long run, but reallocation contributes very little to this figure for two reasons. First, the reduction in τ causes very little reallocation between exporters and non-exporters, as can be seen from the relatively small magnitude of the 'Between' component in Table 1.9. Second, any gains from reallocation between these two groups are largely offset by worsening misallocation within each group. This is because lower trade costs are in this respect similar to the exporter employment subsidy studied in Subsection 1.5.2. The exporters most able to expand in response to falling trade costs are the relatively unproductive, unconstrained ones, who then drag aggregate TFP down.

These results suggest that merely knowing that in the cross-section exporters have relatively high TFPR is insufficient to conclude that trade will raise aggregate productivity via reallocation. To make this point concrete, I repeat this exercise in a model in which misallocation is the result of exogenous wedges in input markets, as in Hsieh and Klenow (2009) — similar models have been used by Berthou et al. (2019) and Bai, Jin, and Lu (2019) to study the gains from trade under misallocation. Formally, I construct a second economy in which labor and capital can be hired freely (i.e., without credit constraints, adjustment costs etc.), but the prices of these factors vary across firms depending on exogenous wedges. I choose distributions for these wedges so that this model exactly replicates the joint distribution of sales, employment, capital and export status from the original model. The two models thus

[15] I choose a relatively small shock so that the first order approximation in (1.20) is reasonably accurate.

have identical cross-sectional implications, and in particular yield identical differences in marginal revenue products between exporters and non-exporters. For complete details see Appendix A.4.

Column (2) of Table 1.9 shows the effect of a 10% reduction in trade costs in this second model. Three things are worth nothing. First, by construction the 'Direct' term is in identical in both models, since this does depend on how firms respond to changes in their environment. Second, much more reallocation occurs in the model with exogenous misallocation — the 'Between' gains in this model are about seven times larger than those from the original model. This is a natural consequence of removing the frictions in the original model — there, inputs were misallocated precisely because reallocation was difficult. Third, falling trade costs do not have any effect on 'Within' misallocation in this model. Conditional on export status, all firms respond symmetrically to falling trade costs, leaving the distributions of inputs within the sets of exporters and non-exporters unchanged. The net result is that the model with exogenous misallocation implies large gains from reallocation; these are roughly half

Table 1.9: Trade Cost Reductions and TFP

	(1) *Full Model*		(2) *Model with Exogenous Misallocation*
	Short Run	*Long Run*	
Direct	0.411	0.411	0.411
Between	0.004	0.029	0.200
Within	-0.003	-0.016	0.000
Overall	0.412	0.428	0.612

Note: Column (1) shows the effect of a 10% reduction in iceberg trade costs in the estimated model, in the short run and in the long run. Column (2) shows the effects of the same reduction in a model in which misallocation is the result of exogenous wedges — see Appendix A.4 for details. Each column decomposes the effect on TFP into three components following (1.20). Note that the 'Overall' effect is not exactly the sum of these three components, because of second order terms absent from (1.20).

as large as the direct effect of falling trade costs.

Overall, the results in Table 1.9 show that a model in which misallocation is the result of exogenous distortions is a poor guide to the effects of falling trade costs. Instead, accurately measuring how policy changes interact with misallocation hinges on explicitly modeling the source of misallocation.

1.6 Conclusion

This paper's main findings can be summarized in three points. First, I used a natural experiment in India to show that exporters responded strongly to an exogenous increase in credit supply, while non-exporters did not. Second, I estimated a model of credit constraints and exporting by targeting the results of this natural experiment. I found that, in an environment with several dimensions of heterogeneity, productivity was the key driver of the decision to export. Since more productive firms are more likely to find credit constraints binding, the result is that credit constraints bind for many exporters; overall, 37% of exporters in the model are constrained, compared to only 8% of non-exporters. Finally, in counterfactual experiments I showed that different policy interventions that target exporters have very different effects on misallocation. Relaxing the credit constraint facing exporting firms significantly raises aggregate TFP because it allows the most productive exporters to expand. Simply subsidizing exporter employment has the opposite effect because the primary beneficiaries are unconstrained and relatively unproductive exporters. The same logic implies that credit constraints limit the productivity gains from reallocation driven by falling trade costs.

Chapter 2

Housing Demand, Inequality, and Spatial Sorting [1]

The skill premium has grown rapidly since 1980 (Acemoglu and Autor 2011). So too has spatial sorting: workers increasingly self-select into different cities on the basis of skill (Berry and Glaeser 2005). In this paper, we demonstrate that income-inelastic housing demand translates aggregate income inequality into spatial sorting across cities, and we quantify the strength of the relationship with a calibrated model, whose key parameters we infer from consumption microdata.

Unskilled households have low incomes, devote a large share of their budgets to housing, and therefore sort into locations with lower housing costs. Skilled households have higher incomes, are relatively insensitive to housing costs, and sort into more expensive locations. A rising aggregate skill premium increases the between-skill difference in housing cost sensitivity and therefore causes ever more divergent location choices. We estimate a model of nonhomothetic housing demand using consumption microdata and find that the interaction of nonhomotheticity and the skill premium explains nearly a quarter of the increase in spatial sorting observed between 1980 and

[1] This is joint work with Trevor Williams

2010.

We begin by estimating nonhomothetic constant elasticity of substitution (NHCES) preferences over housing and nonhousing consumption. The key to our estimation strategy is to control for local housing costs, precisely because households' sorting decisions introduce a positive correlation between prices and incomes at the city level. We find housing demand is moderately income inelastic — for a household in the middle of the expenditure distribution, a 10% increase in total expenditure causes a 2.5% decrease in the housing expenditure share. We reject two alternative preferences used in the literature, Cobb-Douglas and a unit housing requirement.[2] Our estimates are stable across datasets and specifications with different instruments, controls, and fixed effects.

Next, we embed nonhomothetic preferences into a tractable spatial model to study sorting. Heterogeneous households with NHCES preferences trade off wages, amenities, and housing costs. In a partial equilibrium environment, we analytically derive a positive relationship between the aggregate skill premium and the intensity of spatial sorting. When preferences are nonhomothetic, the skill premium creates a wedge between the ideal price indices of skilled and unskilled households. The price indices of less skilled, and hence lower-income, households are endogenously more sensitive to housing costs. An increase in the skill premium increases this wedge in price indices and therefore causes location choices to diverge across skill groups.

To isolate the contribution of the rising skill premium to the increase in spatial sorting since 1980, we build a quantitative model with heterogeneity in productivities, amenities, and housing costs across locations. Even when wages and prices are endogenous, we show that changes in the location-neutral component of the skill premium only cause changes in sorting when preferences are nonhomothetic. Our main counterfactual shuts down the rising aggregate skill premium by scaling the relative

[2] "Unit housing requirement" refers to a model in which each household must purchase one unit of housing, so demand is perfectly price- and income-inelastic.

productivity of skilled workers in every location so that the skill premium, which is endogenous in the model, remains fixed at its 1980 level. In the absence of the rising skill premium, sorting increases 23% less than it did in the data. Our model rationalizes the remaining 77% of the observed increase with location-specific shocks to productivities, amenities, and housing costs. We conclude that the rising aggregate skill premium explains 23% of the observed increase in spatial sorting by skill.

Calibrating the model to our estimated preferences is essential to the result, both qualitatively and quantitatively. Under Cobb-Douglas preferences the skill premium has exactly no effect on sorting because sensitivity to housing costs does not vary with income. Calibrating the model to a unit housing requirement, by contrast, overstates the effect of the skill premium on sorting by a factor of two, because a unit housing requirement is an extreme form of nonhomotheticity in which the housing share declines one-to-one with income.

Finally, we estimate the growth of between-skill inequality in real consumption across space. High housing costs in large cities amplify the urban bias of the skill premium. As a result, real inequality is even more strongly urban-biased than nominal inequality, a result not discussed in prior literature.

By incorporating nonhomothetic preferences into a spatial model (see also Schmidheiny (2006), Eckert and Peters (2018), and Handbury (2019)), we emphasize sorting on prices. In related research, Ganong and Shoag (2017) connect changes in housing-supply regulations to the end of regional wage convergence. Diamond (2016) shows endogenous amenities intensify sorting.[3] Whereas these two papers explore the effects of idiosyncratic shocks — to local regulations and local productivities, respectively — we instead consider an aggregate shock that nevertheless has very different consequences across different locations. In this respect, we are similar to Eckert (2019), Giannone (2019), and Eckert, Ganapati, and Walsh (2020), who study urban-biased

[3]Our baseline model abstracts from endogenous amenities. In Section 2.4, we incorporate endogenous amenities and show our theoretical results are unchanged by this extension.

aggregate productivity shocks. However, in those papers, changes in sorting are driven by wages, whereas our mechanism operates through changes in ideal price indices.

We also contribute to a literature which studies the link between the aggregate income distribution and spatial sorting. A unit housing requirement — which implies housing demand is perfectly income inelastic — is a workhorse assumption in models of within-city sorting. Couture et al. (2019) use such a model to study how rising income inequality triggers a re-sorting of households across heterogeneous neighborhoods within a city, amplified by endogenous amenities. Fogli and Guerrieri (2019) also use a unit housing requirement in a model in which sorting interacts with human capital spillovers. Relative to these papers, we show that the same motive for sorting exists at the between-city level and discipline its quantitative importance by estimating housing demand using consumption microdata. This result is not obvious a priori. At the level of cities a common assumption, even in models with heterogeneous households, is that preferences are Cobb-Douglas and therefore homothetic (see, e.g., Eeckhout, Pinheiro, and Schmidheiny (2014), Diamond (2016), Fajgelbaum and Gaubert (2020)).

The Cobb-Douglas assumption is often justified by the fact that housing expenditure shares vary little across cities with very different income levels (Davis and Ortalo-Magné 2011). We offer an alternative explanation for the similarity of housing expenditure shares across cities: offsetting price and income effects, a view shared by Albouy, Ehrlich, and Liu (2016). Our demand elasticities are broadly similar to those in Albouy, Ehrlich, and Liu (2016), though we estimate housing demand using consumption microdata whereas they rely on city-level variation in incomes, prices and rental expenditures. Our estimation strategy thus avoids any assumptions about aggregating preferences within a city or about the mapping from income to expenditure. Relative to their work, we use the estimated preferences in a quantitative spatial model and perform counterfactuals.

Gyourko, Mayer, and Sinai (2013) is one of the few papers which combines non-homothetic preferences with a model of sorting at the between-city level. They show that an increase in the number of high-income households can create “superstar” cities and explain diverging house prices. Relative to Gyourko, Mayer, and Sinai (2013), our focus is on sorting by skill, rather than the evolution of housing costs. We move beyond a stylized environment and calibrate a spatial equilibrium model. The data reject the unit housing requirement assumed in that paper, and we show this modeling choice dramatically overstates the contribution of income inequality to spatial sorting. Furthermore, we use our model to conduct counterfactuals, supplementing the comparative statics and reduced form approach of Gyourko, Mayer, and Sinai (2013).

The rest of the paper proceeds as follows. Section 2.1 estimates housing demand. Section 2.2 uses the estimated preferences in a simple model to connect changes in the skill premium to changes in sorting. Section 2.3 calibrates a quantitative version of this model, and Section 2.4 uses the calibrated model to quantify the causal role of the skill premium in increasing sorting. Section 2.5 estimates changes in welfare and Section 2.6 concludes.

2.1 Estimating the Income Elasticity of Housing Demand

Estimating the income elasticity of housing demand presents three main challenges. First, we require expenditure data because the key parameter of the model is the elasticity of housing expenditure with respect to total expenditure.[4] Second, OLS estimates are biased by measurement error in expenditure, so we require an

[4] At the risk of ambiguity, we use the familiar term “income elasticity” as shorthand for “expenditure elasticity” throughout the paper.

instrument. Finally, and most importantly, the price of housing varies widely across space, and is correlated with household income. Therefore, we need to control for variation in housing prices. As we show below, failing to do so would strongly bias our results toward homotheticity.

2.1.1 Data

We use the restricted-access Panel Study of Income Dynamics (PSID), which identifies households' county of residence (University of Michigan Institute for Social Research 2021). Since 2005, the PSID has collected information on essentially all consumption covered by the Consumer Expenditure Survey (CEX) (Andreski et al. 2014). We use the 2005-2017 biennial surveys. Our baseline sample is restricted to renting households because they have a clear measure of housing consumption, but we also find similar results using homeowners.

The PSID has two advantages. One, we can link price data to about 90% of households in the PSID. By contrast, the CEX has geographic identifiers only for households in 24 large cities, which is less than half the CEX sample. Two, the PSID follows the same households over time, so we can study how housing expenditure responds to changes in total expenditure within the same household.

We estimate the price of housing for each Metropolitan Statistical Area (MSA) with a hedonic regression as in Albouy (2016). In principle, the price of housing is the market rent for a unit of housing services. In practice, our price indices are the set of MSA dummies in a regression of household rent on observed housing unit characteristics. We construct the price indices using two-year windows in the American Community Survey (ACS), starting in 2005 (Ruggles et al. 2020). For more details of our data, sample selection, and price indices, see Appendix B.1.

2.1.2 Preferences

Households have nonhomothetic constant elasticity of substitution (NHCES) preferences (Comin, Lashkari, and Mestieri 2021) over housing and a numéraire consumption good.The utility U of a household consuming h units of housing and c units of the consumption good is implicitly defined by

$$U^{\frac{\sigma-1}{\sigma}} = \Omega^{\frac{1}{\sigma}} h^{\frac{\sigma-1}{\sigma}} U^{\frac{\epsilon}{\sigma}} + c^{\frac{\sigma-1}{\sigma}}, \tag{2.1}$$

where $0 < \sigma < 1$, $\epsilon \geq \sigma - 1$, and $\Omega > 0$ are parameters.[5] [6] The household maximizes U subject to the budget constraint $ph + c \leq e$, where p is the price of housing and e is total expenditure.

NHCES preferences admit a straightforward Hicksian demand function. Denote the housing expenditure share $\eta \equiv \frac{ph}{e}$. The housing share satisfies

$$\log\left(\frac{\eta}{1-\eta}\right) = \log\Omega + (1-\sigma)\log p + \epsilon \log U. \tag{2.2}$$

We can see σ determines the sensitivity of housing expenditure to prices, and ϵ determines how housing expenditure varies with utility. In particular $\epsilon < 0$ implies the housing expenditure share falls with utility, whereas $\epsilon > 0$ implies the opposite. Because utility is monotonically increasing in total expenditure, ϵ determines the sign of the income elasticity of housing expenditure. If $\epsilon < 0$, housing is a necessity and its expenditure share falls with total expenditure, whereas if $\epsilon > 0$, it is a luxury and its share rises with total expenditure.

[5]The restriction $\sigma < 1$ implies housing demand is price-inelastic, which turns out to be the empirically relevant case. We impose $\sigma < 1$ purely for ease of exposition. NHCES preferences in general do allow $\sigma > 1$.

[6]Note that relative to a fully general formulation, (2.1) normalizes an ϵ-parameter for the numéraire consumption good to zero. Comin, Lashkari, and Mestieri (2021) show that in a single-location model this normalization is without loss of generality. It is also without loss of generality in the multi-location model we develop in Section 2.2, because we assume an isoelastic spatial labor supply function. See Appendix B.3.1 for a proof.

To take equation (2.2) to the data, we substitute out unobservable utility U in (2.1).[7] This yields an expression that implicitly defines η as function of expenditure, prices, and parameters:

$$\eta = \Omega e^{\epsilon} p^{1-\sigma} \left(1-\eta\right)^{1+\frac{\epsilon}{1-\sigma}}. \tag{2.3}$$

NHCES preferences nest two specifications commonly used in the spatial literature. Cobb-Douglas preferences are obtained by taking $\epsilon = 0$ and $\sigma \to 1$ in (2.3). In this case, the expenditure share is constant, equal to

$$\eta = \frac{\Omega}{1+\Omega}. \tag{2.4}$$

The opposite case, a unit housing requirement, is obtained by taking $\epsilon = -1$ and $\sigma \to 0$. Each household consumes Ω units of housing. In this case, the expenditure share is

$$\eta = \Omega \left(\frac{p}{e}\right). \tag{2.5}$$

For values of ϵ and σ between these two extremes, housing demand is income- and price-inelastic, but not perfectly so.

2.1.3 Estimation

We consider households indexed by i in years t. Households reside in MSAs indexed by n. Housing prices vary by location and year and are denoted by p_{nt}, whereas the price of the consumption good is assumed not to vary across space and is normalized to one. We interpret Ω as an idiosyncratic shock to an individual household's taste for housing, so that (2.3) becomes

$$\eta_{it} = \Omega_{it} e_{it}^{\epsilon} p_{nt}^{1-\sigma} \left(1-\eta_{it}\right)^{1+\frac{\epsilon}{1-\sigma}}. \tag{2.6}$$

[7] From the Hicksian demand for the consumption good, $U = (1-\eta)^{\frac{1}{1-\sigma}} e$.

To build intuition for our estimation strategy, we log-linearize (2.6) around the median housing share $\bar{\eta}$ to obtain

$$\hat{\eta}_{it} = \left(\frac{1-\bar{\eta}}{1-\bar{\eta}+(\frac{\epsilon}{1-\sigma}+1)\bar{\eta}} \right) \left(\hat{\Omega}_{it} + \epsilon \hat{e}_{it} + (1-\sigma)\hat{p}_{nt} \right), \tag{2.7}$$

where $\hat{x}$ denotes the log deviation of a variable x from its median. Equation (2.7) simplifies to

$$\hat{\eta}_{it} = \omega_{it} + \beta \hat{e}_{it} + \psi \hat{p}_{nt}, \tag{2.8}$$

where $\omega_{it} \equiv \left(\frac{1-\bar{\eta}}{1-\bar{\eta}+(\frac{\epsilon}{1-\sigma}+1)\bar{\eta}} \right) \hat{\Omega}_{it}$ and β and ψ are defined analogously. Under the null of homothetic preferences, $\epsilon = \beta = 0$. We bring (2.8) to the data by modeling the demand shifter ω_{it} as a function of observable demographic characteristics, year fixed effects, and an additive error. Formally,

$$\hat{\eta}_{it} = \omega_t + \omega' X_{it} + \beta \hat{e}_{it} + \psi \hat{p}_{nt} + \xi_{it}, \tag{2.9}$$

where X_{it} is a vector with the age, gender, and race of the household head, household size, and the number of earners in the household. We observe total expenditure e_{it}, the housing expenditure share η_{it}, and prices p_{nt}. Consistent with our focus on across-city heterogeneity, we assume a common housing market in each location, so prices do not vary within n. The error term ξ_{it} represents measurement error in expenditure and random shocks to housing demand.

2.1.4 Results

Table 2.1, columns (1) - (4), show the estimates of equation (2.9). Note that because columns (1) and (2) do not attempt to estimate the coefficient on prices, they cannot recover the structural parameters ϵ and σ. Column (1) estimates equation (2.9) by OLS without controlling for price $\hat{p}_{nt}$. The point estimate indicates significant nonhomotheticity, but two sources of bias are evident. First, measurement error in expenditure is likely to bias $\hat{\beta}$ downwards.[8] Second, a positive correlation between prices and expenditure, reflecting the sorting of high-income households into high-price MSAs, will bias $\hat{\beta}$ upwards.

