BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel—2024-50

Last updated: December 19, 2024

Consumer Search, Productivity Heterogeneity, Prices, Markups, and Pass-Through: Theory and Estimation

by Alex Chernoff,[1] Allen Head[2] and Beverly Lapham[2]

[1] Canadian Economic Analysis Department
Bank of Canada
AChernoff@bankofcanada.ca

[2] Queen's University
heada@queensu.ca, laphamb@queensu.ca

DOI: https://doi.org/10.34989/swp-2024-50 | **ISSN** 1701-9397

Acknowledgements

We thank seminar participants at the Society for Economic Measurement (2022), the Workshop of the Australasian Macroeconomics Society and Laboratory for Aggregate Economics and Finance (2023), Toronto Metropolitan University, Carleton University, the University of Saskatchewan, and the Canadian Economics Association (2024). The Social Sciences and Humanities Research Council provided funding for this research under Insight Grant 435-2016-0189. The views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

Abstract

We develop and estimate a search model in which identical consumers trade with price-setting firms that differ in productivity. In the model, equilibrium distributions of both prices and markups are non-degenerate and continuous with a firm's price decreasing as its productivity increases. Variation in markups across firms is more complicated and depends on the search process and the distribution of productivity, both of which are estimated using firm-level data on retail industries in Canada. We use the estimated model to characterize the qualitative and quantitative differences in prices and markups across firms. These differences stem from firm-level variation in demand elasticities driven by productivity heterogeneity and by imperfect information about prices. Additionally, we derive analytical expressions to determine how individual firm prices and markups respond to changes in cost and demand. This allows us to empirically analyze the heterogeneity in firms' pass-through of cost and demand shocks to prices and markups. Our findings reveal substantial heterogeneity in pass-through across firms, highlighting the distributional impact of shocks across consumers purchasing at different points of the price distribution. Finally, our analysis underscores the importance of accounting for individual firm price and markup adjustments to fully understand pass-through to average prices.

Topics: Inflation and prices; Service sector
JEL codes: E31, L16

Résumé

Nous élaborons et estimons un modèle de recherche dans lequel des consommateurs identiques font affaire avec des entreprises qui établissent les prix et dont la productivité diffère. Dans ce modèle, les distributions des prix et des marges à l'équilibre sont non dégénérées et continues; le prix d'une entreprise diminue à mesure que sa productivité augmente. La variation des marges entre les entreprises est plus complexe et dépend du processus de recherche et de la distribution de la productivité, qui sont tous deux estimés à l'aide de données sur les entreprises des secteurs du commerce de détail au Canada. Nous utilisons le modèle estimé pour caractériser les différences qualitatives et quantitatives dans les prix et les marges entre les entreprises. Ces différences découlent de la variation des élasticités de la demande entre les entreprises, due à l'hétérogénéité de la productivité et à l'imperfection de l'information sur les prix. En outre, nous dérivons des expressions analytiques pour déterminer comment les prix et les marges des entreprises individuelles réagissent aux variations des coûts et de la demande. Cela nous permet d'analyser empiriquement l'hétérogénéité de la répercussion des chocs de coût et de demande sur les prix et les marges des entreprises. Nos résultats révèlent une hétérogénéité substantielle dans cette répercussion entre les entreprises, ce qui met en évidence les effets différents des chocs sur les consommateurs qui achètent à différents points de la distribution des prix. Enfin, notre analyse souligne l’importance de tenir compte des ajustements de prix et de marges des entreprises individuelles pour comprendre pleinement la répercussion sur les prix moyens.

Sujets : Inflation et prix; Secteur des services
Codes JEL : E31, L16

1 Introduction

In this paper, we develop and estimate a consumer search model for an industry where firms, differing in productivity, compete by setting prices. Using detailed firm-level data from retail industries in Canada, we estimate jointly the parameters of both the productivity distribution and the consumer search process. We then apply the estimated model to study the factors driving variation across firms in prices, markups, and revenues. We also derive analytical expressions to characterize and estimate how individual firm prices and markups respond to cost and demand changes, thereby analyzing heterogeneity in firm-level price and markup pass-through.