Column (2) addresses measurement error by instrumenting for log expenditure using log income, following Lewbel (1996), Davis and Ortalo-Magné (2011), and Aguiar and Bils (2015). As expected, $\hat{\beta}$ rises toward zero. The exclusion restriction here is that income is unrelated to the housing share, conditional on the true level of

Table 2.1: Preference Estimates
Dependent variable: Log housing share

	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) GMM
ϵ			-0.436	-0.291	-0.306
			(0.018)	(0.037)	(0.036)
σ			0.436	0.542	0.522
			(0.039)	(0.079)	(0.075)
Log expenditure	-0.298	-0.162	-0.393	-0.248	
	(0.028)	(0.039)	(0.021)	(0.035)	
Log price			0.508	0.390	
			(0.026)	(0.057)	
Demographic controls	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
R^2	0.12	0.10	0.20	0.16	
First-stage F-stat.		1,264,4		107.3	
N	12,351	12,351	12,351	10,678	10,678
No. of clusters	484	484	484	217	217

Source: PSID, Census, and Saiz (2010)

Note: Renters only. Instrument is log family income. Demographic controls are bins for family size, number of earners, and sex, race, and age of household head. Standard errors clustered at MSA level. See Appendix B.1 for further details of sample construction.

[8]Because expenditure appears in the denominator of $\hat{\eta}$, the bias in $\hat{\beta}$ is not standard classical measurement error. See Appendix B.2 for a short proof.

expenditure. One threat to identification is that if housing expenditure is subject to some adjustment costs, it may react to income changes more slowly than overall expenditure. This would bias our estimates downwards. Another threat is that there may be permanent, unobservable differences in housing demand across households which are correlated with income. We address both these concerns with alternative specifications in Table 2.2. Column (3) of Table 2.1 returns to OLS but addresses omitted variable bias by controlling for prices. Relative to column (1) the coefficient on log expenditure falls, implying very income-inelastic housing demand. This result is consistent with high-income households sorting into high-price MSAs — exactly the pattern the model developed in the next section will predict. Using (2.7) we also back out estimates of the structural parameters ϵ and σ.

Together columns (2) and (3) show that failing to instrument for expenditure and control for prices introduces offsetting biases in the coefficient on expenditure. Column (4) corrects for both biases simultaneously by instrumenting for expenditure using income and controlling for prices. Prices are potentially endogenous because they are a function of housing demand. For example, a city-level shock to housing demand might increase expenditure shares and, consequently, prices. We instrument for prices using Saiz (2010)'s measures of regulatory and geographical constraints on housing construction. These instruments are relevant if tight regulatory or geographical constraints force up local housing costs. They satisfy the exclusion restriction if, conditional on prices and total expenditure, these constraints don't have an effect on housing expenditure. The results in column (4) imply that housing demand is moderately nonhomothetic and price inelastic. For a household in the middle of the expenditure distribution, a 10% increase in total expenditure causes a 2.48% decrease in the housing expenditure share.

Finally, column (5) shows our preferred specification. Here, we estimate the nonlinear equation (2.3) directly by GMM.[9] Similarly to column (4), we instrument for expenditure and prices, and allow ω_{it} to vary with demographic characteristics and year. The estimated ϵ and σ are close to their values in column (4).

We now compare the preferences estimated in Table 2.1 to two benchmarks from the literature: Cobb-Douglas preferences and a unit housing requirement. We begin with formal statistical tests. The NHCES preferences estimated above nest both of these special cases. The null hypothesis of Cobb-Douglas preferences, corresponding to $\epsilon = 0$ and $\sigma = 1$, can be rejected at the 1% level. A unit housing requirement corresponds to $\epsilon = -1$ and $\sigma = 0$ — again, column (5) allows us to reject this null hypothesis at the 1% level. Although our NHCES specification is more flexible than these special cases, it still imposes a particular functional form on the relationship between total expenditure and housing expenditure. To assess the validity of this assumption we construct a binned scatterplot of expenditure against housing shares.[10] The results are shown in Figure 2.1 alongside our estimated preferences (the solid line). Our estimated preferences appear to fit the data well. For comparison we also plot Cobb-Douglas preferences and a unit housing requirement, given by the dashed and dotted lines respectively. Neither alternative comes close to matching the data.

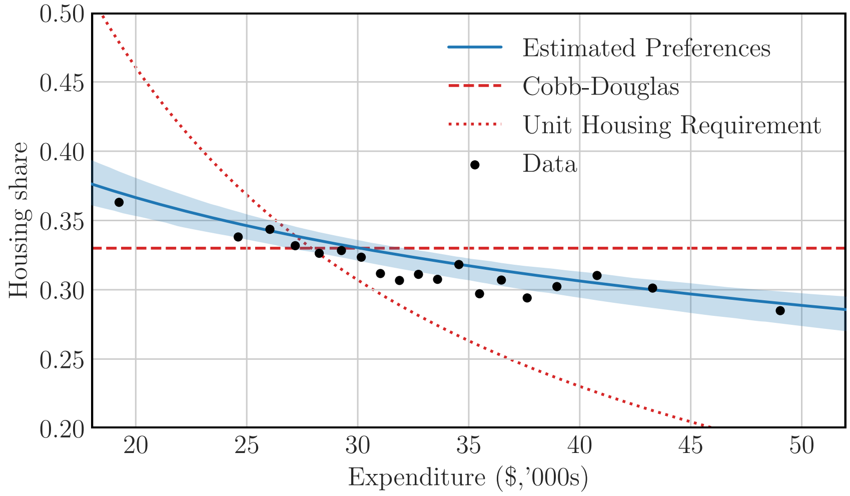
Alternative Specifications

Table 2.2 shows alternative specifications. Housing costs are not the only prices which vary across space, and in column (1) we examine if our results are robust to controlling for this variation. We denote nonhousing prices in location n and year t by q_{nt} and incorporate these into the theory developed in Subsection 2.1.2. Equation

[9] In stating our preferences, we imposed $\epsilon > \sigma - 1$ and $0 < \sigma < 1$. We do not impose these restrictions in our estimation procedure, but they are satisfied by the values obtained in column (5).

[10] We do not use expenditure directly, since as discussed above measurement error contaminates the relationship between expenditure and the housing share. Instead, we predict total expenditure for each household using the instruments and covariates in column (5) of Table 2.1, then split households into twenty bins of predicted expenditure and calculate the average housing share in each bin, partialling out covariates.

Figure 2.1: Housing Expenditure Shares



Notes: 'Estimated Preferences' plots (2.3) at the parameter values obtained in Table 2.1, column (5). The shaded area represents a 95% confidence interval.'Cobb-Douglas' and 'Unit Housing Requirement' plot the preferences described by (2.4) and (2.5), respectively, with the scale parameter Ω chosen to match an expenditure share of 0.33 at the median level of total expenditure. 'Data' plots the average housing share in twenty evenly sized bins defined by predicted total expenditure, whose construction is described in the text.

(2.6) becomes

$$\eta_{it} = \Omega_{it} \left(\frac{e_{it}}{q_{nt}} \right)^{\epsilon} \left(\frac{p_{nt}}{q_{nt}} \right)^{1-\sigma} (1 - \eta_{it})^{1+\frac{\epsilon}{1-\sigma}} . \tag{2.10}$$

The nonhousing price index q_{nt} is the price of a Cobb-Douglas aggregate of goods and nonhousing services constructed from the Bureau of Economic Analysis (BEA) Metropolitan Regional Price Parities (Bureau of Economic Analysis 2020). We estimate (2.10) by GMM. The point estimate of ϵ is virtually unchanged relative to its value in column (5) in Table 2.1, while the estimate of σ is somewhat smaller — but not significantly so.

Another potential concern is that our MSA housing price indices may fail to pick up unobserved differences in housing quality, beyond the dwelling characteristics available in Census data. Column (2) dispenses with prices entirely and instead estimates the linearized equation (2.9) with MSA fixed effects. While excluding prices means we cannot recover the structural parameters ϵ and σ in this specification, we can still estimate the average elasticity of the expenditure share η to log expenditure.

We find a value of -0.261, very close to the value of -0.248 reported in column (4) of Table 2.1. The estimated relationship between the housing expenditure share and overall expenditure is not sensitive to how we control for price variation across MSAs, but only to whether or not we do.

Column (3) considers an alternative instrument for expenditure. A natural concern is that housing expenditure is relatively insensitive to total expenditure because housing expenditure can only be adjusted slowly while total expenditure may fluctuate with transitory income shocks. Column (3) addresses this concern by instrumenting for expenditure using the household's education level. Since differences in education across households are permanent[11], slow adjustment of housing expenditure to transitory shocks is irrelevant in this specification. The point estimates in column (3) are similar to those in our baseline specification and again indicate that housing demand is significantly nonhomothetic.

Finally, we consider the possibility of permanent, unobservable differences in housing demand across households. We parameterize the demand shifter Ω_{it} as follows

$$\log \Omega_{it} = \omega_i + \omega_t + \omega' \tilde{X}_{it} + \xi_{it}$$

where ω_i is a household fixed effect, $\tilde{X}_{it}$ is the subset of demographic controls which are time-varying and ξ_{it} is an idiosyncratic error term. Taking logs of (2.3) then yields

$$\log \eta_{it} = \omega_i + \omega_t + \omega' \tilde{X}_{it} + \epsilon \log e_{it} + (1 - \sigma) \log p_{nt} + \left(1 + \frac{\epsilon}{1 - \sigma}\right) \log (1 - \eta_{it}) + \xi_{it} \tag{2.11}$$

Equation (2.11) allows for permanent unobservable differences in housing demand across households, captured by ω_i. If ω_i happens to be negatively correlated with income, this specification could generate the negative relationship between expenditure and η found in Table 2.1 even when $\epsilon = 0$ and preferences are homothetic.

[11] For 90% of households education level does not change while they are in the sample.

Such permanent differences in housing demand are sometimes used in the literature as a tractable alternative to explicitly nonhomothetic preferences (Diamond 2016; Notowidigdo 2020; Colas and Hutchinson 2021). As we will show in Section 2.2, however, distinguishing between such demand shifters and explicitly nonhomothetic preferences is critical for the mechanism we focus on in this paper.

We demean (2.11) at the household level so that ω_i drops out.[12] We estimate

Table 2.2: Preferences, Alternative Specifications
Dependent variable: Log housing share

	(1) GMM	(2) 2SLS	(3) GMM	(4) GMM
ϵ	-0.300		-0.271	-0.465
	(0.041)		(0.065)	(0.109)
σ	0.389		0.523	0.511
	(0.090)		(0.077)	(0.198)
Log expenditure		-0.261		
		(0.032)		
Demographic controls	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Non-housing prices	✓			
MSA FE		✓		
Household FE				✓
IV	Income	Income	Education	Income
R^2		0.15		
First-stage F-stat.		1,054.6		
N	8,183	12,257	10,271	8,670
No. of clusters	208	390	216	197

Source: PSID, Census, and Saiz (2010)

Note: Renters only. Demographic controls are bins for family size, number of earners, and sex, race, and age of household head. Column (4) includes only time-varying demographic controls. Standard errors clustered at MSA level. See Appendix B.1 for further details of sample construction.

[12]In Appendix Table B.1, column (7), we pursue an alternative estimation strategy by log-linearizing (2.11) and using 2SLS with household fixed effects. We find almost identical point estimates.

the demeaned equation by GMM, using the same instruments as in column (5) of Table 2.1. Since the instruments for p_{nt} do not vary over time, σ is identified only by households who face different prices because they move between MSAs. The results are reported in column (4) of Table 2.2. The point estimate for ϵ falls relative to our baseline, indicating somewhat stronger nonhomotheticity, but the two estimates are not significantly different. The price elasticity σ is very close to its baseline value. We are still able to reject both Cobb-Douglas preferences and a unit housing requirement. We conclude that permanent, unobservable differences in housing demand across households are not driving our baseline results: even within a single household, an increase in total expenditure decreases the housing expenditure share.

Finally, we explore a number of alternative specifications and data sources in Appendix B.2. Continuing to use the PSID, we examine the robustness of our results to to controlling for liquid wealth; to removing demographic controls; to using prices directly rather than instrumenting for them; to using alternative data sources and geographies for prices; to splitting the sample into movers and non-movers; and to using alternative instruments for expenditure. We continue to find that housing demand is moderately income-inelastic. We replicate our results using the CEX, and then extend them to include homeowners (Bureau of Labor Statistics 2020a). The estimated parameters look very similar when we include homeowners.

2.1.5 Connections to prior work

Cobb-Douglas preferences are a workhorse assumption in quantitative spatial models. Davis and Ortalo-Magné (2011) support this assumption by showing that median housing shares were roughly constant across cities from 1980 to 2000. However, constant housing shares over space are necessary but not sufficient to conclude that preferences are Cobb-Douglas. Cross-city comparisons of housing shares reflect both income and price effects. Households in expensive cities have a higher housing share than households in inexpensive cities at every level of income; but the composition of expensive cities is tilted toward high-income households, who tend to have lower housing shares. Previous research documenting homothetic or near-nomothetic housing demand (Davis and Ortalo-Magné 2011; Aguiar and Bils 2015) did not control for variation in local prices. This is consistent with our results: column (2) of Table 2.1 shows failing to control for local prices biases the estimated income elasticity toward homotheticity.

Albouy, Ehrlich, and Liu (2016) allow for nonhomotheticity when estimating preferences over housing and nonhousing consumption, and find that housing demand is moderately income inelastic. That paper aggregates to the MSA level and uses data on income rather than expenditure, while we take individual households as our unit of analysis and use expenditure data. We view our results as complementary to Albouy, Ehrlich, and Liu (2016)'s, but note that our approach avoids some assumptions which are inherent in theirs. Our estimation procedure does not assume that demands can be aggregated across households of different income levels. Furthermore, by directly using data on expenditure we avoid assumptions on the relationship between expenditure and income. Finally, using variation within a household allows us to reject the hypothesis that the observed negative relationship between the housing share and total expenditure is driven by permanent, unobservable household characteristics. This is not possible when the data are aggregated to the MSA level.

2.2 Model

Having established that housing demand is income inelastic in the data, we now explore the implications for spatial sorting by skill. We characterize the relationship between the skill premium and sorting in a simple partial equilibrium model with nonhomothetic preferences. We then construct a quantitative general equilibrium model for the counterfactuals in Section 2.4.

2.2.1 Simple Model

Production and Wages

There are two types of household, skilled and unskilled, with types denoted by $i = s, u$. Households supply labor to tradable-goods producers in their home location, denoted by n. There are no trade costs. Firms are perfectly competitive and produce using skilled and unskilled labor according to the function

$$F_n(l_{sn}, l_{un}) = z_n(A \cdot l_{sn} + l_{un}), \tag{2.12}$$

where z_n is the productivity of region n, A is the relative productivity of skilled labor, and l_{in} is the labor input of type i. Skilled and unskilled labor are perfect substitutes, and their relative productivities do not vary across locations. This assumption implies that the skill premium is exogenous and equal to A in every location. We therefore refer to A as the aggregate skill premium. Households do not save, so wages w_{in} are exactly equal to expenditure e_{in}. Expenditures and wages satisfy

$$e_{sn} = w_{sn} = z_n A \tag{2.13}$$

$$e_{un} = w_{un} = z_n. \tag{2.14}$$

Housing Supply

Each location has a competitive housing sector that transforms p_n units of the consumption good into 1 unit of housing. This means that housing is elastically supplied at an exogenous price p_n.

Location Choice and Preferences

We first describe the problem of a household in a given location, then turn to the household's choice of location. Households have NHCES preferences as in (2.1). The utility of a household of type i in location n, denoted by v_{in}, is

$$\begin{aligned} v_{in} \equiv \quad & \max_{c,h} \quad U \\ & \text{s.t} \quad U^{\frac{\sigma-1}{\sigma}} = \Omega^{\frac{1}{\sigma}} h^{\frac{\sigma-1}{\sigma}} U^{\frac{\epsilon}{\sigma}} + c^{\frac{\sigma-1}{\sigma}}, \\ & e_{in} = c + p_n h. \end{aligned} \tag{2.15}$$

The solution to this problem yields housing expenditure shares $\eta_{in} = \eta(e_{in}, p_n)$ which satisfy our estimating equation (2.3) from Section 2.1. If $\epsilon < 0$, as we estimated in Section 2.1, η_{in} is a decreasing function of total expenditure e_{in}. In estimating (2.3) we allowed Ω to vary across households. Here we instead impose a common Ω across households.[13] Doing so keeps the model tractable and removes a source of sorting that would distract from the effects of the skill premium that we focus on.

We close the model by assuming location n's share of total employment of type i is an isoelastic function of utility v_{in} given by

$$l_{in} = \frac{v_{in}^{\theta} B_n}{\sum_m v_{im}^{\theta} B_m} L_i, \tag{2.16}$$

where L_i is the exogenous national population of households of type i.[14] We refer to B_n as the amenity value of location n and θ as the migration elasticity.

[13] In column (6) of Appendix Table B.1 we drop all demographic controls — equivalent to assuming a common Ω across households — and find that the results are identical to those obtained in our baseline specification. We conclude that controlling for demographics is not important in measuring the income elasticity of housing demand.

[14] One microfoundation of the employment shares (2.16), common in quantitative spatial models, is that each household draws an n-vector of idiosyncratic location preference shocks from independent Fréchet distributions with scale B_n and shape θ (Allen and Arkolakis 2014; Redding 2016). We do not assume a particular microfoundation.

Equilibrium

Given parameters $(\epsilon, \sigma, \Omega, \theta)$, location-specific fundamentals (z_n, B_n, p_n), the aggregate skill premium A, and aggregate labor supplies (L_u, L_s), an equilibrium is a vector of employment levels l_{in}, wages w_{in}, and total expenditures e_{in} satisfying (2.13), (2.14), (2.15) and (2.16).

2.2.2 Analytical Results

We now characterize spatial sorting by skill.[15] We define sorting in terms of the log skill ratio in each location, denoted by s_n and satisfying,

$$s_n \equiv \log\left(\frac{l_{sn}}{l_{un}}\right).$$

As we will shortly see, this object is analytically convenient in the context of our model. Our proposed measure of sorting, which we denote by S, is the variance of the log skill ratio,

$$S = \text{Var}\left(s_n\right).$$

S is zero when skilled workers are distributed in proportion to unskilled workers across space, and rises as they become more clustered. Additionally, S is invariant to proportional increases in the number of skilled workers in all locations. This invariance property is desirable because the number of skilled workers in the US has grown relative to the number of unskilled workers since 1980.

[15] As in Section 2.1, we assume $\sigma \in (0, 1)$. Proofs of all the statements in this section can be found in Appendix B.3.

The Determinants of Sorting

Equation (2.16) yields a simple expression for s_n in terms of utilities v_{in},

$$s_n = \zeta + \theta \log \left(\frac{v_{sn}}{v_{un}} \right), \tag{2.17}$$

where ζ is a function of fundamentals that does not vary across locations. Note B_n is absent by design: because amenities do not differ by type, they do not drive sorting.

To relate v_{in} to wages and prices, consider the ideal price indices, P_{in}, which satisfy

$$v_{in} = \frac{w_{in}}{P_{in}}, \tag{2.18}$$

where we are exploiting the fact that in the simple model wages are equal to expenditure. Substituting (2.18) into (2.17) and using (2.13) and (2.14) to replace wages with productivities yields

$$s_n = \zeta + \theta \log A - \theta \log \left(\frac{P_{sn}}{P_{un}} \right). \tag{2.19}$$

This expression clarifies that wages do not cause sorting conditional on the price indices, because the ratio of skilled to unskilled wages is constant across locations. The skill premium can in fact be absorbed into the constant term. Instead, sorting is only a result of differences in the ideal price indices.