A large literature studies firm heterogeneity and relates it to industry dynamics, mainly focused on manufacturing industries. Here we focus on retail trade, with firms interacting directly with imperfectly informed consumers. The consumer search framework in our theory is based on the *noisy search* environment of Burdett and Judd (1983), extended to allow for firm-level productivity heterogeneity in a manner similar to Herrenbrueck (2017) and Baggs et al. (2018). This approach provides a rich framework for exploring key issues related to variations in market power across firms and across economic conditions. In particular, we are able to characterize the extent to which search frictions permit relatively unproductive firms to survive in the market and to determine how their presence affects the prices and markups of more productive firms. We are also able to address questions about how relatively productive and unproductive firms respond differently to changes in cost and demand conditions, and explore the implications of these differing responses for consumers purchasing in different segments of the price distribution.

In the model, identical consumers search for opportunities to purchase a homogeneous product, knowing the distribution of such opportunities but receiving them randomly in the form of *price quotes* from a finite subset of firms. Meanwhile, an endogenous measure of potential firms pay a cost to draw an idiosyncratic productivity parameter from a fixed distribution. Given the search process of consumers and beliefs regarding the behavior of competing firms, all firms that can achieve non-negative profits operate, choosing prices optimally. Operating firms commit to meet the demand of the consumers who observe their posted price and for whom it is the lowest that they observe.

The search process determines the extent to which individual consumers are informed about the available trading possibilities. The distribution of productivity determines, for each firm, how many potential competitors have costs similar to its own. These components interact to jointly determine the characteristics of equilibrium distributions of posted prices, realized markups, and revenue, all of which are non-degenerate and continuous on a connected support in equilibrium.

More precisely, in the search model, a firm's sales can be divided into two categories: those made to consumers with no alternative and those made to consumers with higher-price alternatives. When all firms have the same cost, as in the model of Burdett and Judd (1983), the only factor affecting the elasticity of demand a firm faces, and hence its markup, is the relative importance of these two groups of consumers. With productivity heterogeneity, there is an additional force at work. Firms with similar productivity post similar prices, and firms in denser regions of the firm productivity distribution face a higher risk of losing sales to nearby firms. Thus, both the fraction of consumers with no alternatives and the firm productivity distribution affect firm's pricing decisions. While prices fall monotonically with firm productivity, markups can vary non-monotonically in a complex manner that reflects both of these factors.

To obtain empirically relevant implications of the theory for firm behavior at both the individual and industry level, we estimate the parameters of both the search process and productivity distribution. We do this using detailed firm-level data on revenue, costs of goods sold, and profits for eleven retail sub-sectors in Canada over the period 2001–2013. We develop a novel structural estimation approach and illustrate the effectiveness of this estimation procedure through Monte Carlo exercises.

Several common patterns emerge when comparing our estimates of the search process across retail sub-sectors. A significant but relatively small fraction of consumers (between 5% and 18% across sub-sectors) observe only a single price. In all sub-sectors, the most frequently observed number of price quotes is two, and at least 59% of consumers receive three or fewer price quotes. Furthermore, in all cases, a small but significant fraction of consumers are "well informed," receiving a large number of quotes. With regard to the distribution of productivity, while our theoretical and empirical framework accommodates a general class of productivity distributions, we restrict attention to Pareto distributions, and our estimates of shape parameters range between 2.5 and 6.5. These values are in line with estimates in the literature based on manufacturing data.

Using our estimated parameters, we characterize numerically the implied distributions of prices and markups for the eleven retail sub-sectors in our data. In all sub-sectors, firm markups are increasing in firm size through most of the distribution. In all but two sub-sectors, markups are decreasing in firm size at the low end of the productivity distribution, generating a u-shaped relationship between markups and firm size. These results are driven by the non-monotonicity in the firm-level demand elasticities resulting from the interaction between search and productivity heterogeneity considerations, as described previously. We provide supplementary reduced-form empirical evidence of the presence of this u-shaped relationship in our retail sub-sectors.

We also study firm-level price and markup responses to changes in the economic environment. In particular, given the search process and extent of productivity heterogeneity, firms respond differently to changes in both production costs and demand conditions. In discussing these reactions, we distinguish between those which hold industry composition constant (*short-run* responses) and those taking account of firm entry and exit (*long-run* responses). In both cases, we characterize a rich pattern of price and markup responses at both the firm and aggregate levels. Again, variation in the elasticity of demand drives differences in the responsiveness of firm-level prices and markups throughout the productivity distribution.

More specifically, in our estimated model, small firms tend to change their prices more in response to a change in their *own* cost than do large firms in the short and long run. In fact, the smallest firms tend to exhibit more than complete pass-through. This occurs because low-productivity firms generally face a higher elasticity of demand than do high-productivity firms.

This finding, that large retail firms exhibit relatively low levels of pass-through of a change in their own cost while small firms fully pass-through those changes, is consistent with the empirical findings from studies using manufacturing data.

In contrast, while short-run pass-through of *common* marginal cost movements to prices is incomplete for all firms, it is the larger, high-productivity ones that respond more strongly, passing through more of a common cost change than their smaller low-productivity competitors. This is intuitive as the latter price close to the consumers' reservation price and thus have less incentive to change prices than do their higher-productivity competitors who have higher markups. In the long run (i.e. with entry and exit), firms' responses to a common cost shock is identical to their short-run response to an own cost shock. Hence, long-run price responses when all firms' costs change are characterized by stronger responses by smaller firms than by larger firms.

Turning to a change in demand conditions, we show that short-run price pass-through of an increase in consumers' common reservation price is incomplete and is decreasing in firm size. This is consistent with the intuition afforded by the model: when the reservation price increases, small, low-productivity firms have fewer competitors pricing above them and thus raise their prices significantly. Their larger, high-productivity competitors, in contrast, restrain their price increases in order to not lose sales to their competitors who were pricing above them. We also examine long-run responses of prices to changes in the reservation price and changes in fixed costs and demonstrate that the direction and strength of firms' responses depends on firm productivity.

While our main focus is on variations in firm-level responses, we also demonstrate that the direction and degree of pass-through of cost and demand changes to *average* prices differs with the type of shock and between the short and long run. In the short run, common cost movements are passed through more strongly than are changes in the reservation price, but this pattern is reversed in the long run. Because consumers purchase only one unit but receive quotes from multiple retailers, our framework allows us to distinguish between the pass-through to average posted and transaction prices. Whether transaction prices or posted prices respond more strongly depends on variation in pass-through and market share across firms. Therefore, our analysis suggests that capturing the price and markup adjustments of individual firms is essential for understanding pass-through to average prices.

Our analysis and results contribute to the literature on search, price dispersion, pass-through, and inflation, following in the vein of Head et al. (2010), Head et al. (2012), and Wang et al. (2020). Our work here is distinguished in this literature by the inclusion of firm heterogeneity, estimation of the productivity distribution and search process, and the focus on heterogeneity in both the levels and responsiveness to shocks of prices and markups. Others have extended the framework of Burdett and Judd (1983) in related but different directions. Examples include Kaplan et al. (2019), Menzio and Trachter (2018), and Menzio (2023).

In recent work, Menzio (2024) considers theoretical implications for prices and markups in a model with a Poisson search process and firm heterogeneity. Our work differs in that we generalize the search process and estimate the model using firm-level data and consider pass-through. Firm heterogeneity is also present in the theoretical work of Herrenbrueck (2017) and both theoretical and empirical work of Baggs et al. (2018). Both papers focus on international issues, and the latter uses firm-level data but takes a different estimation approach, exploiting cross-border travel and exchange rate movements.

We also view our approach and results as complementary to those in the large literature on industry dynamics following Hopenhayn (1992), Melitz (2003), and others, but our approach differs in significant ways. First, imperfect competition in our framework stems from search frictions, specifically consumers' incomplete information regarding trading opportunities, rather than from product differentiation. Second, markups here are endogenous, are heterogeneous across firms, and respond differentially to a wide range of parameter changes and shocks. Moreover, the overall degree of market power is determined by the distribution of markups, which in turn is driven by the properties of demand which depend on both the search process and the entire distribution of firm productivity.[1] Third, we focus empirically on retail trade rather than manufacturing, and our pass-through results speak directly to changes in consumer prices.

The remainder of the paper is organized as follows. Section 2 describes the economic environment and the search equilibrium. In Section 3, we introduce a class of productivity distributions for which we illustrate analytically several properties of the theoretical model with regard to heterogeneity in prices, markups, and the responses of both to cost and demand changes. Section 4 presents our estimation procedure and results of Monte Carlo exercises. Section 5 describes the data used in estimation, and Section 6 the estimation results. In Section 7 we examine variation in prices and markups across firms in the estimated model. In Section 8, we analyze firm-level differences in pass-through to prices and markups of cost and demand changes. Section 9 concludes.