To see how these price indices depend on the wages and prices, we use expressions for P_{un} and P_{sn} implied by our NHCES preferences:

$$P_{un} = \left(1 + \Omega \left(\frac{z_n}{P_{un}} \right)^{\epsilon} p_n^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{2.20}$$

$$P_{sn} = \left(1 + \Omega A^{\epsilon} \left(\frac{z_n}{P_{sn}} \right)^{\epsilon} p_n^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \tag{2.21}$$

These price indices resemble ordinary CES price indices, except the weight placed on housing is a function of real income as long as $\epsilon \neq 0$. In particular, if $\epsilon < 0$ — as we found in Section 2.1 — this weight is decreasing in real income. Moreover, the housing weight for skilled workers is always lower than for unskilled workers and decreases with A. Inspection of (2.20) and (2.21) shows each can be written as a function $P_i(c_n)$, where $c_n \equiv z_n^{\frac{\epsilon}{1-\sigma}} p_n$ is defined as productivity-adjusted housing cost. Intuitively, z_n should appear in c_n when $\epsilon < 0$ because a higher income lowers the burden of higher house prices when housing demand is income inelastic.

We consider the implications for sorting, starting with the homothetic case. In this case, $\epsilon = 0$ and $P_u(c) = P_s(c)$ for all c. Inspection of (2.19) then shows s_n does not depend on n. Skilled and unskilled workers are distributed in proportion to one another in every location and $S = 0$.

When preferences are nonhomothetic and $\epsilon < 0$, $P_u(c)$ is a steeper function of c than $P_s(c)$. As c grows, the wedge between unskilled and skilled price indices grows and high c locations look increasingly unattractive to unskilled workers. Lemma 1 formalizes this argument.

Lemma 1. *Suppose housing demand is income inelastic so that $\epsilon < 0$. Then, $\log \frac{P_u(c)}{P_s(c)}$ is a strictly increasing function of productivity-adjusted housing cost c. Equation* (2.19) *then implies the skill ratio s_n is a strictly increasing function of c.*

The dotted and solid lines in panel (a) of Figure 2.2 illustrates this mechanism. The ideal price index for unskilled households is a steep function of c, whereas the ideal price index for skilled households is flatter. By (2.19), the skill ratio s_n in Panel (b) is just an affine transformation of the gap between the dotted and solid lines in Panel (a).

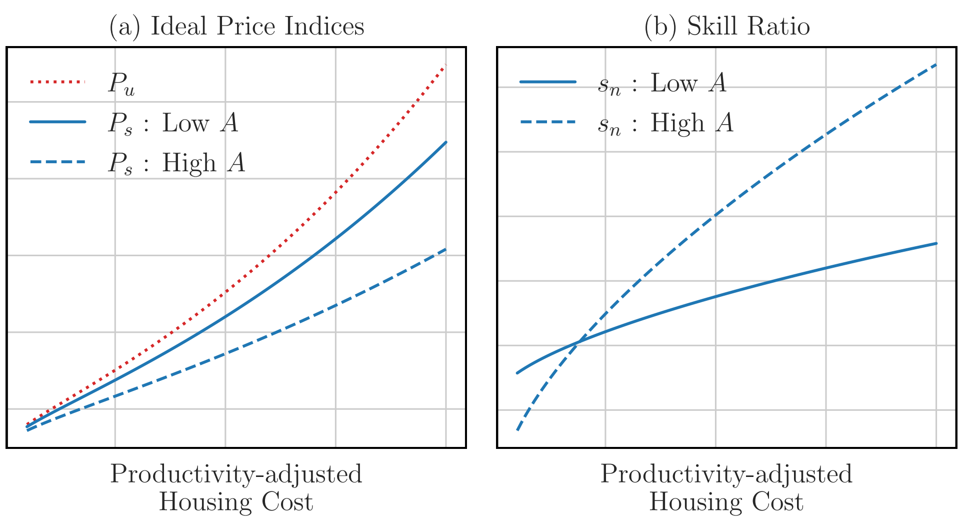
Sorting and Changes in the Skill Premium

So far, we have focused on the level of sorting. Now we turn to changes in sorting caused by changes in the skill premium. Proposition 1 states our main result by studying a small increase in the skill premium, $d \log A > 0$.

Proposition 1. *Suppose housing demand is income inelastic so that $\epsilon < 0$. Consider an increase in the skill premium, $d \log A > 0$. Then, ds_n is a strictly increasing function of s_n. Sorting rises, $dS > 0$. If, instead, $\epsilon = 0$, then $ds_n = 0$ for all n and $dS = 0$.*

Equations (2.20) and (2.21) show skilled and unskilled ideal price indices differ only because of the aggregate skill premium A. As A rises, skilled households place less weight on housing costs and become more willing to live in locations with a high productivity-adjusted housing cost c. This flattening of the ideal price index is illustrated by the dashed line in Panel (a) of Figure 2.2, which increases the skill premium relative to the solid line. The gap between P_u and P_s grows and (2.19) tells us s_n must then become a steeper function of c, as shown by the dashed line in Panel (b). Skilled households, newly insensitive to housing costs, flee cheap locations toward the left of Panel (b), and instead cluster in expensive ones on the right. The

Figure 2.2: Ideal Prices Indices and the Skill Ratio



Note: The dotted and solid lines in Panel (a) plot the price indices defined by (2.20) and (2.21), respectively, as functions of productivity-adjusted housing cost. The dashed line plots (2.21) again but uses a higher value of the skill premium A. The solid line in Panel (b) plots the log skill ratio s_n given by (2.19), corresponding to the dotted and solid lines in Panel (a). The dashed line plots s_n but uses the value for P_s given by the dashed line in Panel (a).

higher skill premium reinforces pre-existing patterns of sorting and so S rises.[16]

Comparison to Cobb-Douglas

Above we have emphasized that when $\epsilon = 0$ and preferences are homothetic, households do not sort on prices and their sorting decisions do not diverge as the skill premium rises. It is reasonable to ask whether the same mechanism might be captured by allowing for exogenous, skill-specific differences in the housing expenditure share while retaining a Cobb-Douglas specification within each skill group. This has appeared in the literature as a tractable stand-in for nonhomothetic preferences (Diamond 2016; Notowidigdo 2020; Colas and Hutchinson 2021). The answer to this question is no.

Cobb-Douglas preferences imply the price index $P_{in} = p_n^{\kappa_i}$, where $\kappa_i \in (0, 1)$ is the housing share for type i. We assume $\kappa_u \geq \kappa_s$. Equation (2.17) becomes

$$s_n = \zeta + \theta \log A + \theta(\kappa_u - \kappa_s) \log p_n. \tag{2.22}$$

Equation (2.22) shows that when $\kappa_u > \kappa_s$, skilled households will sort into high price locations, just as in the model above. However, unlike in our explicitly nonhomothetic model, changes in the aggregate skill premium A do not cause changes in spatial sorting by skill.[17] From (2.22), we can see that the skill premium enters identically in every location n and does not interact with prices p_n. We conclude that in order to capture the mechanism we focus on, it is not enough to impose different expenditure shares by type — instead changes in income must alter the weight each skill group places on housing costs.

[16] This logic may fail when prices are endogenous, since price changes are not guaranteed to be monotonic with respect to the initial level of sorting. However, we account for endogenous prices in the quantitative model.

[17] This result is only exactly true in our simple model with exogenous housing costs and wages. In Appendix B.5.1 we repeat our main counterfactual using this form of Cobb-Douglas preferences, and find that the implied relationship between the skill premium and spatial sorting is quantitatively negligible.

2.2.3 Quantitative Model

To take the model to the data, we enrich it on several dimensions.

The simple model deliberately shut down sorting based on wages. We relax this assumption by replacing (2.12) with a CES production function,

$$F_n(l_s, l_u) = Zz_n \left((Aa_n l_s)^{\frac{\rho-1}{\rho}} + l_u^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}, \tag{2.23}$$

where ρ is the elasticity of substitution between skilled and unskilled labor. Implied wages are

$$w_{un} = Zz_n l_{un}^{\frac{-1}{\rho}} \left((Aa_n l_{sn})^{\frac{\rho-1}{\rho}} + l_{un}^{\frac{\rho-1}{\rho}} \right)^{\frac{1}{\rho-1}} \tag{2.24}$$

$$w_{sn} = (Aa_n)^{\frac{\rho-1}{\rho}} Zz_n l_{sn}^{\frac{-1}{\rho}} \left((Aa_n l_{sn})^{\frac{\rho-1}{\rho}} + l_{un}^{\frac{\rho}{\rho}} \right)^{\frac{1}{\rho-1}}. \tag{2.25}$$

As in the simple model, the quantitative model contains a location-specific productivity shock z_n and an economy-wide skill bias A. We additionally allow for location-specific skill bias using the shifter a_n, so that skilled households may have a comparative advantage working in, say, San Francisco relative to Detroit. The economy-wide productivity shifter Z is for notational convenience when we conduct counterfactuals.

We allow amenities B_{in} to differ by skill, so that (2.16) becomes

$$l_{in} = \frac{v_{in}^{\theta} B_{in}}{\sum_m v_{im}^{\theta} B_{im}} L_i. \tag{2.26}$$

Note that amenities B_{in} do not enter the problem (2.15) which defines v_{in} and η_{in} and so do not directly affect housing demand.[18] With these modifications, our model can capture changes in sorting driven by location-specific changes in a location's attractiveness to skilled versus unskilled households, through both wages and amenities.

[18]Of course, amenities may still influence housing demand through their effect on endogenous wages and prices, but this poses no threat to the identification strategy pursued in Section 2.1.

As in the simple model, our focus remains on changes in overall sorting driven by location-neutral changes in A.

In the simple model, the price of housing p_n was exogenous. In reality, increases in A push skilled households toward expensive cities, putting upward pressure on housing costs and crowding out unskilled households. The quantitative model captures this feedback to house prices by including inelastic housing supply as in Hsieh and Moretti (2019). The price of housing in location n is given by

$$p_n = \Pi_n \left(HD_n\right)^{\gamma_n}, \tag{2.27}$$

where HD_n is (physical) housing demand in n and Π_n is an exogenous price shifter. γ_n, the inverse elasticity of housing supply, is allowed to vary by location to reflect different physical or regulatory constraints on building. Housing demand is the sum of housing consumption by both types of households:

$$HD_n = p_n^{-1} \sum_i \eta_{in} e_{in} l_{in}. \tag{2.28}$$

Finally we enrich the mapping from income w_{in} to expenditure e_{in}. There are two differences between income and expenditure. The first is that the relevant quantity for expenditure is permanent income, but in the data we only observe current income. However, aggregating to the level of a skill group averages away any transitory income shocks, making this less of a concern. Second, taxes create a wedge between income and expenditure. We incorporate this wedge into our model following Heathcote, Storesletten, and Violante (2017),

$$e_{in} = \lambda w_{in}^{1-\tau}. \tag{2.29}$$

where we impose that expenditure is equal to after-tax income. τ determines the progressivity of the tax system and λ is chosen so that the government budget balances.

Equilibrium

Given parameters $(\epsilon, \sigma, \Omega, \theta, \rho, \tau, \lambda, \{\gamma_n\})$, location-specific fundamentals $(a_n, z_n, B_{un}, B_{sn}, \Pi_n)$ for all n, aggregate fundamentals (Z, A), and labor supplies (L_u, L_s), an equilibrium is a vector of populations l_{in}, wages w_{in}, total expenditures e_{in}, expenditure shares η_{in}, housing demands HD_n, and prices p_n satisfying equations (2.15), (2.25), (2.26), (2.27), (2.28), and (2.29).

A Neutrality Result

We conclude this section by extending part of Proposition 1 to the quantitative model. Crucially, although our quantitative model accommodates rich patterns of sorting based on location-specific skill biases a_n and amenities B_{in}, homotheticity shuts down any relationship between aggregate productivity A and sorting. Proposition 2 formalizes this point.

Proposition 2. *Suppose $\epsilon = 0$ so that preferences are homothetic. Then, S, the level of sorting, does not depend on A.*

See Appendix B.3.4 for a proof. To gain intuition for this result, return to the expression for the log skill ratio s_n derived in the simple model. In the quantitative model with $\epsilon = 0$, (2.17) is modified to

$$s_n = \zeta + \theta \log\left(\frac{e_{sn}}{e_{un}}\right) + \log\left(\frac{B_{sn}}{B_{un}}\right). \tag{2.30}$$

Using (2.29) to replace expenditures with wages, and then using (2.25) to replace wages with productivities and the skill ratio, we obtain

$$s_n = \vartheta_0 + \vartheta_1 \log a_n + \vartheta_2 \log\left(\frac{B_{sn}}{B_{un}}\right), \tag{2.31}$$

where ϑ_0, ϑ_1, and ϑ_2 are constants. The striking feature of (2.31) is that s_n is entirely determined by location-specific fundamentals a_n and B_{in}, up to the additive constant $\tilde{\zeta}$. Changes in A have no impact on the distribution of skill ratios, and thus no impact on sorting, when preferences are homothetic. Proposition 2 is useful because it implies any changes in sorting in our quantitative model following changes in A are ultimately the result of nonhomothetic housing demand.

2.3 Calibration

Section 2.1 estimated nonhomothetic preferences over housing consumption, and Section 2.2 embedded these preferences in a simple model to make our key theoretical point — increases in the aggregate skill premium cause increases in spatial sorting. To determine the importance of this force in explaining trends in sorting since 1980, we now calibrate the quantitative model. The crucial preference parameters — ϵ and σ — are set at the values obtained in Section 2.1. The scale parameter Ω is not identified separately from the scale of prices (discussed below), so we normalize it to 1 in each year. The calibration of the remaining parameters is discussed in detail below: the elasticity of substitution ρ is calibrated from the literature; we derive estimating equations from the model for the tax-progressivity parameter τ and the housing-supply elasticities γ_n; and we calibrate the migration elasticity θ by targeting literature estimates. The results of this exercise are summarized in Table 2.3.

2.3.1 Data

Location-level information on wages, rents and employment are from IPUMS (Ruggles et al. 2020). We use the 5% population samples of the 1980, 1990, and 2000 decennial censuses and the 3% population sample from the 2009-2011 ACS. We have a balanced panel of 269 locations: 219 MSAs and the 50 non-metropolitan por-

Table 2.3: Parameters

Parameter	Value	Role	Source
ϵ	-0.306	Income elasticity	PSID
σ	0.522	Price elasticity	PSID
ρ	3.850	Production	Card (2009)
τ	0.174	Taxation	PSID
$\{\gamma_n\}$	0.630[a]	Housing supply	Census
θ	5.106	Migration	Indirect inference

[a] Employment-weighted mean

tions of states. Our census sample consists of prime-age adults who report strong labor-force attachment. Wages and rents are deflated by the Consumer Price Index (CPI) excluding shelter (Bureau of Labor Statistics 2020b). Location-level price indices are constructed each year from a hedonic rents regression as in Section 2.1. The hedonic approach does not recover the level of prices, so we scale prices to match the average housing share from the CEX in each year, which rose from 0.32 in 1980 to 0.41 in 2010.[19]

See Appendix B.1 for more details on the data.

2.3.2 Parameters

Elasticity of substitution

The production side of the model is standard and we externally calibrate $\rho = 3.85$ to match Card (2009).[20] That paper estimates the elasticity of substitution between workers of different skill groups at the MSA level using immigration as an instrument for labor-supply changes. The elasticity is larger than canonical estimates from Katz and Murphy (1992) and Acemoglu and Autor (2011), who report values close to 1.6. However, Katz and Murphy (1992) estimate an aggregate production function on

[19]Rent equivalent for owners was not surveyed until 1984. Owners' housing shares grew only somewhat slower than renters' over time.

[20]See Table 5, column (7), in Card (2009) for the negative inverse elasticity of -0.26.

time-series data, whereas Card (2009) estimates a city-level production function on cross-sectional data. Studies estimating the elasticity of substitution at the city level tend to find values between 3 and 5 (Bound et al. 2004; Beaudry, Doms, and Lewis 2010; Baum-Snow, Freedman, and Pavan 2018; Eckert, Ganapati, and Walsh 2020).[21]

Tax system

To calibrate the progressivity parameter τ, we follow Heathcote, Storesletten, and Violante (2017). From (2.29), log post-tax income for household i in year t is equal to

$$\log y_{it} = \log \lambda_t + (1-\tau) \log w_{it}. \tag{2.32}$$

Regressing log post-tax income on log pre-tax income and a year fixed effect in the PSID for 1980, 1990, 2000, and 2010 yields $\hat{\tau} = 0.174$, close to the value of 0.181 reported by Heathcote, Storesletten, and Violante (2017) for 1978-2006.

Housing-supply elasticities

The housing-supply equation (2.27) is specified in terms of the physical quantity of housing, HD_n, which is not observed. To obtain an estimating equation, rewrite (2.27) with HD_n expressed in terms of price and expenditure,

$$p_n = \tilde{\Pi}_n \left(\sum_i \eta_{in} e_{in} l_{in} \right)^{\chi_n},$$

where $\chi_n = \frac{\gamma_n}{1+\gamma_n}$ and $\tilde{\Pi}_n = \Pi_n^{\frac{1}{1+\gamma_n}}$. Taking logs and differencing over time yields an equation linear in χ_n

$$\Delta \log p_n = \Delta \log \tilde{\Pi}_n + \chi_n \Delta \log \left(\sum_i \eta_{in} e_{in} l_{in} \right). \tag{2.33}$$

[21] An exception is Diamond (2016), who estimates an elasticity close to 1.6 in line with the time-series results.

Following Saiz (2010), we parameterize χ_n as a function of geographical and regulatory constraints, $\chi_n = \chi + \chi_L UNAVAL_n + \chi_R WRLURI_n$. $UNAVAL_n$ is a measure of geographic constraints from Saiz (2010) and $WRLURI_n$ is the Wharton Residential Land Use Regulation Index developed by Gyourko, Saiz, and Summers (2008). Substituting the expression χ_n into (2.33) yields an estimating equation for χ, χ_L and χ_R,

$$\Delta \log p_n = \Delta \log \tilde{\Pi}_n + (\chi + \chi_L UNAVAL_n + \chi_R WRLURI_n)\, \Delta \log \left(\sum_i \eta_{in} e_{in} l_{in} \right).$$

with η_{in} constructed using (2.3). We take the long difference of all variables between 1980 and 2010. Because rents and employment are endogenous to unobserved housing supply shocks, we follow Diamond (2016) and use Bartik shocks, as well as their interactions with $UNAVAL_n$ and $WRLURI_n$, as instruments. We set $\gamma_n = \frac{\chi_n}{1-\chi_n}$. The employment-weighted average of the γ_n obtained using this procedure is 0.63, comparable to the value of 0.77 reported by Saiz (2010).

Migration elasticity

The migration elasticity is calibrated by indirect inference as in Greaney (2020). We match the long-run elasticity of employment to nominal wages reported by Hornbeck and Moretti (2019), who use local TFP estimated from plant-level data as an instrument for labor-demand shocks. They report an elasticity of 2.76. We solve for the value of θ that produces the same response in our model, implying $\theta = 5.11$.[22]

2.3.3 Fundamentals

We now turn to the fundamentals of the quantitative model: the location-specific productivity, amenity and housing-supply shifters $(a_n^t, z_n^t, B_{un}^t, B_{sn}^t, \Pi_n^t)$ and the ag-

[22]Because we use a NHCES utility function rather than the usual Cobb-Douglas specification, the value of this parameter is not comparable to other values reported in the literature.

gregate productivity parameters A^t and Z^t. Note we have added a time superscript because we allow all fundamentals to vary by year. We obtain these fundamentals for each year $t \in \{1980, 1990, 2000, 2010\}$ by inverting the model so that it exactly matches Census data on wages and employment by skill and MSA and the MSA price indices.[23]

2.4 The Skill Premium and Sorting

How did the increase in the aggregate skill premium alter the spatial distribution of skill 1980-2010? To answer this question, we perform a counterfactual experiment using the quantitative model developed in Section 2.2 and calibrated in Section 2.3. For each census year between 1980 and 2010, we solve for the spatial distributions of skilled and unskilled workers that would have occurred had the aggregate skill premium remained constant at its 1980 level. In the implementation, we choose values for aggregate productivity Z^t and aggregate skill bias A^t such that (i) average skilled wages grow at the same rate as average unskilled wages, and (ii) average unskilled wages grow at the same rate as they did in the data. All location-specific fundamentals — productivities, skill biases, amenities, and housing-supply shifters — evolve as they did in the data. Because only Z^t and A^t are changed, the difference between the data and the model represents a *location-neutral* shock. This difference identifies the causal effect of the rising aggregate skill premium.

2.4.1 Main Results

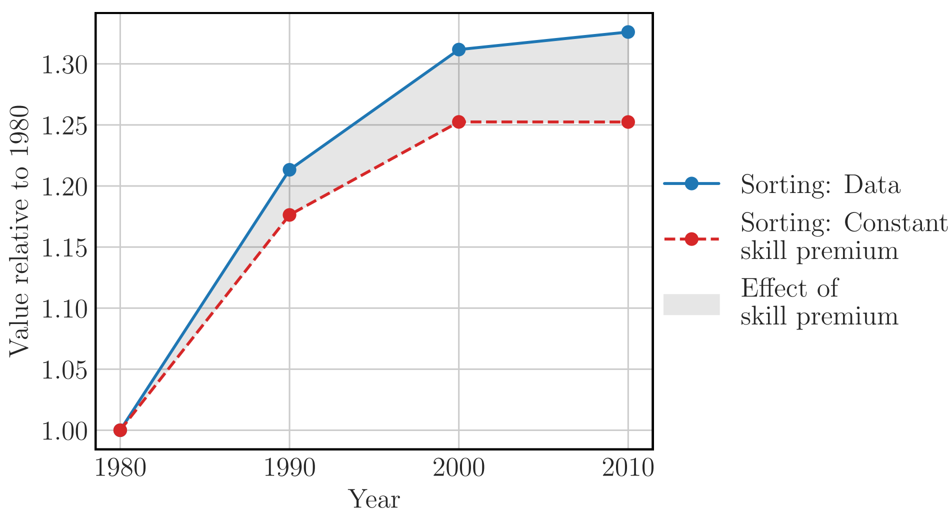
Figure 2.3 shows our main result. In the absence of a rising national skill premium, sorting only rises by 25.2%, whereas in the data it rose by 32.6%. We conclude that without the increase in the skill premium, sorting would have risen by 7.4 percentage

[23]In each year, A^t and Z^t are not separately identified from the scale of a_n^t and z_n^t. We normalize the mean of a_n^t and z_n^t to each be 1 in every year, which has no impact on our results.

points less — 23% of the overall increase between 1980 and 2010. The shaded area in Figure 2.3 shows this difference. Our model attributes the remaining 77% to idiosyncratic amenity, productivity, and housing-supply shocks such as those highlighted by Diamond (2016) or Ganong and Shoag (2017).

Figure 2.4 explores the model mechanism. In the simple model of Section 2.2, the key driver of sorting was the productivity-adjusted cost of housing, c_n. There, we showed that increases in the skill premium make skilled households less sensitive to housing costs, thereby pushing them toward high c_n locations. Because these locations are already relatively skill intensive, sorting rises. Figure 2.4 shows each of these steps in the quantitative model. Panel (a) plots the causal effect of the higher skill premium in 2010 against the productivity-adjusted housing cost in 2010. As in the simple model, the higher skill premium encouraged skilled workers to move toward high cost locations. Panel (b) translates the skill-housing cost relationship into a statement about sorting by plotting the causal effect of the higher skill premium against city-

Figure 2.3: Sorting since 1980



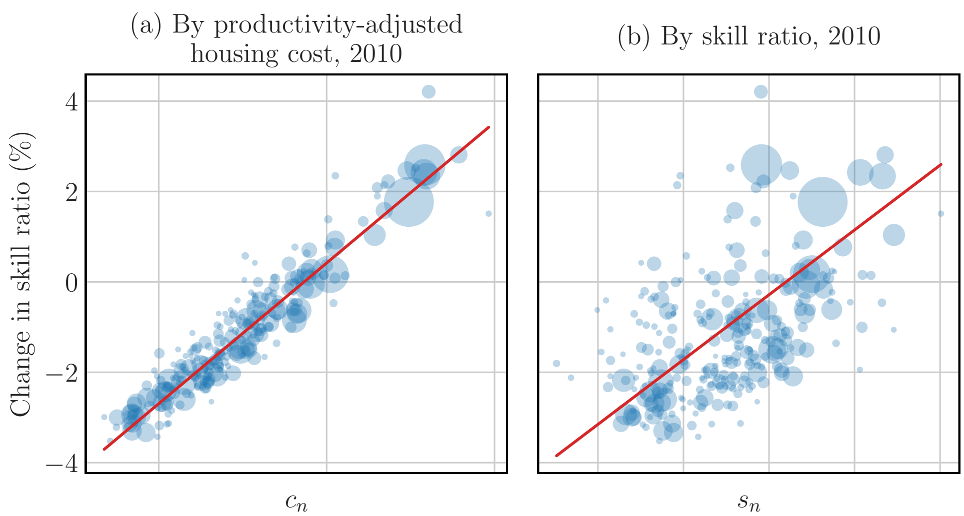
Note: "Data" calculates sorting using Census data on employment by education level, 1980-2010. "Constant skill premium" shows the level of sorting in an economy with the same fundamentals as in the data, except that aggregate productivities are changed to eliminate the increase in the skill premium. Shaded area shows the causal effect of the rising skill premium. Sorting is defined as the variance of the log skill ratio across cities, weighted by 1980 employment.

level skill ratios in 2010. The positive slope of the regression line shows skill-intensive locations generally gained skilled workers as a result of the rising skill premium, and so sorting increased. The relationship in (b) is noisy because in the quantitative model, c_n is only one determinant of sorting, alongside wages and amenities.

The Role of Preferences

We have emphasized the importance of estimating preferences, rather than assuming an extreme case. We now investigate how assuming different preferences would change our results. We repeat the main counterfactual experiment for different values of the income elasticity ϵ and the price elasticity σ. For each value of ϵ and σ we hold other parameters constant at the values in Table 2.3, invert the model again using 1980-2010 data on wages, employment and prices as in Subsection 2.3.3 and then feed the model the change in skill-neutral productivity Z and skill-biased productivity A

Figure 2.4: City Level Effects of the Skill Premium



Note: Panels (a) and (b) plot the causal effect of the increase in the skill premium in each MSA in 2010 against the productivity-adjusted housing cost in 2010 and log skill ratio in 2010, respectively. The causal effect of the skill premium is defined as the difference between the economy with all fundamentals changing as in the data, and the economy with the same fundamentals except for aggregate productivities Z and A, which are changed to eliminate the increase in the skill premium 1980-2010. The dots are sized proportionally to 1980 employment. The solid line is the regression line.

obtained above. The outcome of interest is the fraction of the increase in sorting 1980-2010 caused by the location neutral shock to A.

Figure 2.5 plots the results of this exercise for $0 \leq \sigma \leq 1$ and $\epsilon \geq \sigma - 1$.[24] The color of the figure shows the effect of the rising skill premium at each value of ϵ and σ, with a darker shade indicating a larger effect. As ϵ falls towards -1 and housing becomes more of a necessity, the effect of the skill premium rises, reaching a maximum of about 48% of the increase in sorting observed in the data. In contrast, we can see that varying σ — i.e moving horizontally — does not have a large impact on the relationship between the skill premium and sorting. Since our theory emphasizes the role of nonhomotheticity in generating sorting, this pattern makes intuitive sense.

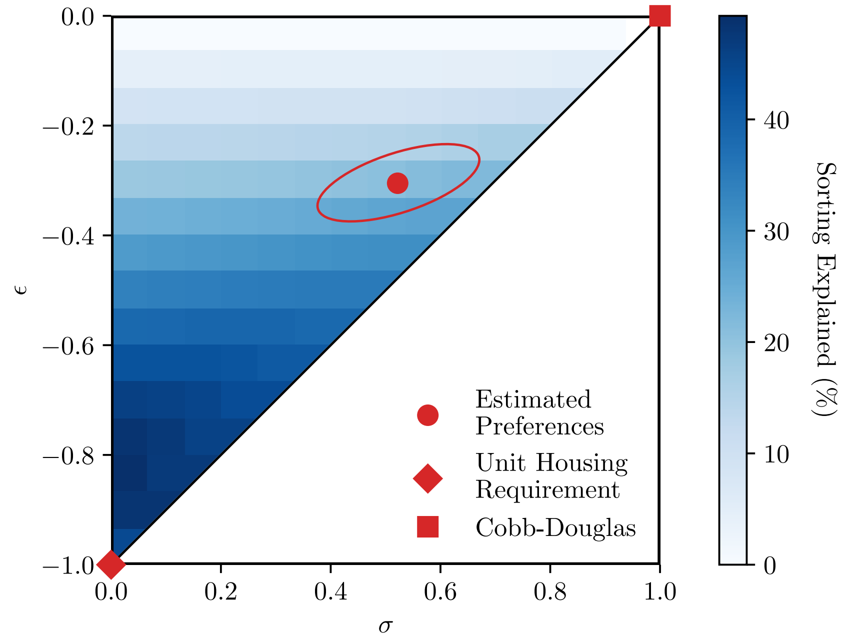
The three markers in Figure 2.5 show important special cases of our NHCES preferences. The square at $\sigma = 1$ and $\epsilon = 0$ shows Cobb-Douglas preferences, where the rising skill premium explains none of the observed increase in the skill premium. The diamond at $\sigma = 0$ and $\epsilon = -1$ shows results under a unit housing requirement. For these preferences the share explained rises to 45%. Finally the circle shows our estimated preferences, at which the rising skill premium explains 23% of the observed increase in sorting. The ellipse around this point shows a 95% confidence set for ϵ and σ. Within this set the share explained ranges from 17% to 28% . Figure 2.5 thus shows that our estimated preferences produce results which are qualitatively different than those obtained under Cobb-Douglas preferences, and quantitatively far from those obtained under a unit housing requirement.

2.4.2 Extensions and Robustness

In Appendix B.5 we run a number of robustness checks; we discuss the results briefly here. We consider alternative measure of sorting — the Theil index, the dissimilarity index, and the 90/10 ratio of the skill ratio distribution — and find similar

[24]Recall from Section 2.1 that the second inequality is needed for the preferences to be well-defined.

Figure 2.5: The Role of Preferences



Note: Share of the increase in sorting observed in the data attributable to the skill premium, as a function of the income and price elasticities ϵ and σ, respectively. Lower left-hand corner corresponds to a unit housing requirement, and upper right-hand corner to Cobb-Douglas. Estimated preferences are represented by the red circle, with 95% confidence ellipse.

results to those obtained using our baseline measure. We reimplement the counterfactual so that the aggregate terms (A^t, Z^t) are chosen to match the growth of average wages, rather than the growth of unskilled wages alone, and find very similar results. We experiment with an alternative specification of nonhomothetic preferences by recalibrating the model to Price Independent Generalized Linear (PIGL) preferences. This change has little effect on out main results.

We consider a lower value of ρ equal to 1.6, taken from Acemoglu and Autor (2011). Using this value of ρ, the share of the increase in sorting explained by the rising skill premium falls to 13.9%. The fact that the share explained falls is intuitive — when skilled and unskilled labor are not close substitutes, the influx of skilled workers into expensive cities is dampened by falling skilled wages. Our quantitative results are therefore fairly sensitive to changes in this parameter. Nevertheless, as we argued in Section 2.3, city-level estimates of the elasticity of substitution are consistently higher than the aggregate time-series estimates, and our baseline value of $\rho = 3.85$ is around the middle of such estimates.

Finally, we consider endogenous amenities. Diamond (2016) allows amenities to respond endogenously to the local skill mix and concludes skilled workers value amenities more, causing sorting. In Appendix B.5.5, we extend our model to incorporate endogenous amenities; here, we summarize the key points. First, in the simple model of Section 2.2, endogenous amenities amplify the effect of the skill premium on spatial sorting in the presence of nonhomothetic housing demand. An increase in the skill premium makes skilled households less sensitive to housing costs and encourages movement toward more expensive cities, just as in our baseline model. Then, amenities endogenously rise in expensive cities, encouraging further skilled in-migration. Second, we extend the neutrality result in Proposition 2 to endogenous amenities. If housing demand is homothetic, changes in the skill premium continue to have no effect on sorting. In summary, endogenous amenities are likely to amplify the effect of the skill premium on spatial sorting when preferences are nonhomothetic, but they do not create an independent link between the skill premium and spatial sorting.

2.5 Real inequality

In this section we apply our estimated preferences to measure the evolution of inequality in real consumption over space and time. Recent work by Baum-Snow and Pavan (2013), Autor (2019), Giannone (2019), Eckert (2019), and Eckert, Ganapati, and Walsh (2020) has shown the skill premium grew fastest in large cities after 1980, part of what Eckert, Ganapati, and Walsh (2020) label the "new urban bias" of economic growth. We go beyond nominal earnings and estimate welfare gaps between skilled and unskilled households across space (see also Moretti 2013; Albouy, Ehrlich, and Liu 2016, for efforts to reconcile real and nominal aggregate inequality). We find the real skill premium is even more urban-biased than implied by nominal figures, because unskilled households are disproportionately hurt by rising housing costs in large cities.

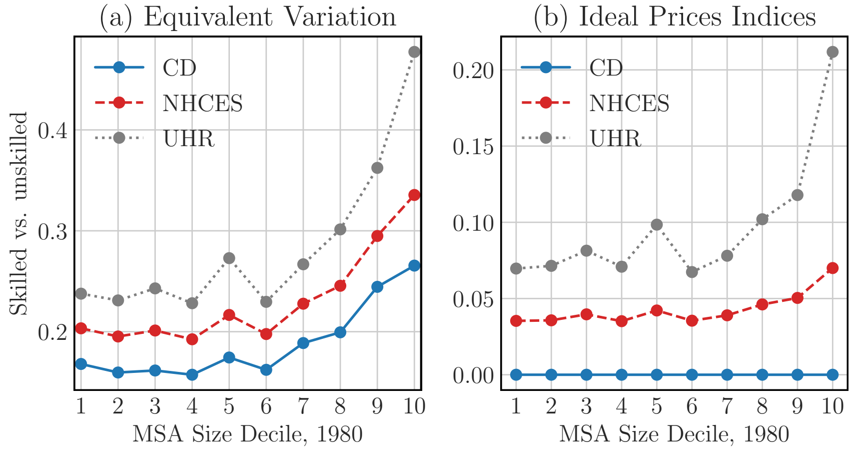
We consider changes in expenditure and housing costs between 1980 and 2010.[25] For each skill group i and MSA n, we calculate the equivalent variation EV_{in}: how much would a household of type i in MSA n need to be paid to make them indifferent between receiving this payment and facing 1980 prices and expenditure levels, versus facing 2010 prices and expenditure levels. We express this quantity as a fraction of 1980 expenditure. Our measure of the change in the real skill premium in MSA n is the log difference between EV_{sn} and EV_{un}.

Figure 2.6 displays changes in the real skill premium across the city-size distribution. We bin MSAs into ten deciles by 1980 employment, weighted so that each bin is one tenth of national employment. The solid line assumes that real housing costs in every MSA were constant 1980-2010. In this case, EV_{in} for each group is exactly equal to expenditure growth. The difference in equivalent variation is positive in every decile because the skill premium rose everywhere. Consistent with the literature mentioned above, the gap is particularly large in the largest MSAs, where the nominal skill premium grew by roughly 26%.

The dashed and dotted lines show the real skill premium under the assumptions of NHCES preferences and a unit housing requirement, respectively. Accounting for housing costs raises the skill premium everywhere, but disproportionately so in large cities, as shown in the right panel of Figure 2.6. NHCES preferences (dashed line) imply a gap of 3 log points in the smallest cities compared to 7 in the largest. By contrast, a unit housing requirement (dotted line) inflates the skill premium by 7 log points at the bottom decile and 21 at the top, a pattern stemming from excessive differences in implied price sensitivity between skill groups. Disciplining preferences

[25]In our framework, this is identical to after-tax income.

Figure 2.6: Real Inequality by Size Decile



Notes: Panel (a): 'CD' plots the difference in average expenditure growth 1980-2010 between skilled and unskilled households in each size decile — as described in the text, this coincides with the difference in equivalent variation across skill groups when preferences are Cobb-Douglas. Expenditure is equal to after-tax income, which is calculated using the tax function (2.29) and Census data on wages. 'NHCES' adjusts expenditure for changes in ideal price indices prices. 'UHR' performs the same adjustment under the assumption that $\epsilon = -1$ and $\sigma = 0$. Panel (b) shows the contribution of the ideal price indices to each of the three cases in Panel (a).

with data is key when attempting to measure the effects of income and price changes.[26]

The rise of inequality in large cities—locations with high housing prices and high shares of skilled households—is unsurprising in light of the findings in Section 2.4. Sorting increased, at least in part, because skilled households were willing to move to expensive locations. Rising incomes made their welfare, hence their choice of where to live, less responsive to housing costs. In this sense we can interpret sorting as a barometer of welfare inequality. All else equal, sorting is high when real returns to skill are very different across space.

2.6 Conclusion

When housing demand is income inelastic, the skill premium causes spatial sorting because skilled and unskilled households face different ideal price indices in the same

[26] Albouy, Ehrlich, and Liu (2016) conclude that aggregate real inequality is higher than nominal inequality under estimated nonhomothetic preferences, but they do not study sorting or space.

location. Skilled households have low housing shares and are insensitive to high house prices. The opposite is true for unskilled households. The growth in the skill premium since 1980 has amplified the cost-of-living wedge, causing skilled households to flock to expensive cities and unskilled households to flee to cheap cities. Our model attributes about one quarter of the observed increase in spatial sorting to the growth of the skill premium. While our paper is not the first to explore this housing demand channel, our contribution is to justify its relevance at the level of cities, to connect it explicitly to the income elasticity, and to gauge its magnitude in a quantitative spatial model.

Bibliography

Acemoglu, Daron, and David Autor. 2011. "Skills, tasks and technologies: Implications for employment and earnings". In *Handbook of Labor Economics*, 4:1043–1171.

Aguiar, Mark, and Mark Bils. 2015. "Has consumption inequality mirrored income inequality?" *American Economic Review* 105 (9): 2725–56.

Albouy, David. 2016. "What are cities worth? Land rents, local productivity, and the total value of amenities". *Review of Economics and Statistics* 98 (3): 477–487.

Albouy, David, Gabriel Ehrlich, and Yingyi Liu. 2016. "Housing demand, cost-of-living inequality, and the affordability crisis". Working Paper.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the topography of the spatial economy". *The Quarterly Journal of Economics* 129 (3): 1085–1140.

Amiti, Mary, and David E. Weinstein. 2011. "Exports and Financial Shocks*". *The Quarterly Journal of Economics* 126, no. 4 (): 1841–1877.

Andreski, Patricia, Geng Li, Mehmet Zahid Samancioglu, and Robert Schoeni. 2014. "Estimates of annual consumption expenditures and its major components in the PSID in comparison to the CE". *American Economic Review* 104 (5): 132–35.