2 Theoretical Environment

2.1 Retailers and Households

We consider a model of an industry comprised of firms that we refer to as *retailers*. The structure of this industry is similar to that studied by Burdett and Judd (1983), in that retailers sell a homogeneous good and households (consumers) have incomplete information regarding prices. We extend the environment developed by Burdett and Judd (1983) in two primary directions: retailers have heterogeneous technologies, and the measure of active retailers is endogenous.

There is an exogenous measure, λ, of *ex-ante* identical households who consume the homogeneous good. Households do not observe all prices, but rather each receives a random number of price quotes from retailers. Let $q_k \in [0, 1]$ for $k \in \{1, 2, ..., K\}$ denote the exogenous probability with which a randomly selected household observes k prices before purchasing, where $\sum_{k=1}^{K} q_k = 1$. A household purchases one unit of the good from the retailer who is posting the lowest price that

[1] Other frameworks in the literature accommodate variable markups, with many building on the models of Atkeson and Burstein (2008) and Melitz and Ottaviano (2008). In these frameworks, market power and endogenous markups stem from product differentiation rather than incomplete information and search.

it observes if that price is less than or equal to the exogenous common reservation price, $\tilde{p}$.

There is a large measure of prospective retailers, each of which may pay a fixed cost, f_e, to draw a firm-specific productivity parameter, z, from a common distribution, $J(z)$, which is continuous on a connected support, $[z_L, z_H]$, where $z_H = \infty$ is possible. The marginal cost to a retailer with productivity parameter z is constant and given by $\frac{\phi}{z}$, where $\phi > 0$ is common across firms.

To be *active*, a retailer that has paid for a productivity draw must pay a common fixed production cost, Ψ. This cost is endogenously determined and is proportional to the tightness of the market as described below. Previewing the equilibria we consider, there will be a cut-off productivity parameter, $\tilde{z}$, determined endogenously, such that only those retailers having productivity parameters greater than or equal to $\tilde{z}$ will produce. Hence, we denote the equilibrium distribution of productivity across *active* retailers by $F(z)$, where

$$F(z) \equiv \frac{J(z) - J(\tilde{z})}{1 - J(\tilde{z})}, \tag{1}$$

with support $[\tilde{z}, z_H]$.

2.2 Retailer Optimization

Given their productivity parameter and beliefs about their competitors' prices, each active retailer posts a price to maximize profits while committing to produce to meet demand. We describe retailers' optimal pricing problems in detail below. For now, let the endogenous cumulative distribution function of profit maximizing prices posted by retailers be denoted by $L(p)$. We assume for now and later show that $L(p)$ is continuous on a connected support, $[\underline{p}, \tilde{p}]$.

We denote the ratio of the measure of households (buyers) to the measure of active retailers (sellers) as $\mu \equiv \frac{\lambda}{N}$ and refer to μ as market tightness. Note that μ is endogenous as it is determined in equilibrium by the measure of active retailers, N. We assume that the common fixed cost of production, Ψ, is proportional to market tightness (i.e. $\Psi = \delta\mu$) where the degree of proportionality, $\delta > 0$, is exogenous. As described below, this assumption is consistent with the properties of fixed costs as measured in our data and is also convenient for estimation of the model. We interpret this feature as capturing the idea that markets with a relatively high ratio of buyers to sellers (μ) will be characterized by relatively high per-firm sales, thus requiring relatively high fixed and semi-variable cost expenditures on items such as rent and advertising.[2]

As noted above, households receive price quotes from a random sample of retailers. Given the probabilities with which each household receives k price quotes, the expected measure of households that receive a quote from any one retailer is given by $\mu \sum_{k=1}^{K} q_k k$. Hence, given the distribution of posted prices, $L(p)$, the total expected measure of households that see a retailer's posted price, p,

[2] Examples of other papers which allow for a positive correlation between market size and fixed operating costs include Arkolakis (2010) and Das et al. (2007).

and for which this is the lowest price they observe is given by

$$\mu A\left(1-L(p)\right), \tag{2}$$

where

$$A\big(1-L(p)\big) \equiv \sum_{k=1}^{K} q_k k(1-L(p))^{k-1}. \tag{3}$$

Note that as households purchase at most one unit, $\mu A(1-L(p))$ is also the total expected output in units for a retailer posting price p. For example, a firm that posts the maximum (i.e. household reservation) price, $\tilde{p}$, expects to sell to $\mu A(0) = \mu q_1$ households.

Now, given $L(p)$ and market tightness, μ, expected revenue and profit for a firm with productivity parameter z that posts price p are

$$R(p) = p\mu A\big(1-L(p)\big), \tag{4}$$

$$\Pi(z,p) = \left(p - \frac{\phi}{z}\right)\mu A\big(1-L(p)\big) - \Psi. \tag{5}$$

As noted above, each retailer chooses its price to maximize expected profits given beliefs about the prices of its competitors. The first-order condition for an interior solution to this problem determines the pricing function, $p(z)$, and is given by

$$A\left(1-L(p(z))\right) - \left[p(z) - \left(\frac{\phi}{z}\right)\right] A'\big(1-L(p(z))\big)L'(p(z)) = 0. \tag{6}$$

From (6), we note that retailers' prices are independent of market tightness, μ, and fixed costs, Ψ.

The following proposition states an important relationship between a firm's productivity parameter and its profit-maximizing price.

Proposition 1. *The retailer pricing function, $p(z)$, is monotonically decreasing in z.*

Proof: See Appendix B.

Proposition 1 establishes that retailers with a higher productivity parameter, z, and therefore a lower unit cost, optimally post lower prices. As prices are monotonically decreasing in z, we have $1-L(p(z)) = F(z)$ and $-L'(p(z))p'(z) = F'(z)$, where $F(z)$ is the productivity distribution given by (1). Note that the distribution of posted prices, $L(p)$, thus inherits certain properties from the distribution of productivity. For example, given the assumed properties of $J(z)$ and hence $F(z)$, the distribution of posted prices will be continuous on a connected support, $[\underline{p}, \tilde{p}]$.

Given the relationship between $L(p(z))$ and $F(z)$, we rewrite the first-order condition (6):

$$A\left(F(z)\right) + \left[p(z) - \left(\frac{\phi}{z}\right)\right]\left[\frac{A'\big(F(z)\big)F'(z)}{p'(z)}\right] = 0. \tag{7}$$

Recalling that $\tilde{p}$ is the (exogenous) maximum price and $\tilde{z}$ is the (endogenous) minimum productivity parameter, the solution to this first-order differential equation is

$$p(z) = \left(\frac{1}{A\big(F(z)\big)}\right)\left(\tilde{p}q_1 + \phi\int_{\tilde{z}}^{z} A'\big(F(x)\big)\left(\frac{F'(x)}{x}\right)dx\right). \tag{8}$$

For details of the solution, see Appendix A.

The markup of price over marginal cost for a firm with productivity parameter z is given by

$$mkup(z) \equiv \frac{p(z)}{\left(\frac{\phi}{z}\right)} = \left[\frac{z}{A\big(F(z)\big)}\right]\left[\frac{\tilde{p}q_1}{\phi} + \int_{\tilde{z}}^{z} A'\big(F(x)\big)\left(\frac{F'(x)}{x}\right)dx\right]. \tag{9}$$

Letting $elas(z)$ denote the negative of the elasticity of demand facing a firm with productivity parameter z, from (7), we have the following standard relationship here between the $elas(z)$ and the markup:

$$elas(z) \equiv -\left(\frac{\partial \ln\left(\mu A(1-L(p(z)))\right)}{\partial \ln\left(p(z)\right)}\right) = \frac{mkup(z)}{mkup(z)-1}. \tag{10}$$

Substituting the pricing function, (8), into the revenue function, (4), we can derive expected revenue as a function of z:

$$R(z) = \mu\left(\tilde{p}q_1 + \int_{\tilde{z}}^{z}\left(\frac{\phi}{x}\right)A'\left(F(x)\right)F'(x)dx\right). \tag{11}$$

We can also use (5) to derive profits as a function of z:

$$\Pi(z) = \mu\left(\tilde{p}q_1 + \int_{\tilde{z}}^{z}\left(\frac{\phi}{x}\right)A'\left(F(x)\right)F'(x)dx - \left(\frac{\phi}{z}\right)A(F(z))\right) - \Psi. \tag{12}$$