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Dynamic Inputs and Resource (Mis)Allocation". *Journal of Political Economy* 122:1013–1063.

Atkeson, Andrew, and Ariel Tomás Burstein. 2010. "Innovation, Firm Dynamics, and International Trade". *Journal of Political Economy* 118 (3): 433–484.

Autor, David H. 2019. "Work of the past, work of the future". *AEA Papers and Proceedings* 109 (): 1–32.

Aw, Bee Yan, Sukkyun Chung, and Mark Roberts. 2000. "Productivity and Turnover in the Export Market: Micro-level Evidence from the Republic of Korea and Taiwan (China)". *World Bank Economic Review* 14 (1): 65–90.

Bai, Yan, Keyu Jin, and Dan Lu. 2019. *Misallocation under trade liberalization.* Tech. rep. National Bureau of Economic Research.

Banerjee, Abhijit V., and Esther Duflo. 2014. "Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program". *The Review of Economic Studies* 81, no. 2 (): 572–607.

Baum-Snow, Nathaniel, Matthew Freedman, and Ronni Pavan. 2018. "Why has urban inequality increased?" *American Economic Journal: Applied Economics* 10 (4): 1–42.

Baum-Snow, Nathaniel, and Ronni Pavan. 2013. "Inequality and city size". *Review of Economics and Statistics* 95 (5): 1535–1548.

Beaudry, Paul, Mark Doms, and Ethan Lewis. 2010. "Should the personal computer be considered a technological revolution? Evidence from US metropolitan areas". *Journal of Political Economy* 118 (5): 988–1036.

Bernard, Andrew, and J. Jensen. 1999. "Exceptional exporter performance: cause, effect, or both?" *Journal of International Economics* 47 (1): 1–25.

Berry, Christopher R, and Edward L Glaeser. 2005. "The divergence of human capital levels across cities". *Papers in Regional Science* 84 (3): 407–444.

Berthou, Antoine, John Jong-Hyun Chung, Kalina Manova, and Charlotte Sandoz Dit Bragard. 2019. "Trade, productivity and (mis) allocation". *Available at SSRN 3502471*.

Bound, John, Jeffrey Groen, Gabor Kezdi, and Sarah Turner. 2004. "Trade in university training: cross-state variation in the production and stock of college-educated labor". *Journal of Econometrics* 121 (1-2): 143–173.

Brooks, Wyatt, and Alessandro Dovis. 2020. "Credit market frictions and trade liberalizations". *Journal of Monetary Economics* 111:32–47. ISSN: 0304-3932.

Buera, Francisco J, Joseph P Kaboski, and Yongseok Shin. 2011. "Finance and development: A tale of two sectors". *American economic review* 101 (5): 1964–2002.

— . 2021. "The macroeconomics of microfinance". *The Review of Economic Studies* 88 (1): 126–161.

Bureau of Economic Analysis. 2020. *Metropolitan Area Regional Price Parities*. Tech. rep. `https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=30&isuri=1&major_area=5&area=xx&year=-1&tableid=104&category=8101&area_type=0&year_end=-1&classification=non-industry&state=5&statistic=1&yearbegin=-1&unit_of_measure=levels`, last accessed 06/01/21. Washington, DC.

Bureau of Labor Statistics. 2020a. *Consumer Expenditure Survey Public Use Microdata*. Tech. rep. `bls.gov/cex/pumd.htm`, last accessed 12/01/20. Washington, DC.

— . 2020b. *Consumer Price Index for All Urban Consumers: All Items Less Shelter in U.S. City Average*. Tech. rep. Retrieved from FRED, Federal Reserve Bank of St. Louis, `https://fred.stlouisfed.org/series/CUSR0000SA0L2`, last accessed 06/01/21.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs". *Econometrica* 82 (6): 2295–2326.

Calonico, Sebastian, Matias D Cattaneo, Max H Farrell, and Rocio Titiunik. 2019. "Regression discontinuity designs using covariates". *Review of Economics and Statistics* 101 (3): 442–451.

Card, David. 2009. "Immigration and inequality". *American Economic Review* 99 (2): 1–21.

Chaney, Thomas. 2016. "Liquidity constrained exporters". *Journal of Economic Dynamics and Control* 72 (C): 141–154.

Clerides, Sofronis K., Saul Lach, and James R. Tybout. 1998. "Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco". *The Quarterly Journal of Economics* 113 (3): 903–947. ISSN: 00335533, 15314650.

Colas, Mark, and Kevin Hutchinson. 2021. "Heterogeneous workers and federal income taxes in a spatial equilibrium". *American Economic Journal: Economic Policy* 13 (2): 100–134.

Comin, Diego A, Danial Lashkari, and Martí Mestieri. 2021. "Structural change with long-run income and price effects". *Econometrica* 89 (1): 311–374.

Cooper, Russell W., and John C. Haltiwanger. 2006. "On the Nature of Capital Adjustment Costs". *The Review of Economic Studies* 73, no. 3 (): 611–633.

Couture, Victor, Cecile Gaubert, Jessie Handbury, and Erik Hurst. 2019. "Income growth and the distributional effects of urban spatial sorting". Working Paper.

Davis, Morris A, and François Ortalo-Magné. 2011. "Household expenditures, wages, rents". *Review of Economic Dynamics* 14 (2): 248–261.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. 2016. "Prices, Markups, and Trade Reform". *Econometrica* 84 (2): 445–510.

Diamond, Rebecca. 2016. "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000". *American Economic Review* 106 (3): 479–524.

Eckert, Fabian. 2019. "Growing apart: Tradable services and the fragmentation of the US economy". Working Paper.

Eckert, Fabian, Sharat Ganapati, and Conor Walsh. 2020. "Skilled scalable services: the new urban bias in recent economic growth". Working Paper.

Eckert, Fabian, and Michael Peters. 2018. "Spatial structural change". Working Paper.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. 2014. "Spatial sorting". *Journal of Political Economy* 122 (3): 554–620.

Fajgelbaum, Pablo D, and Cecile Gaubert. 2020. "Optimal spatial policies, geography, and sorting". *The Quarterly Journal of Economics* 135 (2): 959–1036.

Fogli, Alessandra, and Veronica Guerrieri. 2019. "The end of the American Dream? Inequality and segregation in US cities". Working Paper.

Ganong, Peter, and Daniel Shoag. 2017. "Why has regional income convergence in the US declined?" *Journal of Urban Economics* 102:76–90.

Giannone, Elisa. 2019. "Skilled-biased technical change and regional convergence". Working Paper.

Greaney, Brian. 2020. "The distributional effects of uneven regional growth". Working Paper.

Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. "Superstar cities". *American Economic Journal: Economic Policy* 5 (4): 167–99.

Gyourko, Joseph, Albert Saiz, and Anita Summers. 2008. "A new measure of the local regulatory environment for housing markets: the Wharton Residential Land Use Regulatory Index". *Urban Studies* 45 (3): 693–729.

Handbury, Jessie. 2019. "Are poor cities cheap for everyone? Non-homotheticity and the cost of living across u.s cities". Working Paper.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante. 2017. "Optimal tax progressivity: an analytical framework". *The Quarterly Journal of Economics* 132 (4): 1693–1754.

Hornbeck, Richard, and Enrico Moretti. 2019. "Who benefits from productivity growth? Direct and indirect effects of local TFP growth on wages, rents, and inequality". Working Paper.

Hsieh, Chang-Tai, and Peter J Klenow. 2009. "Misallocation and manufacturing TFP in China and India". *The Quarterly journal of economics* 124 (4): 1403–1448.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing constraints and spatial misallocation". *American Economic Journal: Macroeconomics* 11 (2): 1–39.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice". *Journal of Econometrics* 142 (2): 615–635.

Itskhoki, Oleg, and Benjamin Moll. 2019. "Optimal development policies with financial frictions". *Econometrica* 87 (1): 139–173.

Kaboski, Joseph P, and Robert M Townsend. 2011. "A structural evaluation of a large-scale quasi-experimental microfinance initiative". *Econometrica* 79 (5): 1357–1406.

Kapoor, Mudit, Priya Ranjan, and Jibonayan Raychaudhuri. 2017. "The impact of credit constraints on exporting firms: Evidence from the provision and subsequent removal of subsidised credit". *The World Economy* 40 (12): 2854–2874.

Katz, Lawrence F, and Kevin M Murphy. 1992. "Changes in relative wages, 1963–1987: supply and demand factors". *The Quarterly Journal of Economics* 107 (1): 35–78.

Kohn, David, Fernando Leibovici, and Michal Szkup. 2014. *No Credit, No Gain: Trade Liberalization Dynamics, Production Inputs and Financial Development.* Tech. rep. 2.

Lee, David S. 2008. "Randomized experiments from non-random selection in U.S. House elections". The regression discontinuity design: Theory and applications, *Journal of Econometrics* 142 (2): 675–697. ISSN: 0304-4076.

Leibovici, Fernando. 2021. "Financial development and international trade". *Journal of Political Economy* 129 (12): 000–000.

Lewbel, Arthur. 1996. “Demand estimation with expenditure measurement errors on the left and right hand side”. *The Review of Economics and Statistics*: 718–725.

Manova, Kalina. 2012. “Credit Constraints, Heterogeneous Firms, and International Trade”. *The Review of Economic Studies* 80, no. 2 (): 711–744.

McCrary, Justin. 2008. “Manipulation of the running variable in the regression discontinuity design: A density test”. *Journal of econometrics* 142 (2): 698–714.

Melitz, Marc J. 2003. “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”. *Econometrica* 71 (6): 1695–1725.

Midrigan, Virgiliu, and Daniel Yi Xu. 2014. “Finance and Misallocation: Evidence from Plant-Level Data”. *American Economic Review* 104, no. 2 (): 422–58.

Moll, Benjamin. 2014. “Productivity losses from financial frictions: Can self-financing undo capital misallocation?” *American Economic Review* 104 (10): 3186–3221.

Moretti, Enrico. 2013. “Real wage inequality”. *American Economic Journal: Applied Economics* 5, no. 1 (): 65–103.

Notowidigdo, Matthew J. 2020. “The incidence of local labor demand shocks”. *Journal of Labor Economics* 38 (3): 687–725.

Paravisini, Daniel, Veronica Rappoport, Philipp Schnabl, and Daniel Wolfenzon. 2015. “Dissecting the Effect of Credit Supply on Trade: Evidence from Matched Credit-Export Data”. *The Review of Economic Studies* 82 (1 (290)): 333–359.

Redding, Stephen J. 2016. “Goods trade, factor mobility and welfare”. *Journal of International Economics* 101:148–167.

Rotemberg, Martin. 2019. “Equilibrium effects of firm subsidies”. *American Economic Review* 109 (10): 3475–3513.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2020. *IPUMS USA: Version 10.0*. `usa.ipums.org/usa`, last accessed 12/01/20, Minneapolis, MN.

Saiz, Albert. 2010. “The geographic determinants of housing supply”. *The Quarterly Journal of Economics* 125 (3): 1253–1296.

Schmidheiny, Kurt. 2006. “Income segregation from local income taxation when households differ in both preferences and incomes”. *Regional Science and Urban Economics* 36 (2): 270–299.

University of Michigan Institute for Social Research. 2021. *Panel Study of Income Dynamics*. Tech. rep. `psidonline.isr.umich.edu`, last accessed 05/01/21. Ann Arbor, MI.

Zia, Bilal H. 2008. “Export incentives, financial constraints, and the (mis) allocation of credit: Micro-level evidence from subsidized export loans”. *Journal of financial economics* 87 (2): 498–527.

Contents

A Appendix to: Exporters, Credit Constraints, and Misallocation **102**

A.1 Empirical Appendix 102

A.2 Theory Appendix 108

A.3 Estimation Appendix 109

A.4 A Model with Exogenous Misallocation 112

B Appendix to: Housing Demand, Inequality, and Spatial Sorting [1] **116**

B.1 Data 116

B.2 Estimation 120

B.3 Theory 127

B.4 Calibration 134

B.5 Counterfactual 137

[1] This is joint work with Trevor Williams

Appendix A

Appendix to: Exporters, Credit Constraints, and Misallocation

A.1 Empirical Appendix

Data

I use data from the Prowess dataset compiled by CMIE. Specifically, I consider firms in the 'manufacturing superset', as defined by CMIE. Among these firms, I drop any which have missing information on total borrowing, wage bills or sales. I also drop any missing plant and machinery in 2007, the baseline year for my regression discontinuity design. I also drop any firms which have very large growth rates in absolute value for any of these variables; in practice I drop firms for which these growth rates are in the top or bottom 1% across all firms.

Firms in Prowess differ in the dates on which they report financial statements. I adopt the following convention: if a report is made in the first six months of year t, I date it to year $t-1$ on the grounds that most of the production the report refers to took place in $t-1$. Firms also differ in the time span their financial statements cover: although most report information covering 12 months, a few report information for shorter timespans. Where this is the case, I rescale the flow variables (wage bills and sales) to a yearly frequency.

Regression Discontinuity Design Details

I implement the RDD estimation using the `rdrobust` Stata package created by **cct_stata**. In particular, in each regression I use the MSE optimal one-sided bandwidth, a triangular kernel, and a first-order local polynomial. Standard errors are based on plug-in residuals, and are clustered at the firm level. In Table 1.5 I report the 'conventional' point estimates and standard errors, i.e.,without the robust bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). However I also report results with this correction in Column (4) of Table A.2.

Borrowing Costs

In Table A.1, I investigate whether PSL eligibility lowered firms' borrowing costs. Here I replicate the specification of Columns (3) and (4) of Table 1.2, but use as the

Table A.1: Effect of Priority Sector Lending on Borrowing Costs

	(1) *Exporters*	(2) *Non-exporters*
	0.094 (0.073)	−0.006 (0.083)
Years	2008-12	2008-12
N	10145	6857

Source: Prowess Dataset, all manufacturing firms, 2005-2012

Note: Each column reports the estimated discontinuity in log borrowing costs at plant and machinery equal to 50 million rupees in 2007. The outcome is measured between 2008 and 2012. All regressions control for year, industry and firm age fixed effects. Additionally I control for 2005 loans, sales and employment, and 2005 log borrowing costs. (1) shows results for exporters and (2) shows results for non-exporters. Note that a positive number implies that firms eligible for PSL — i.e those to the left of 50 million rupees — faced higher borrowing costs.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

outcome (log) borrowing costs. Prowess measures this as total expenses on financial services — i.e interest payments and fees — divided by total borrowing. Column (1) shows results for exporters and (2) for non-exporters. The positive number in Column (1) implies that if anything, exporters eligible for PSL faced higher borrowing costs relative to those that were ineligible. The discontinuity for non-exporters in Column (2) is much smaller and has the opposite sign. Neither estimate is statistically significant. I conclude that PSL did not change eligible firms' borrowing costs; in particular, lower borrowing costs cannot explain the positive effects of PSL eligibility on exporters in Table 1.2.

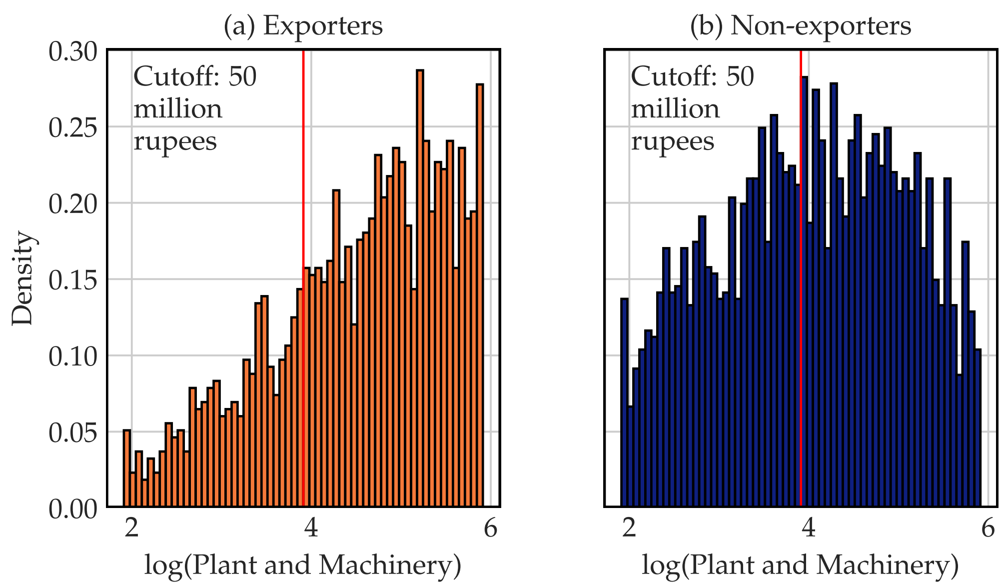
Bunching Checks

Figure A.1 plots the density of (log) plant and machinery in 2007 for firms close to the cutoff. Panel (a) shows exporters and (b) shows non-exporters. Visual inspection shows that in neither figure is there a mass of firms just to the left of the cutoff, as one would expect if firms were strategically choosing their plant and machinery in order to become or remain eligible for PSL. To formalize this observation, I use the test statistic of **cattaneo_density**. I find values of 1.5729 (p-value = 0.1157) for exporters and 1.0793 (p-value = 0.2804) for non-exporters. The positive values indicate that if anything there are slightly too few firms to the left of the cutoff — the opposite of what we would expect if firms were choosing their plant and machinery to become eligible for PSL. Neither estimate is significantly different from zero, therefore I do not reject the null hypothesis of no bunching for either group.

Robustness Checks

Table A.2 shows the results of a number of robustness checks. The baseline these results should be compared against is Columns (3) and (4) of Table 1.2. Each column shows results for loans, employment and sales among exporters in Panel (a) and non-exporters in Panel (b). The first three columns assess the sensitivity of my results to technical aspects of the regression discontinuity design specification. Column (1) halves the optimally chosen bandwidth and Column (2) doubles it. Although the point estimates move around a bit, qualitatively my results do not seem overly sensitive to the choice of bandwidth. Column (3) performs the 'donut hole' check suggested by **cattaneo_practical**, in which I drop the 5% of observations closest to the 50 million rupee cutoff. The idea here is that these are the observations most susceptible to manipulation, and so (3) acts as a check of how sensitive my results are to manipulation. Dropping these observations does not make a dramatic difference to my results. Column (4) applies the bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014); this does not change my point estimates much, although because the bias must be estimated it inflates the standard errors. Column (5) shows that my results are not sensitive to the choice of years I use for measuring

Figure A.1: Bunching Checks



Source: Prowess dataset, manufacturing firms, 2007. *Notes*: Each panel shows a histogram of log plant and machinery for firms close to the 50 million rupee cutoff for PSL eligibility. Panel (a) shows exporters and Panel (b) shows non-exporters. In neither plot is there an obvious mass to the left of the cutoff, i.e there is no visual evidence of bunching.

outcomes; I drop the first and last years and continue to find very similar results. Finally Column (6) drops all controls other than year fixed effects. The results for exporters are noisier and not statistically significant, although qualitatively similar. For non-exporters the point estimates are somewhat larger than my baseline results, but still not statistically significant.