2.3 The Cutoff Productivity for Operation

Previewing equilibrium, the firm with the lowest productivity parameter, $\tilde{z}$, posts the maximum price, $\tilde{p}$, and earns zero profits. For such a firm, $A(F(\tilde{z})) = A(0) = q_1$. To guarantee that such a firm exists and that $\tilde{z} > z_L$, we assume that fixed costs, Ψ, satisfies

$$\left(\tilde{p} - \frac{\phi}{z_L}\right)\mu q_1 < \Psi < \tilde{p}\mu q_1. \tag{13}$$

The first inequality implies that fixed costs are high enough so that some potential firms will not find it profitable to operate: that is, $\tilde{z} > z_L$. The second inequality implies that fixed costs are low enough so that some potential firms will find it profitable to operate because the expected revenue from charging the reservation price is sufficient to cover their fixed costs.

Under these restrictions, the following zero-profit condition binds for a retailer with $\tilde{z}$:

$$\left(\tilde{p} - \frac{\phi}{\tilde{z}}\right)\mu q_1 - \Psi = 0, \tag{14}$$

Substituting $\Psi = \delta\mu$ into (14) allows us to determine the cutoff productivity for operation as a function of exogenous parameters:

$$\tilde{z} = \frac{\phi q_1}{\tilde{p} q_1 - \delta}. \tag{15}$$

Using (15) we derive the markup and the elasticity of demand facing the least productive firm:

$$mkup(\tilde{z}) = \frac{\tilde{p} q_1}{\tilde{p} q_1 - \delta} = \frac{\left[\frac{\tilde{p}\mu q_1}{\Psi}\right]}{\left[\frac{\tilde{p}\mu q_1}{\Psi}\right] - 1}, \tag{16}$$

$$elas(\tilde{z}) = \frac{\tilde{p} q_1}{\delta} = \frac{\tilde{p}\mu q_1}{\Psi}, \tag{17}$$

where we define $elas(\tilde{z})$ as the limit of (10) as z approaches $\tilde{z}$ from above.[3] These expressions clarify that the markup and the elasticity of demand for a firm with $\tilde{z}$ do *not* reflect any strategic considerations. This is because the lowest level of productivity among operating firms is determined endogenously such that the expected variable profits emanating from that firm's exogenously determined price, $\tilde{p}$, exactly cover its fixed cost: $\left(\tilde{p} - \frac{\phi}{\tilde{z}}\right)\mu q_1 = \Psi$. Hence, the spread between that firm's price and its marginal cost depends *only* on the ratio between that firm's revenue and its fixed operating cost. This feature results from the assumptions of an exogenous reservation price and free entry and exit, resulting in zero profits for the lowest productivity firm that produces.

In accordance with intuition, the larger the spread between that firm's revenue and its fixed cost, the lower its markup must be to cover its fixed cost. In particular, a higher fraction of households which receive one quote, q_1, which one might link with more market power, is associated here with higher revenue and hence with a *lower* markup for this firm.

2.4 Free Entry

A prospective retailer must pay $f_e > 0$ to draw their idiosyncratic productivity parameter from the distribution given by $J(z)$ with support $[z_L, z_H]$. Under the restrictions in (13), in equilibrium, expected profits before entry will equal the cost of entry:

$$J(\tilde{z})0 + (1 - J(\tilde{z}))\bar{\Pi}(\tilde{z}) = f_e, \tag{18}$$

where $\bar{\Pi}(\tilde{z})$ is expected profits as a function of the cutoff productivity for operation.

Using (12) and recalling that $\mu = \lambda/N$, we can rewrite the free entry condition (18) as

$$N = \left(\frac{\lambda}{f_e}\right)\left[(\tilde{p}q_1 - \delta)(1 - J(\tilde{z})) + \phi\left(\int_{\tilde{z}}^{z_H}\left(\int_{\tilde{z}}^{z}\left(\frac{A'(F(x))F'(x)}{x}\right)dx - \frac{A(F(z))}{z}\right)J'(z)dz\right)\right]. \tag{19}$$

Substituting in $\tilde{z}$ from (15) into this expression gives the equilibrium measure of active retailers, N, as a function of exogenous parameters.

[3] This elasticity is valid only for a price decrease. If this firm raises its price above $\tilde{p}$, their sales drop to zero.

2.5 The Role of Productivity Heterogeneity

Here we briefly compare prices and markups in our model to their analogs in an environment with noisy household search and *ex-ante* homogeneous firms similar to that studied by Burdett and Judd (1983) to provide a better understanding of the role of firm-level productivity heterogeneity in our model. In the Burdett and Judd environment, price and markup dispersion exists solely because households have incomplete information about prices. In our model, productivity heterogeneity also contributes to variation in prices and markups across firms. For comparative purposes, we consider a homogeneous firm model in which all firms have marginal cost $\frac{\phi}{\tilde{z}}$ and where the parameters of the price quote distribution are identical to those in the model with productivity heterogeneity.

Let $\hat{L}(\hat{p})$ denote the equilibrium cumulative distribution function of profit-maximizing prices in the model with homogeneous firms. We evaluate differences in prices between the two models point-wise at the same percentile values, which implies that we compare $\hat{p}$ from the homogeneous firm model to $p(z)$ from the heterogeneous firm model at points where $1-\hat{L}(\hat{p}) = F(z)$. In Appendix A, we demonstrate the following relationship between those prices:

$$p(z) = \hat{p} + \left(\frac{(\phi/\tilde{z})}{A(F(z))}\right) \int_{\tilde{z}}^{z} A'(F(x))F'(x) \left(\frac{\tilde{z}}{x} - 1\right) dx. \tag{20}$$

Using (20), we see that at all points in the percentiles of the pricing distribution except at the reservation price, the price in the heterogeneous firm environment is below that in the homogeneous firm model. This is perhaps to be expected since all but the least productive firm in the heterogeneous environment have lower marginal costs than those in the homogeneous firm case. The strategic considerations for selecting an optimal price, however, differ significantly between the two environments, which could have made it challenging to predict the direction of price differences.

Comparing markups across the two models at the same point in the price distribution, we derive the following in Appendix A:

$$mkup(z) = \widehat{mkup} + \left(\frac{1}{A(F(z))}\right) \left(mkup(\tilde{z})q_1 \left(\frac{z}{\tilde{z}} - 1\right) + \int_{\tilde{z}}^{z} A'(F(x))F'(x) \left(\frac{z}{x} - 1\right) dx\right), \tag{21}$$

where $\widehat{mkup} = \hat{p}/(\phi/\tilde{z})$ is the markup in the homogeneous firm model at $\hat{p}$ where $1 - \hat{L}(\hat{p}) = F(z)$. Upon inspection of (21), we note that markups are higher in the heterogeneous model at all points in the percentiles of the pricing distribution except at $\tilde{p}$. This occurs because firm heterogeneity strengthens the effects on markups which emanate from buyers' incomplete information. Specifically, the existence of lower-productivity firms in the market posting relatively high prices raises the expected alternative price posted by a firm's competitors. This enables them to realize higher markups than they would in the absence of productivity heterogeneity.

Another important difference between the two models is that in the homogeneous firm model, markups are negatively related to revenue whereas with productivity heterogeneity, as we demonstrate below, markups are either positively related to revenue or exhibit a u-shaped relationship. We provide empirical evidence below showing that markups measured by the ratio of revenue to variable costs vary non-monotonically with revenue. This evidence aligns with the model that accounts for productivity heterogeneity but is inconsistent with the homogeneous firm environment.

3 A Specialized Theoretical Framework

3.1 A Class of Productivity Distributions

In order to give a more detailed account of prices, markups, revenue, and profits in equilibrium; to provide a framework for structural estimation of the model; and to more fully characterize the responses of prices and markups to changes in economy parameters, we analyze the theoretical environment for a specific class of productivity distributions. Specifically, we consider an environment in which the cumulative distribution function (CDF) for productivity parameters can be written as a monotonic function of the ratio $\tilde{z}/z$ only. We will refer to this as the *monotone-$\tilde{z}$-ratio property.* Distributions in this class have CDFs for which there exists a monotonic function, $G(\tilde{z}/z)$, such that $F(z;\tilde{z}) = G(\tilde{z}/z)$. Here, for clarity, we have denoted the CDF of the productivity distribution of active retailers to explicitly account for its dependence on $\tilde{z}$. We also have $F'(z;\tilde{z}) = -G'(\tilde{z}/z)\left(\tilde{z}/z^2\right)$.

Intuitively, for this class of distributions, a retailer's productivity *relative* to the least productive retailer and, by extension, to other retailers is crucial for their pricing decisions in equilibrium. Distributions that satisfy this property include the bounded and unbounded Pareto distributions, the triangular distribution, and the uniform distribution.