Table A.2: Robustness Checks

	(1) *Narrow Bw.*	(2) *Wide Bw.*	(3) *'Donut Hole'*	(4) *CCT*	(5) *Short Times-pan*	(6) *No Con-trols*
(a) Exporters						
Loans	0.391*	0.275**	0.380**	0.331*	0.319*	0.259
	(0.208)	(0.114)	(0.175)	(0.182)	(0.181)	(0.184)
Employment	0.375***	0.161*	0.271*	0.282**	0.320***	0.172
	(0.136)	(0.087)	(0.164)	(0.135)	(0.119)	(0.151)
Sales	0.301**	0.177*	0.323**	0.227	0.255*	0.112
	(0.149)	(0.093)	(0.161)	(0.145)	(0.132)	(0.164)
(b) Non-Exporters						
Loans	−0.055	0.023	−0.037	0.021	0.086	0.128
	(0.155)	(0.092)	(0.139)	(0.136)	(0.132)	(0.131)
Employment	−0.013	−0.003	−0.181	0.008	0.006	0.122
	(0.097)	(0.061)	(0.127)	(0.086)	(0.087)	(0.132)
Sales	0.028	0.050	−0.038	0.142	0.126	0.193
	(0.145)	(0.089)	(0.138)	(0.132)	(0.124)	(0.160)
Years	2008-12	2008-12	2008-12	2008-12	2009-11	2008-12
Pre-policy controls	Yes	Yes	Yes	Yes	Yes	No

Source: Prowess Dataset, all manufacturing firms, 2005-2012

Note: Columns show results for different specifications; rows show results for different outcomes. Each estimate reports the discontinuity in the outcome at plant and machinery equal to 50 million rupees. Plant and machinery measured in 2007; export status defined using sales in 2007. All outcomes are measured in logs and a positive number indicates a positive effect of being eligible for Priority Sector Lending. (1) and (2) vary the bandwidth used in estimation; (3) drops the 5% of observations closest to the cutoff; (4) applies the bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014); (5) measures outcomes in a shorter time window; and (6) drops all controls except year fixed effects.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

A.2 Theory Appendix

Proof. Proof of Theorem 1

Let θ^0 and ϑ^1 denote two different values of ϑ, the productivity-fixed cost elasticity. Let μ_f^0 and μ_f^1 denote corresponding levels of the export fixed cost. Suppose $\vartheta^1 < \vartheta^0 \leq 0$ and suppose μ_f^1 is chosen so that the probability a firm exports does not change. Define

$$\Omega_x^0 = \{(z,a)|\Delta\pi(z,a) \geq \exp(\mu_f^0)z^{\vartheta^0}\},$$

$$\Omega_x^1 = \{(z,a)|\Delta\pi(z,a) \geq \exp(\mu_f^1)z^{\vartheta^1}\},$$

as the sets of exporters under these two parameterizations. Finally let the sets of 'entering' and 'exiting' exporters be

$$\text{Entering} = \overline{\Omega}_x^0 \cap \Omega_x^1$$

$$\text{Exiting} = \Omega_x^0 \cap \overline{\Omega}_x^1$$

where the bars denote complements. Suppose a firm with state (z,a) belongs to 'Exiting'. Then from the definitions of each of the sets Ω_x^0 and Ω_x^1,

$$\log z \leq \log \bar{z} \equiv (\vartheta^0 - \vartheta^1)^{-1} \left(\mu_f^0 - \mu_f^1\right).$$

The same reasoning implies that for any (z,a) in 'Entering'

$$\log z \geq \log \bar{z}.$$

Thus, every firm which starts exporting has a higher productivity than every firm which ceases exporting.

Now, let (z', a') be the productivity and assets of an arbitrary exiting firm, and

(z'', a'') the productivity and assets of an arbitrary entering firm. We have established that $z'' \geq z'$ above. Because the entering firm did not initially export while the exiting firm did, the follow inequality must hold

$$\Delta\pi(z', a')\,(z')^{-\vartheta^0} < \Delta\pi(z'', a'')\,(z'')^{-\theta^1}\,.$$

Each side is clearly increasing in z, and $\Delta\pi(z, a)$ is increasing in a. So this inequality can only hold if $a'' < a'$, i.e entering firms must have lower assets than exiting firms.

An exporting firm is constrained if

$$\log z > \left(\frac{1}{\sigma - 1}\right) \log a + B$$

where B is a constant. I have shown above that when ϑ falls, any firm which starts exporting has higher z and lower a than any firm which stops exporting. Thus if any firm which stops exporting is constrained, every new exporter is also constrained. If any firm which starts exporting is unconstrained, every firm which stops exporting is unconstrained. Either way, the net effect must be that the share of exporters who are constrained (weakly) rises. The same logic applies to non-exporters. When ϑ falls, every new non-exporter has lower productivity and higher assets than every exiting non-exporter. Thus the share of constrained non-exporters must (weakly) fall.

□

A.3 Estimation Appendix

Targeted Moments

The estimation targets 11 moments. 6 of these are the PSL treatment effects in Column (3) and (4) of Table 1.2 and 5 are descriptive statistics calculated using the Prowess dataset. The descriptive statistics are:

(i) The standard deviation of log sales growth: I calculate this by first regressing log sales in 2006 and 2007 on industry fixed effects, to absorb variation created by industry level shocks absent from my model. I then calculate the standard deviation of the differences in the residuals from this regression. Note that I drop drop observations with very large growth rates in absolute value; in practice I drop firms with growth rates in the top or bottom 1% of the distribution of growth rates.

(ii) The autocorrelation of log sales: I follow the same procedure as in (i), but calculate the correlation between residuals in 2006 and 2007.

(iii) The fraction of exporters: I simply calculate the fraction of firms in Prowess in 2007 with positive export sales.

(iv) The difference in log sales between exporters and non-exporters: I regress log sales on industry fixed effects and an exporter dummy, again using firms from Prowess in 2007.

(v) The frequency of investment 'spikes': For each firm I calculate the change in log capital between 2006 and 2007. I define a spike as a log change greater than 0.20 in absolute value.

Estimation Procedure

As mentioned in the text, I choose a vector of parameters Ψ to minimize

$$L(\Psi) = \left(\widehat{M} - M(\Psi)\right)' W \left(\widehat{M} - M(\Psi)\right).$$

$M(\Psi)$ must be calculated by simulation.

I implement this by first solving the entrepreneur's dynamic programming problem from Section 2.2. I do this using discrete grids for z and f. In each case I use a grid with 15 points, having verified that my results are not sensitive to this choice. Instead of directly using capital k and liquid assets a as state variables, I define the state as total assets $t = k + a$ and the share of these held as capital, s. Clearly this is without loss of generality. For these endogenous state variables, I use cubic spline interpolation across discrete grid points. For total assets I use a grid which is evenly spaced in logs, and use 50 grid points. For the share of assets held as capital, I use an evenly spaced grid with 20 points.

Having solved this problem for a policy function g, I then draw a sample of N observations on the exogenous states f and z from their respective stationary distributions. Starting from an arbitrary distribution of k and a, I then simulate the policy function g, along with shocks to z, until the the joint distribution of states converges (this takes roughly 200 periods). I then calculate the descriptive moments (i)-(v) and simulate the PSL policy as described in the text. In practice I set $N = 20,000$, having verified that my results are not sensitive to this choice. I collect these moments into a vector $M(\Psi)$. To choose the value of Ψ which minimizes $L(\Psi)$, I use a global search algorithm (controlled random search with local mutation) from the Fortran implementation of NLOPT.

Untargeted Moments

In 1.6 I report values of 5 untargeted moments. The first of these is from Table 1.4, Column (1). I now describe the construction of the remaining four moments:

(i) The persistence of exporting: I calculate the probability that an exporting firm in Prowess in 2006 is still exporting in 2007, conditional on remaining active (i.e having positive overall sales).

(ii) Log sales growth of new exporters: I construct a dummy equal to 1 if a firm was active in 2006 but not an exporter, and exports in 2007. Call this

New Exporter$_{it}$. I then run the regression

$$\Delta \log R_{it} = \beta_0 + \beta_1 \textit{New Exporter}_{it} + \epsilon_{it}$$

where $\Delta \log R_{it}$ is the growth rate of sales between t and $t-1$. I use data from Prowess in 2007. Table 1.6 reports an estimate of β_1.

(iii) Differences in input demands: I run the regression

$$\log y_{its} = \gamma_s + \gamma_1 e_{its} + \epsilon_{its}$$

where y_{its} is an input — either capital or labor — and e_{its} is a dummy equal to 1 if the firm exports. s indexes sectors and γ_s is a sector fixed effect. Table 1.6 reports estimates of γ_1 for labor and capital.

A.4 A Model with Exogenous Misallocation

In this Appendix, I develop a model in which misallocation is the result of exogenous, firm-specific wedges in input markets. I show that the wedges can be chosen so that the model has exactly the same cross-sectional implications as the model with endogenous misallocation developed in Section 2.2 and estimated in Section 2.1 — I refer to this as the full model. I also show that the model with exogenous misallocation has very different implications for the counterfactuals considered in Section 1.5. In particular I show that, for both the employment subsidy and the reduction in trade costs, this model predicts exactly zero change in the 'Within' component of misallocation. This explains the fact that in Column (2) of Table 1.9, the 'Within' component is exactly zero.

Firms differ in their productivities z, which are drawn from the stationary distribution of the AR(1) process (1.1). Production functions are Cobb-Douglas, as in

(1.2), and demand is CES as in (1.3). Labor is hired at a wage w, without any credit constraints. However, firms differ in a labor wedge τ_ℓ. A firm with wedge τ_ℓ faces an effective wage $w\tau_\ell$. Likewise, capital is hired at a rental rate r_k, and firms have capital wedges τ_k. This implies that firms set their marginal revenue products equal to

$$MRPK = \tau_k r_k, \quad MRPL = \tau_\ell w.$$

Since $MRPK$ and $MRPL$ will generally differ across firms, the equilibrium allocation will not maximize total sales, and therefore aggregate TFP. Hence, these wedges give rise to misallocation. Finally, I assume that a firm's export status is exogenously given.

Now, suppose we observe a cross-section of firms generated by the full model. In particular, suppose we observe their sales, employment, capital and export statuses. For each firm from the full model, we can generate an identical firm in the model with exogenous misallocation. To see this, note that observing a firm's inputs, sales and export status pins down its productivity z. Then we can choose wedges τ_ℓ and τ_k for this firm so that its employment and capital exactly match those generated by the full model. Finally, choose the firm's export status to match that of the full model. Following this procedure, for any cross-section of firms generated by the full model, the model with exogenous misallocation can generate an identical cross-section. An immediate implication is that average $MRPK$ and $MRPL$ among exporters and non-exporters will be the same across the two models.

By aggregating over the decision of individual firms in the model with exogenous misallocation, we can arrive at the same expression for TFP as in (1.17). However, this is where the similarities between the two models end. Below I show that Z_x and Z_d, the productivity of a representative exporter and non-exporter, respectively, are entirely determined by the joint distributions of τ_ℓ, τ_k and z. These distributions are not altered by any policy interventions, i.e., changes in trade costs or the exporter employment subsidy. Hence, in this model, $d \log Z_x = d \log Z_d = 0$ and changes in the 'Within' component of misallocation are always zero. Intuitively, this is because exogenous wedges shift a firm's labor or capital demand up or down, but do not change the elasticities with which these demands respond to changes in prices or wages. Therefore, conditional on export status, all firms in a given group (i.e., exporters or non-exporters) respond symmetrically, leaving the productivity of the representative exporter or non-exporter unchanged.

In the model with exogenous wedges, the labor demand ℓ of any non-exporter is

$$\ell = c_d z^{\sigma-1} \tau_\ell^{\sigma-\alpha(\sigma-1)} \tau_k^{-\alpha(\sigma-1)}$$

where c_d is a constant common to all non-exporters. Denote the (exogenous) joint distribution of z, τ_ℓ and τ_k among non-exporters be $\mathcal{G}_d$. Integrate to obtain total employment in non-exporting firms

$$L_d = c_d \int z^{\sigma-1} \tau_\ell^{\sigma-\alpha(\sigma-1)} \tau_k^{-\alpha(\sigma-1)} d\mathcal{G}_d.$$

Now the share of any non-exporter is

$$s_\ell = c_d z^{\sigma-1} \tau_\ell^{\sigma-\alpha(\sigma-1)} \tau_k^{-\alpha(\sigma-1)} L_d^{-1}.$$

Observe that this share does not depend on any endogenous variables because c_d cancels. By the same steps, we can obtain an analogous expression for s_k — this also does not depend on any endogenous variables.

Now, consider the definition of Z_d (1.18). This is an integral over productivities z and the shares s_ℓ and s_k. We have just seen that that these shares are entirely determined by the (exogenous) joint distributions of productivities and wedges. So Z_d is also exogenously determined, and does not depend on iceberg costs or the exporter employment subsidy. Therefore $d\log Z_x = 0$ in response to either of these policies. A similar argument applies to Z_x, and implies $d\log Z_x = 0$. Since the 'Within' component is just a weighted average of these it is also zero.

Appendix B

Appendix to: Housing Demand, Inequality, and Spatial Sorting [1]

B.1 Data

B.1.1 PSID

The primary consumption microdata come from the Panel Study of Income Dynamics. The PSID is administered biannually, with about 9,000 households in each wave. It included a consumption module starting in 1999 and added several categories in 2005. The survey now covers about 70% of spending in the national accounts (**blundell2016consumption**). Total expenditure is computed as the sum of all reported consumption categories: rent, food, utilities, telephone and internet, automobile expenses (including car loans, down payments, lease payments, insurance, repairs, gas, and parking), other transportation expenses, education, childcare, healthcare, home repairs, furniture, computers (2017 only), clothing, travel, and recreation. The PSID imputes a small number of observations to handle invalid responses. To match the definition in IPUMS, housing expenditure is equal to rent plus utilities.

[1] This is joint work with Trevor Williams

Homeowners were not asked to estimate the rental value of their home until 2017, so we restrict attention to renters and analyze homeowners with the CEX.

We use the 2005-2017 waves of the PSID and select our sample according to the following criteria. We drop respondents in the top and bottom 1% of the pre-tax income distribution in each year to guard against serious misreporting errors. We then select households in which the head is prime-age (25-55, inclusive) and attached to the labor force (head or spouse reports usually working at least 35 hours per week). The controls included in the regressions are dummies for family size bins, number of earners, age bins, sex of household head, race of household head, and year. Education is defined as years of schooling of the highest-earning household member. We use the PSID sample weights in all regressions.

Using the restricted access county identifiers, we can assign local prices to 92% of households in the PSID sample. The remaining households live in rural counties for which we do not construct rental price indices.

B.1.2 Rental Price Indices

We compute metropolitan area rental price indices from ACS data following Albouy (2016). We estimate a standard hedonic regression model of the form

$$r_{int} = p_{nt} + X'_{int}\beta_t + \varepsilon_{int} \tag{B.1}$$

where i denotes households, n denotes cities, r is log rent, and X is a set of observed dwelling characteristics: number of rooms, number of bedrooms, the interaction of the two, building age, number of units in the building, type of kitchen, type of plumbing, plot size, a dummy for whether the unit is a condo, and a dummy for whether the unit is a mobile home. The estimate of p, an MSA by year fixed effect, is the rental price index. The $\hat{p}_{nt}$ are mean zero in every year by construction, so we include year fixed effects in all specifications. We run the regression separately for each two-year window starting in 2005 and restrict the sample to renting households in the ACS.

Regressing $\hat{p}_{nt}$ on MSA average log rent yields a slope coefficient of 0.79 (population-weighted) and an R^2 of 0.90. In a robustness exercise, we use the Metropolitan Regional Price Parities published by the BEA (Bureau of Economic Analysis 2020). The BEA estimates MSA-level price indices for rents, goods and other services. Regressing our rental index on the BEA rental index yields a slope coefficient of 0.84 and an R^2 of 0.98.

B.1.3 CEX

We append the 2006-2017 Consumer Expenditure Surveys (CEX) together and annualize at the household level. We define rental expenditure as actual rent paid for renters (`rendwe`) and self reported rental-equivalent (`renteqvx`) for owners. As in PSID, we add utilities `util` to be consistent with the data available in the Census. To solicit rental equivalent, homeowners are asked "If someone were to rent your home today, how much do you think it would rent for monthly, unfurnished and without utilities?" We define total consumption expenditure as equal to total reported expenditure `totexp` *less* retirement and pension savings `retpen`, cash contributions `cashco`, miscellaneous outlays `misc` (which includes mortgage principal), and life and personal insurance `lifins`. For homeowners, we subtract `owndwe` and add `renteqvx`. We apply the exact same sample selection criteria and controls in the CEX as in the PSID (see Section B.1.1). We use CEX sample weights in all regressions.

In 2006, the CEX added more detailed geographic identifiers in the variable `psu`. The primary sampling unit, i.e. the MSA of residence, is available for a subset of households. The CEX identifies twenty-four large MSAs, which cover about 45% of households in the survey.

B.1.4 Census

We use the 5% public use samples from the 1980, 1990, and 2000 Censuses. For the final period of data, we use the 2009-2011 American Community Survey, a 3% sample. For convenience we refer to this as the "2010 data." IPUMS attempts to concord geographic units across years, although complete concordance is not possible because of data availability and disclosure rules. We classify MSAs according to the variable `metarea`. We produce a balanced panel using the following rule: if an MSA appears in all four years, then it is kept. If an MSA does not appear in all four years, then we assign all individuals in that MSA across all years to a residual state category. For example, Charlottesville, VA appears in 1980, 2000, and 2010, but not in 1990. Therefore we assign all individuals in Charlottesville in every year to "Virginia." This procedure gives us 219 MSAs (including Washington, D.C.) and 50 residual state categories, for a total of 269 regions. The share of national employment which can be assigned to an MSA, rather than a state residual, is 70% in 1980, 72% in 1990, and 75% in 2000 and 2010.

A worker is considered skilled if she or he has completed at least a four year college degree according to the variable `educ`. By this metric, the national fraction of workers who are skilled is 22.5% in 1980, 26.5% in 1990, 30.2% in 2000, and 35.7% in 2010.

We compute average wages and employment for each region, skill level, and year. Wages are from the IPUMS variable `incwage`. To be included in the wage and employment sample, workers must be between 25 and 55 years old, inclusive; not have any business or farm income; work at least 40 weeks per year and 35 hours per week; and earn at least one-half the federal minimum wage. Wages are adjusted to 2000 real values using BLS' Non-Shelter CPI.

Households within a skill level, location, and year are assumed to have expenditure given by the average post-tax wage income of group, $e_{int} = \overline{y}_{int} \equiv \lambda_t \left(\overline{w}_{int}\right)^{1-\tau}$, where λ_t is chosen to balance the budget. This assumes that the elasticity of expenditure to permanent post-tax income is unity. Households save in the data, but savings wash out in the aggregate since we focus on permanent income.

We could relax this assumption following **straub2019**. Suppose that expenditure were given by $e_{int} = \overline{y}_{int}^{\phi}$. If $\phi < 1$, expenditure would be nonhomothetic in permanent income. Qualitatively, this feature would increase the strength of our sorting mechanism. If consumption were less important for high-income households (relative to income), then they would be less sensitive to the price of local housing consumption (again, relative to income). We do not pursue a quantitative treatment of nonhomothetic total expenditure, which would require a dynamic quantitative spatial model beyond the scope of our paper.

In order to obtain instrumental variables for labor demand, we construct Bartik shift-share variables. The share is a region's industrial composition in 1980, and the shift is change in average wages nationwide (excluding the region itself).

We use the industry categories in the Census variable `ind1990`. Harmonizing the industries with our own crosswalk yields 208 industries which are consistently defined over all four periods. We drop individuals who cannot be classified into any industry ($\approx 0.3\%$ of workers) or who are in the military ($\approx 0.9\%$ of workers).

B.2 Estimation

We first describe how measurement error biases OLS estimates of the log-linearized estimating equation (2.9). We then describe alternative specifications to estimate the preferences in Section 2.1.

B.2.1 Measurement error

Recall that the log-linearized estimating equation is

$$\hat{\eta}_{int} = \omega_t + \omega' X_{int} + \beta \hat{e}_{int} + \psi \hat{p}_{nt} + \xi_{int}$$

We address measurement error in expenditure in the following way. First, partialling out observable demographics and prices, the reduced-form relationship between expenditure shares and total expenditure is

$$\eta = \beta e + \xi \tag{B.2}$$

where each variable is residualized, and hats and subscripts are suppressed for notational convenience. Expenditure and rental expenditure are measured with error: $\tilde{e} \equiv e + \upsilon^e$, $\tilde{r} \equiv r + \upsilon^r$, and $\tilde{\eta} \equiv \tilde{r} - \tilde{e}$. υ^e and υ^r are assumed to be uncorrelated with e, r, and ξ.

The OLS estimate of β is asymptotically

$$\begin{aligned}
\widehat{\beta}_{OLS} &= \frac{\text{cov}(\tilde{\eta}, \tilde{e})}{\text{var}(\tilde{e})} \\
&= \frac{\beta \sigma_e^2 + \sigma_{\upsilon^r,\upsilon^e} - \sigma_{\upsilon^e}^2}{\sigma_e^2 + \sigma_{\upsilon^e}^2} \\
&= \beta \frac{\sigma_e^2}{\sigma_e^2 + \sigma_{\upsilon^e}^2} + \frac{\sigma_{\upsilon^r,\upsilon^e} - \sigma_{\upsilon^e}^2}{\sigma_e^2 + \sigma_{\upsilon^e}^2}
\end{aligned}$$

The attenuation bias $\sigma_e^2/(\sigma_e^2 + \sigma_{\upsilon^e}^2)$ is familiar from classical measurement error. There are two additional sources of bias: (1) measurement error in expenditure appears on both the left- and right-hand sides of (B.2) and (2) measurement errors in expenditure and rent are mechanically correlated. The direction of the bias is ambiguous, but is likely to be downward if measurement error in expenditure is large and not too highly correlated with measurement error in rent.

B.2.2 Alternative specifications in PSID

We present several alternative specifications in Table B.1, still using our baseline sample of renters in the PSID.

In column (1), we include liquid wealth as a control (we use the inverse hyperbolic sine transformation to include households with zero wealth). Liquidity constraints feature in some models of nonhomothetic housing demand such as **bilal2018location**. The estimates are unchanged, suggesting that liquidity constraints are not first order. In column (2), we instrument for expenditure using job tenure. The exclusion restriction is that job tenure affects the housing share only by shifting total expenditure, conditional on controls including family size and age. The estimates are similar. Columns (3) and (4) split the sample into movers and non-movers, respectively, in order to explore a key margin of adjustment to housing expenditure. At annual frequency, households can adjust their housing expenditure either by moving or by re-negotiating their rent. The fact that the estimated ϵ in columns (3) and (4) are similar suggest that both margins appear to be operative. Non-movers' housing expenditure is only slightly more inelastic than movers'. Column (5) uses a county-level rental price index from Zillow, a real estate analytics company (**zillow**). Reassuringly, the estimates are similar even with different data and a different level of geography. Column (6) does not instrument for price. The results are similar to the baseline, suggesting that endogeneity of prices is not first-order. Column (7) shows that the coefficient estimates are robust to excluding demographic controls, which is evidence against households' sorting on variables other than income and price. Column (8) estimates (2.11) by 2SLS with household fixed effects, and yields very similar results to our fixed-effect GMM estimates. Column (9) repeats 2SLS with household and MSA fixed effects.

Table B.1: Preferences, alternative specifications (PSID)

Dependent variable: Log housing share

	(1) GMM	(2) GMM	(3) GMM	(4) GMM	(5) GMM	(6) GMM	(7) GMM	(8) 2SLS	(9) 2SLS
ϵ	-0.303	-0.400	-0.275	-0.363	-0.256	-0.350	-0.308	-0.434	
	(0.037)	(0.128)	(0.043)	(0.045)	(0.043)	(0.030)	(0.030)	(0.140)	
σ	0.523	0.506	0.616	0.369	0.686	0.461	0.530	0.522	
	(0.077)	(0.089)	(0.094)	(0.105)	(0.066)	(0.037)	(0.078)	(0.228)	
Log expenditure								-0.415	-0.430
								(0.148)	(0.137)
Log price								0.457	
								-0.173	
Demographic controls	✓	✓	✓	✓	✓	✓		✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controlling for wealth	✓								
Household FE								✓	✓
MSA FE									✓
Expenditure IV	Income	Tenure	Income	Income	Income	Income	Income	Income	Income
Price	Census	Census	Census	Census	Zillow	Census, no IV	Census	Census	
Sample	Full	Full	Movers	Non-movers	Full	Full	Full	Full	Full
R^2								0.17	0.18
First-stage F-stat.								26.8	34.8
N	10,678	10,353	6,729	3,964	6,572	12,351	10,678	8,670	9,985
No. of clusters	217	217	212	176	192	484	217	197	350

Source: PSID, Census, Zillow, and Saiz (2010).

Note: Column (1) instruments for expenditure using job tenure. Columns (2) and (3) split the sample into households which moved addresses and those which did not. Column (4) use median price per square foot at the county level from Zillow. Column (5) repeats the main GMM specification without instrumenting for price. Column (6) omits demographic controls. Column (7) reports the linearized model with household fixed effects. Column (8) reports the linearized model with household and MSA fixed effects. Price instruments are geographic and regulatory constraints. Standard errors clustered at MSA level.

B.2.3 Consumer Expenditure Survey (CEX)

In this section we present additional results from the Consumer Expenditure Survey (CEX). Reassuringly, all findings are close to our main results.

In the first column of Table B.2, we re-estimate our baseline specification in the CEX. The estimated expenditure elasticity is slightly higher, but the difference is not statistically significant.

Homeowners

Thus far we have focused on renting households because we do not observe expenditure on owner-occupied housing. In this section we explore whether our results extend to homeowners too. An appropriate measure of housing expenditure by homeowners is *rent equivalent*, which is the market rate for the flow of housing services consumed. The PSID consumption module did not elicit rent equivalent until 2017, but rent equivalent is available in all recent waves of the CEX. Therefore we use the CEX to study homeowners.

Column (2) of Table B.2 pools renting and owning households together. The estimate is consistent with significant nonhomotheticity. Restricting attention only to owners (column (3)) yields even stronger nonhomotheticity than the baseline estimate for renters.

In columns (5) and (6), we use an alternative measure of housing expenditure for homeowners, *out-of-pocket expenses*. We define out-of-pocket expenses as the sum of mortgage interest, property tax, insurance, maintenance, and repairs. We omit payments on mortgage principal since these payments are savings, not consumption. Out-of-pocket expenses reflect the user cost of housing, which is equal to the rental value of the house in equilibrium. The estimates are close to our baseline results.

In our main analysis of homeowners, we restrict our sample to households who own a single home, which includes 94% of homeowners in the CEX. The reason

Table B.2: Preferences (CEX)

Dependent variable: Log housing share

	(1) GMM	(2) GMM	(3) GMM	(4) GMM	(5) GMM	(6) GMM
ϵ	-0.312	-0.253	-0.443	-0.440	-0.312	-0.228
	(0.043)	(0.031)	(0.041)	(0.040)	(0.026)	(0.038)
σ	0.319	0.330	0.114	0.143	0.272	0.338
	(0.198)	(0.146)	(0.155)	(0.157)	(0.109)	(0.147)
Sample	Renters	Pooled	Owners	Owners	Pooled	Owners
Rent measure for owners		Rent equivalent	Rent equivalent	Rent equivalent	Out of pocket	Out of pocket
Including homeowners with second homes				✓		
Demographic controls	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
N	2,995	8,269	5,274	5,659	8,269	5,274

Source: CEX, Census, and Saiz (2010).

Note: Column (1) replicates our baseline specification of Table 2.1 column (5), using the CEX. Column (2) adds homeowners owning one home, measuring housing expenditure by self-reported rental equivalent. Column (3) restricts to homeowners only. Column (4) includes households who own second homes. Columns (5) and (6) measures housing expenditure as out-of-pocket expenses, defined as mortgage interest, property taxes, insurance, maintenance, and repairs; mortgage principle is excluded. Instrument is log household income. Standard errors clustered at MSA level.

is that expenditure on second homes does not reflect the local cost of living, but rather is a luxury more akin to recreation or vacations. That said, it is possible that second homes are a substitute for primary homes in expensive markets: for example, a household could live in a small house in the city and maintain a larger house in the country. Including second homeowners in column (4) leaves our results virtually unchanged.

Imputing rents from home values

Because data on rent equivalent is (until very recently) unavailable in the PSID, the standard approach has been to impute rents as a constant fraction of self-reported home value, generally six percent (**attanasiopistaferri2016**; **straub2019**). We argue that this is not an appropriate strategy if housing demand is nonhomothetic. The six percent figure is from **poterba2008tax**, who compute the user cost of housing with data from the Survey of Consumer Finances. **poterba2008tax** document considerable variation in the user cost across different types of homeowners, with a mean of six percent. The Residential Financial Survey, used by the BEA to impute rents in the national accounts, shows that the rent-to-value ratio is strongly decreasing in home value, a fact that we replicate from the CEX in Figure B.1.

Imputing rent as a constant fraction of home value would tend to deflate the housing shares of households with low home values and inflate the housing shares of households with high home values, obscuring nonhomotheticity in the data. Therefore, our preferred approach is to use reported rent equivalent expenditure.

B.2.4 Income elasticities from the literature

Table B.3 summarizes estimates of the income elasticity of housing demand from the literature. Controlling for local prices, using expenditure on the right hand side, and accounting for measurement error with an IV are all key in obtaining a consistent estimate of the elasticity.

B.3 Theory

B.3.1 Irrelevance of income elasticity normalization

In Subsection 2.1.2, we introduced NHCES preferences as

$$U^{\frac{\sigma-1}{\sigma}} = \Omega^{\frac{1}{\sigma}} h^{\frac{\sigma-1}{\sigma}} U^{\frac{\epsilon}{\sigma}} + c^{\frac{\sigma-1}{\sigma}}. \tag{B.3}$$

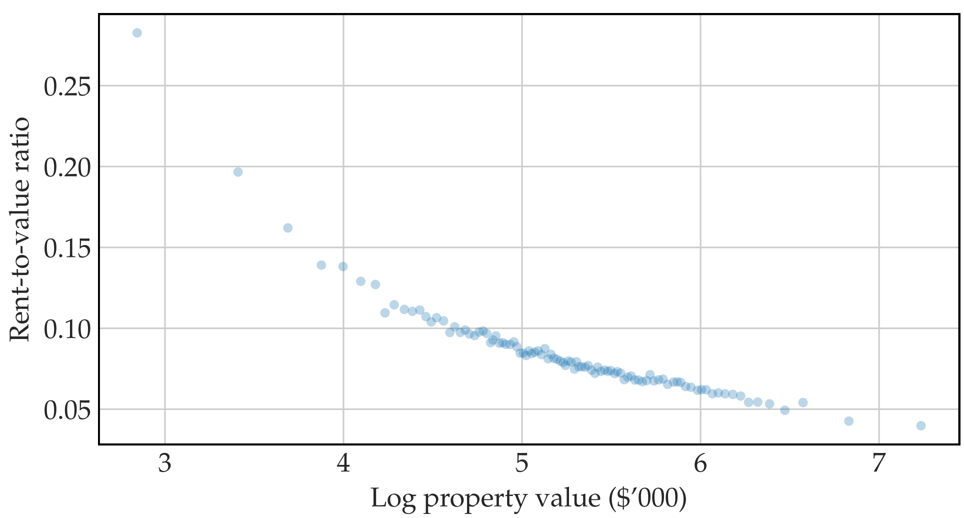
Here, we show that this is equivalent to the more general formulation

$$\tilde{U}^{\frac{\sigma-1}{\sigma}} = \Omega_h^{\frac{1}{\sigma}} h^{\frac{\sigma-1}{\sigma}} \tilde{U}^{\frac{\epsilon_h}{\sigma}} + \Omega_c^{\frac{1}{\sigma}} c^{\frac{\sigma-1}{\sigma}} \tilde{U}^{\frac{\epsilon_c}{\sigma}} \tag{B.4}$$

where ϵ_h, ϵ_c, Ω_h and Ω_c are parameters.

First, observe that (B.3) is a special case of (B.4) with $\epsilon_c = 0$, $\Omega_c = 1$, $\Omega_h = \Omega$, and $\epsilon_h = \epsilon$. Second, let us take ϵ_c and $\Omega_c > 0$ as given. It is straightforward to

Figure B.1: Rents and Property Values



Source: CEX, 2006-2017. Average ratio of self-reported rent equivalent to self-reported property value computed for 100 property value bins.

Table B.3: Income elasticities in the literature

Paper	Elasticity	Sample	Local prices?	Expenditure?	IV?
rosenthal2014private[a]	-0.88	Renters	✓		
ioannides2008interactions[b]	-0.79	Owners	✓		
hansen1996income[c]	-0.73	Renters			
larsen2014engel[d]	-0.67	Owners			
zabel2004demand[e]	-0.52	Owners	✓		
Albouy, Ehrlich, and Liu (2016)[f]	-0.28	Renters	✓		
lewbel2009tricks[g]	-0.28	Renters		✓	✓
attanasio2012modelling[h]	-0.22	Both		✓	
Aguiar and Bils (2015)[i]	-0.08	Both		✓	✓
Davis and Ortalo-Magné (2011)[j]	-0.01	Both		✓	✓
Paper benchmark[k]	-0.25	Renters	✓	✓	✓

[a] American Housing Survey, 1985-2011. Table 5, column 1.
[b] American Housing Survey, 1985-1993. Table 5, column 1.
[c] American Housing Survey, 1989. Table 5, column 2, last row.
[d] Norweigian Rental Survey and Consumer Expenditure Survey, 2007. Table 2, row 5.
[e] American Housing Survey, 2001. Table 3, row 3.
[f] US Census, 1970-2014. Table 1, column 3.
[g] Canadian Family Expenditure Surveys, 1969-1996. Median uncompensated elasticity computed using authors' replication file following their Appendix VII.1.
[h] British Household Panel Survey, 1991-2002. Table 4, panel B. Estimates for high- and low-education groups are averaged with weights one-third and two-thirds, respectively.
[i] US CEX, 1980-2010. Table 2, column 1.
[j] US CEX, 1982-2003. Text, page 253.
[k] PSID, 2005-2017.

show that by choosing Ω_h and ϵ_h correctly, we can produce preferences which yield identical housing demand functions as in (B.3). To see this, divide both sides of (B.4) by $\Omega_c^{\frac{1}{\sigma}} \tilde{U}^{\frac{\epsilon_c}{\sigma}}$ to obtain

$$\Omega_c^{-\frac{1}{\sigma}} \tilde{U}^{\frac{\sigma-1}{\sigma}-\frac{\epsilon_c}{\sigma}} = \Omega_h^{\frac{1}{\sigma}} \Omega_c^{-\frac{1}{\sigma}} h^{\frac{\sigma-1}{\sigma}} \tilde{U}^{\frac{\epsilon_h-\epsilon_c}{\sigma}} + c^{\frac{\sigma-1}{\sigma}}. \tag{B.5}$$

Now set

$$\Omega_h = \Omega_c^{\epsilon(1-\sigma)+1}\Omega, \quad \epsilon_h = \epsilon + \epsilon_c \left(1 - \frac{\epsilon}{\sigma-1}\right).$$

Inserting these expressions into (B.5), we obtain

$$\left(\Omega_c^{1-\sigma}\tilde{U}^{1-\frac{\epsilon_c}{\sigma-1}}\right)^{\frac{\sigma-1}{\sigma}} = \Omega^{\frac{1}{\sigma}} h^{\frac{\sigma-1}{\sigma}} \left(\Omega_c^{1-\sigma}\tilde{U}^{1-\frac{\epsilon_c}{\sigma-1}}\right)^{\frac{\epsilon}{\sigma}} + c^{\frac{\sigma-1}{\sigma}}. \tag{B.6}$$

By comparing this with (B.3), we can see that

$$\tilde{U} = \Omega_c^{\frac{1}{\sigma-1-\epsilon_c}} U^{\frac{\sigma-1}{\sigma-1-\epsilon_c}}. \tag{B.7}$$

That is, $\tilde{U}$ is a monotonically increasing transformation of U and so represents the same preferences over housing and non-housing consumption.

Finally, it could in principle be the case that (B.4), when incorporated into the quantitative spatial model developed in Section 2.2, might lead to different preferences over locations than (B.3). But for our isolastic model of labor supply, this is not the case. To see this, consider the location choice equation (2.16) using the preferences defined in (B.4). We obtain

$$l_{in} = \frac{\tilde{v}_{in}^{\theta} B_n}{\sum_m \tilde{v}_{im}^{\theta} B_m} L_i, \tag{B.8}$$

where

$$\tilde{v}_{in} = \Omega_c^{\frac{1}{\sigma-1-\epsilon_c}} v_{in}^{\frac{\sigma-1}{\sigma-1-\epsilon_c}}. \tag{B.9}$$

Combining these two expressions, we obtain

$$l_{in} = \frac{v_{in}^{\tilde{\theta}} B_n}{\sum_m v_{im}^{\tilde{\theta}} B_m} L_i, \tag{B.10}$$

where

$$\tilde{\theta} = \theta \left(\frac{\sigma - 1}{\sigma - 1 - \epsilon_c} \right). \tag{B.11}$$

That is, choosing $\epsilon_c > 0$ just proportionally rescales the migration elasticity θ. Following the calibration strategy outlined in Section 2.3 with $\epsilon_c > 0$, we would just estimate a rescaled version of θ, and all of the model's predictions would be unchanged. Therefore, assuming $\epsilon_c = 0$ and $\Omega_c = 1$ is without loss of generality.

B.3.2 Proof of Lemma 1

To derive the ideal price indices in (2.20) and (2.21), substitute the expression for P_{in} in (2.18) into the Hicksian demand function (2.2) to obtain an expression in terms of η_{in}

$$\frac{\eta_{in}}{1 - \eta_{in}} = \Omega p_n^{1-\sigma} \left(\frac{e_{in}}{P_{in}} \right)^{\epsilon}.$$

Substituting this into the expression for η_{in} in (2.3) and rearranging yields

$$P_{in}^{1-\sigma} = \left(1 + \Omega p_n^{1-\sigma} \left(\frac{e_{in}}{P_{in}} \right)^{\epsilon} \right).$$

Replacing expenditures with productivities following (2.13) and (2.14) yields the ideal price indices (2.20) and (2.21), reproduced below

$$P_{un} = \left(1 + \Omega \left(\frac{z_n}{P_{un}} \right)^{\epsilon} p_n^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$$
$$P_{sn} = \left(1 + \Omega A^{\epsilon} \left(\frac{z_n}{P_{sn}} \right)^{\epsilon} p_n^{1-\sigma} \right)^{\frac{1}{1-\sigma}}.$$

Defining $c_n = z_n^{\frac{\epsilon}{1-\sigma}} p_n$, we define the functions $P_u(c)$ and $P_s(c)$ implicitly

$$P_u(c) = \left(1 + \Omega c^{1-\sigma} P_u(c)^{-\epsilon}\right)^{\frac{1}{1-\sigma}} \tag{B.12}$$

$$P_s(c) = \left(1 + A^{\epsilon} \Omega c^{1-\sigma} P_s(c)^{-\epsilon}\right)^{\frac{1}{1-\sigma}} . \tag{B.13}$$

Clearly $P_i(c_n) = P_{in}$.

To prove Lemma 1, we establish that $\log P_u(c) - \log P_s(c)$ is strictly increasing in c. This is equivalent to showing that

$$\delta_u(c) > \delta_s(c)$$

where $\delta_i(c)$ is the elasticity of P_i with respect to c. Differentiating (2.20) and (2.21) and rearranging yields

$$\delta_i(c) = \frac{(1-\sigma)\eta_i(c)}{1-\sigma+\epsilon\eta_i(c)}$$

where η_i is the housing expenditure share of type i when facing productivity adjusted housing cost c. δ_i is clearly a strictly increasing function of $\eta_i(c)$. Whenever $\epsilon < 0$, so that housing demand is income inelastic, $\eta_u(c) > \eta_s(c)$ because the expenditure share of the unskilled household is always higher. Therefore $\delta_u > \delta_s$ and $\log P_u(c) - \log P_s(c)$ is strictly increasing in c. Lemma 1 follows.

B.3.3 Proof of Proposition 1

We first assume $\epsilon < 0$. P_{un} is unaffected by changes in A, so

$$ds_n = -\theta d \log P_{sn} + d\zeta$$

from (2.19). Inspection of (B.13) shows that $A^{\epsilon(1-\sigma)^{-1}}$ appears isomorphically to c, and so

$$ds_n = -\theta\epsilon\,(1-\sigma)^{-1}\,\delta_s(c_n)d\log A + d\zeta$$

where δ_s is the elasticity of P_s with respect to productivity adjusted housing costs. In Appendix B.3.2 we showed that δ_s is a strictly increasing function of c. Since $\epsilon < 0$, this implies that ds_n is also strictly increasing in c. Since by Lemma 1 s_n is strictly increasing in c, we then have that ds_n is strictly increasing in s_n. Now we turn to sorting S, defined as the variance of s_n. We prove this statement for the weighted variance with positive (and fixed) weights ω_n which sum to 1 since we will weight by 1980 employment shares in our empirical application. By definition

$$S = \sum_n \omega_n\,(s_n - \bar{s})^2$$

$$\bar{s} = \sum_n \omega_n s_n.$$

Differentiating

$$dS = 2\sum_n \omega_n\,(ds_n - d\bar{s})\,(s_n - \bar{s}) = 2Cov(ds_n, s_n).$$

Since ds_n is a strictly increasing function of s_n, this covariance is positive and so $dS > 0$. This completes the proof for the case of $\epsilon < 0$. When $\epsilon = 0$, $P_{sn} = P_{un}$ and (2.19) then implies that $ds_n = 0$ for all n. $dS = 0$ follows.

B.3.4 Proof of Proposition 2

We start by taking logs of (2.26)

$$\log l_{in} = \theta\log v_{in} + \log B_{in} - \log U_i$$

where U_i is just the denominator in (2.26), divided by L_i. We difference this across types in the same location n and use the definition of the log skill ratio s_n

$$s_n = \theta \log\left(\frac{v_{sn}}{v_{un}}\right) + \log\left(\frac{B_{sn}}{B_{un}}\right) - \log\left(\frac{U_s}{U_u}\right).$$

Now when $\epsilon = 0$ preferences are homothetic, and v_{in} is given by

$$v_{in} = e_{in}\left(1 + \Omega p_n^{1-\sigma}\right)^{\frac{-1}{1-\sigma}}.$$

This implies that the ratio v_{sn}/v_{un} is just the ratio of expenditures. Therefore

$$s_n = \theta \log\left(\frac{e_{sn}}{e_{un}}\right) + \log\left(\frac{B_{sn}}{B_{un}}\right) - \log\left(\frac{U_s}{U_u}\right).$$

Now use (2.29) to replace expenditures with wages

$$s_n = \theta(1-\tau) \log\left(\frac{w_{sn}}{w_{un}}\right) + \log\left(\frac{B_{sn}}{B_{un}}\right) - \log\left(\frac{U_s}{U_u}\right).$$

Next, we replace wages with productivities and labor supplies, using (2.25), and rearrange

$$(1 + \theta(1-\tau)\rho^{-1})s_n = \theta(1-\tau) \log a_n + \log\left(\frac{B_{sn}}{B_{un}}\right) - \log\left(\frac{U_s}{U_u}\right).$$

Differencing this equation between any two locations shows that the difference between s_n and s_m for any two locations n and m depends only on location specific fundamentals and not on A. So changes in A have no effect on the variance of s_n, i.e on sorting S.

B.4 Calibration

Tax system

We use data from the 1981/91/2001/11 waves of the PSID (each containing summary information on the *prior* year's income). Using the same sample restrictions as in section 2.1, we run the PSID data through the NBER's TAXSIM program. For each household, pre-tax income is computed as adjusted gross income minus Social Security transfers. Post-tax income is computed as pre-tax income minus federal and state taxes (including payroll taxes) plus Social Security transfers. We estimate (2.32) by pooled OLS over the four periods. Our estimated $\hat{\tau}$ is 0.174 (robust s.e. 0.003). The R^2 of the regression is 0.98, suggesting that, despite its parsimony, a log-linear tax equation is a good approximation to the actual tax system in the United States. Our estimate is close to Heathcote, Storesletten, and Violante (2017), who estimate $\hat{\tau} = 0.181$.

Housing Supply Elasticities

Our estimating equation is

$$\Delta \log p_n = \Delta \log \tilde{\Pi}_n + (\chi + \chi_L UNAVAL_n + \chi_R WRLURI_n)\, \Delta \log \left(\sum_i \eta_{in} e_{in} l_{in} \right). \tag{B.14}$$

Changes are between 1980 and 2010. Saiz (2010) reports values of land unavailability $UNAVAL_n$ and regulatory constrains $WRLURI_n$ for a subset of MSAs. After dropping those for which these measures are missing, we are left with 193 MSAs. Prices p_n are obtained from hedonic regressions in the Census data as described in the text. We use Census data on employment, wages, and (2.3) to construct housing expenditure $\sum_i \eta_{in} e_{in} l_{in}$ for each MSA. Finally we use the Bartik shifter Z_{int} (and its interactions with $UNAVAL_n$ and $WRLURI_n$) as an instrument for housing expenditure. Table B.4 reports the result of estimating (B.14) by 2SLS. For the 193 locations with complete data, we then define

$$\gamma_n = \frac{\chi_n}{1 - \chi_n}$$

where

$$\chi_n = \chi + \chi_L UNAVAL_n + \chi_R WRLURI_n.$$

Of the remaining locations, 50 are the nonmetro portions of states and 26 are MSAs for which $UNAVAL_n$ and $WRLURI_n$ are not available. For the 26 MSAs, we define γ_n to be the median among the 193 MSAs with complete information. For the 50 state residuals, we set γ_n to the lowest value among the 193 MSAs with complete information, on the assumption that supply is likely to be more elastic in nonmetro areas.

Migration elasticity

We estimate θ by requiring our model to match the results of Hornbeck and Moretti (2019). That paper estimates the causal effect of TFP shocks between 1980 and 1990 on employment and wages. We mimic their setting by shutting down all shocks other than shocks to productivity and then repeating their regressions using the output of our model. Our target is the ratio of the effect on employment to

Table B.4: Housing Supply Elasticity Estimates
Dependent variable: Log price change, 1980-2010

χ	0.209
	(0.069)
χ_L	0.090
	(0.055)
χ_R	0.230
	(0.057)

Source: Census. Robust standard errors in parentheses.

the effect on wages by 2010 — the long run elasticity of employment to wages. This implies a target of $4.03/1.46 = 2.76$ (see Table 2, Column (3) of Hornbeck and Moretti (2019)). Formally we proceed as follows:

(i) Guess θ

(ii) Invert the model in 1980 and 1990 to obtain fundamentals $(A^t_{in}, B^t_{in})_{i,n}, (\Pi^t_n)_n, (L^t_i)_i$ for $t = 1980, 1990$.

(iii) Solve the model with fundamentals $(A^{90}_{in}, B^{80}_{in})_{i,n}, (\Pi^{80}_n)_n, (L^{80}_i)_i$ to obtain $\left(\hat{l}^{90}_{in}, \hat{w}^{90}_{in}\right)_{i.n}$.

(iv) Define $L^{80}_n = \sum_i l^{80}_{in}$, $W^{80}_n = \sum_i l^{80}_{in} w^{80}_{in} / \sum_i l^{80}_{in}$ and $\log Z^{80}_n = \sum_i l^{80}_{in} \log A^{80}_{in} / \sum_i l^{80}_{in}$ and likewise for $\hat{L}^{90}_n$, $\hat{W}^{90}_n$ and $\log \hat{Z}^{90}_n$

(v) Estimate the models below by OLS, weighting by 1980 employment:

$$\log \hat{L}^{90}_n - \log L^{80}_n = \pi^L \left(\hat{Z}^{90}_n - Z^{80}_n\right) + \upsilon^L_n$$
$$\log \hat{W}^{90}_n - \log W^{80}_n = \pi^W \left(\hat{Z}^{90}_n - Z^{80}_n\right) + \upsilon^W_n.$$

The fact that we only study changes between 1980 and 1990 is innocuous, because our model has no transitional dynamics.

(vi) Calculate π^L/π^W.

(vii) Update θ until π^L/π^W converges to the target value.

This procedure yields $\theta = 5.11$.

B.5 Counterfactual

B.5.1 Cobb-Douglas preferences with type-specific parameters

In Section 2.2 we considered Cobb-Douglas preferences with different expenditure shares by skill type. There we showed that in the simple model, skill-specific Cobb-Douglas preferences do not link changes in the skill premium to changes in spatial sorting. In our quantitative model this is no longer true, because endogenous changes in housing costs will cause changes in sorting by skill when the weight on housing differs across skill groups. We repeat our main counterfactual experiment under the assumption that each skill group has Cobb-Douglas preferences with potentially different expenditure shares. We set the expenditure share for each group equal to its employment-weighted average across MSAs in 1980. This model explains only 1.78% of the observed increase in spatial sorting, compared to 23% for our explicitly nonhomothetic model. We conclude that even extending Cobb-Douglas preferences to accommodate different expenditure shares by skill cannot capture the link between the rising skill premium and spatial sorting by skill.

B.5.2 Alternative Measures of Sorting

Here we define the alternative measures of sorting discussed in subsection 2.4.2.

The Theil index for a non-negative variable x with weights ω_n is defined as

$$T = \sum_i \omega_n \left(\frac{x_n}{\bar{x}}\right) \log\left(\frac{x_n}{\bar{x}}\right)$$

where $\bar{x}$ is the weighted average of x_n. We use this as a measure of sorting by setting $x_n = \exp(s_n)$, where s_n is the log-skill ratio, and weight by 1980 employment.

The dissimilarity index D for two populations u and s, spread over geographical units indexed by n is given by

$$D = \frac{1}{2} \sum_{n} \left| \left(\frac{l_{sn}}{L_s} \right) - \left(\frac{l_{un}}{L_u} \right) \right|$$

Note that employment weights are already implicit in this expression.

The results of our main counterfactual using these alternative measures, as well as the 90/10 ratio of the log skill ratio distribution, are shown in Table B.5. All measures of sorting have increased since 1980, and for columns (2) and (4), the effect of the skill premium is quite similar to our baseline result. For the dissimilarity index, we find a somewhat lower value.

B.5.3 Alternative Counterfactual Implementation

In our baseline counterfactual, the values A^t and Z^t are chosen to (i) fix the skill premium at its 1980 level and (ii) match the growth of average unskilled wages from the data. As an alternative, we modify (ii) to match the growth of average wages (pooling unskilled and skilled together). Although the implied sequence of (A^t, Z^t) is

Table B.5: Alternative Measures of Sorting

	(1) Var. log skill ratio	(2) Theil	(3) Dissimilarity	(4) 90-10
Change, 1980-2010				
Data	32.6%	34.5%	19.0%	8.28 pp.
Model	25.2%	27.1%	16.0%	6.10 pp.
Effect of skill premium (% of observed change)	22.6	21.3	15.5	26.3

Note: Each column reports the change in sorting in the data and in the economy in which all fundamentals change as in the data, apart from the aggregate productivity parameters A and Z. The aggregate parameters A and Z are changed to eliminate the observed increase in the skill premium, 1980-2010. Column (1) is our preferred measure of sorting, while columns (2)-(4) present alternative measures of sorting. The final row reports the difference between the data and the model economy, which measures the causal effect of the rising skill premium on each measure of sorting. See Appendix B.5.2 for details of each measure of sorting.

somewhat different, the counterfactual result is similar at 24%.

B.5.4 Alternative Parametrization of Preferences

We recalibrate our model to Price Independent Generalized Linear (PIGL) utility, a leading case of nonhomothetic preferences (**boppart2014**; Eckert and Peters 2018). PIGL admits a closed form for the indirect utility function (2.15),

$$v_{in} = \frac{1}{\varepsilon}(e_{in}^{\varepsilon} - 1) - \frac{\Omega}{\varsigma}(p_n^{-\varsigma} - 1)$$

for parameters $0 < \varepsilon < \varsigma < 1$ and $\Omega > 0$. By Roy's identity, the housing share is

$$\eta_{in} = \Omega e_{in}^{-\varepsilon} p_n^{\varsigma} \tag{B.15}$$

Taking logs, adding a time subscript, and interpreting the scalar Ω as an idiosyncratic household demand shifter Ω_{int}, (B.15) is equivalent to the linearized estimating equation (2.7) for NHCES utility. The income elasticity is ε and the price elasticity is ς, which correspond to β and ψ, respectively, in (2.8). We can therefore read the parameters directly off column (4), Table 2.1, setting $\varepsilon = 0.248$ and $\varsigma = 0.390$. After recalibrating the full model we find that the skill premium explains 19.6% of the increase in sorting since 1980, comparable to our baseline results. More generally, (2.8) is a first order approximation to any demand system. We conclude that our findings are not sensitive to the parametrization of utility.

B.5.5 Endogenous Amenities

Diamond (2016) shows the importance of endogenous amenities for understanding the location choices of skilled versus unskilled workers. In this subsection we consider how our results might change in the presence of endogenous amenities.

We start by incorporating them into the simple model described in Section 2.2. Following Diamond (2016) we model amenities as

$$B_{in} = b_{in} \left(\frac{l_{sn}}{l_{un}} \right)^{\beta_s}. \tag{B.16}$$

That is, for both types amenities depend on the skill ratio, but different types may value them differently — this is captured by β_s. In the context of our simple model, we do not allow exogenous differences in amenities across types, and so we impose $b_{sn} = b_{un} = b_n$. Diamond (2016) shows that skilled households value endogenous amenities more than unskilled households, implying $\beta_s > \beta_u$. We also impose $\beta_s - \beta_u < 1$ to avoid endogenous amenities so strong that they cause perfect sorting (i.e a situation in which skilled and unskilled workers inhabit totally different locations).

It is helpful to compare two economies with the same fundamentals — one without endogenous amenities, whose variables are denoted by $\bar{x}$, and one with endogenous amenities, whose variables are denoted by $\tilde{x}$. In the economy with endogenous amenities (2.19) becomes

$$\tilde{s}_n = \tilde{\zeta} - \theta \left(\log \tilde{P}_{sn} - \log \tilde{P}_{un} \right) + (\beta_s - \beta_u)\, \tilde{s}_n. \tag{B.17}$$

Notice that in our model the ideal price indices are independent of the presence of endogenous amenities, and so $\bar{P}_{in} = \tilde{P}_{in}$. This implies

$$\bar{s}_n = \bar{\zeta} - \theta \left(\log \tilde{P}_{sn} - \log \tilde{P}_{un} \right)$$

and therefore

$$\tilde{s}_n = (1 - (\beta_s - \beta_u))^{-1} \left(\bar{s}_n + (\tilde{\zeta} - \bar{\zeta}) \right). \tag{B.18}$$

That is, skill ratios in the economy with endogenous amenities are simply an affine transformation of skill ratios in the economy without endogenous amenities. In particular given our assumption on β_s and β_u, the slope of $\tilde{s}_n$ with respect to $\bar{s}_n$ is above one. This leads to the first result of this section

Proposition 3. *Suppose $\beta_s > \beta_u$. If $\epsilon < 0$, sorting is higher in the presence of endogenous amenities, i.e $\tilde{S} > \bar{S}$. If instead $\epsilon = 0$ then $\tilde{S} = \bar{S} = 0$.*

This follows directly from observing that $\tilde{s}_n$ is an affine transformation of $\bar{s}_n$ with a coefficient on $\bar{s}_n$ above 1. Proposition 3 tells us that endogenous amenities amplify the effects of nonhomothetic housing demand. Nonhomothetic housing demand ensures that high price locations have a higher skill ratio. Endogenous amenities then encourage even more skilled workers to locate there. But it is important to note that when $\epsilon = 0$, there is no sorting even with endogenous amenities, showing that they do not create an independent motive for sorting in our model, but rather amplify existing ones.

We now proceed to our next result, concerning the effect of an increase in the skill premium, $d \log A > 0$. Differentiating (B.17) yields

$$d\tilde{s}_n = (1 - (\beta_s - \beta_u))^{-1} \left(d\tilde{\zeta} - \theta \left(d \log \tilde{P}_{sn} - \log \tilde{P}_{un} \right) \right).$$

Again substituting out prices using the economy without endogenous amenities, we obtain

$$d\tilde{s}_n = (1 - (\beta_s - \beta_u))^{-1} \left(d\bar{s}_n + \left(d\tilde{\zeta} - d\bar{\zeta} \right) \right).$$

Now $d\tilde{s}_n$ is an affine function of $d\bar{s}_n$ with a coefficient on $d\bar{s}_n$ above 1. Following the same steps as above, we obtain the result.

Proposition 4. *Suppose $\beta_s > \beta_u$. If $\epsilon < 0$, sorting increases more when amenities are endogenous. Formally, $d\tilde{S} > d\bar{S}$. When $\epsilon = 0$ then $d\tilde{S} = d\bar{S} = 0$*

Proposition 4 shows that endogenous amenities amplify the mechanism we focus on in this paper — diverging incomes causing diverging sensitivities to housing costs and thus diverging location choices – but do not independently link the skill premium to spatial sorting.

Finally, we extend Proposition 2 to a richer environment with endogenous amenities. We drop the assumption that $b_{sn} = b_{un}$. Adding endogenous amenities does not change the derivation presented in the proof of Proposition 2, so we start from

$$(1 + \theta(1 - \tau)\rho^{-1})s_n = \theta(1 - \tau)(\log a_n + \log A) + \log\left(\frac{B_{sn}}{B_{un}}\right) - \log\left(\frac{U_s}{U_u}\right).$$

Inserting our definition of B_{in} and rearranging, we obtain

$$(1 + \theta(1 - \tau)\rho^{-1} - (\beta_s - \beta_u))s_n = \theta(1 - \tau)(\log a_n + \log A) + \log\left(\frac{b_{sn}}{b_{un}}\right) - \log\left(\frac{U_s}{U_u}\right).$$

Following exactly the same steps as in Proposition 2, we obtain our final result:

Proposition 5. *Suppose $\beta_i \neq 0$. Suppose also $\epsilon = 0$ so that preferences are homothetic. Then changes in aggregate skill-bias A have no effect on sorting S.*

Proposition 5 tells us that even in the quantitative model, if preferences are homothetic then endogenous amenities do not independently link the skill premium to spatial sorting.