This class of productivity distributions offers two specific advantages for our analysis. First, we are able to estimate five parameters of the structural model using firm-level micro-economic data for various retail sub-sectors. Our structural estimation methodology and parameter estimates are presented in Sections 4 and 6.

Second, we are able to prove several propositions analytically, thus providing sharp characterizations of the theoretical nature of price and markup responses to changes in economy parameters, highlighting how those responses vary across firms. In addition, because the five estimated parameters are sufficient to calculate firm-level markups and pass-through rates, we are able to provide estimates of those variables for the various retail sub-sectors we study. These empirical implications of our analysis are presented alongside the theoretical implications in Sections 7 and 8.

3.2 Equilibrium Price and Markup Functions

For this class of distributions, we may rewrite the pricing function, (8) as

$$p(z;\tilde{z}) = \left(\frac{1}{A\big(G(\tilde{z}/z)\big)}\right)\left(\tilde{p}q_1 - \phi\int_{\tilde{z}}^{z} A'\big(G(\tilde{z}/x)\big)G'(\tilde{z}/x)\left(\frac{\tilde{z}}{x^3}\right)dx\right). \tag{22}$$

Letting $u \equiv \tilde{z}/x$, we may perform a change of variable in the integral above to derive the following:

$$p(z;\tilde{z}) = \left(\frac{1}{A\big(G(\tilde{z}/z)\big)}\right)\left(\tilde{p}q_1 - \left(\frac{\phi}{\tilde{z}}\right)\int_{\tilde{z}/z}^{1} uA'\big(G(u)\big)G'(u)du\right). \tag{23}$$

Henceforth we index retailers by their *relative marginal cost*, $(\phi/z)/(\phi/\tilde{z}) = (\tilde{z}/z) = v$, allowing us to rewrite the pricing function (23) as

$$p(v) = \left(\frac{1}{A\big(G(v)\big)}\right)\left(\tilde{p}q_1 - \left(\frac{\phi}{\tilde{z}}\right)\int_{v}^{1} h(u)du\right), \tag{24}$$

where $h(u) = uA'(G(u))G'(u) < 0$.

We may also write firm markups as a function of v :

$$mkup(v) = \left(\frac{1}{vA(G(v))}\right)\left(\left(\frac{\tilde{p}}{\phi/\tilde{z}}\right)q_1 - \int_{v}^{1} h(u)d(u)\right). \tag{25}$$

This expression implies that all firms' markups are increasing functions of the markup of the least productive firm, $\frac{\tilde{p}}{\phi/\tilde{z}}$. It also indicates that other firms' markups may be greater than or less than the markup of the least productive firm. We also note that under free entry, the markup of the least productive firm is independent of $\tilde{z}$, as can be seen from (16). Hence, in the long run, a retailer's price depends on both its relative cost and the level of $\tilde{z}$, whereas a firm's markup depends on $\tilde{z}$ only through the retailer's relative cost.

In the next section, we turn to estimation of this specialized model. Deferring presentation of the complete theoretical implications of the model until after the estimation enables a more thorough discussion, allowing the theoretical results to be interpreted in light of the empirical findings. This sequence of analysis ensures that the conclusions we draw regarding price, markup, and pass-through heterogeneity are both theoretically sound and empirically validated.

4 Estimation Methodology

4.1 Links between Our Estimation Methodology and the Existing Literature

Our estimation methodology is related to an approach used in the international trade literature developed for estimating the parameters of a constant-elasticity-of-substitution (CES) demand model with monopolistically competitive firms with heterogeneous productivity. This methodology involves deriving an equation that relates the productivity CDF and model parameters to an observable revenue variable. Since firm-level revenue is monotonic in productivity in the CES demand model, the empirical distribution function (EDF) of revenue is used to proxy for the productivity CDF, and the methodology is implemented by regressing the log of revenue on a function of the revenue EDF. This approach is most commonly applied under the assumption that productivity follows a Pareto distribution, although Head et al. (2014) show how the method can also be applied when productivity follows a log-normal distribution.

In our setting, firm-level revenue is also monotonic in productivity and depends on the consumer search parameters of the model. As we show below, the theoretical revenue equation in our model can be transformed to yield an equation that remains monotonic in firm-level productivity and can be estimated by non-linear least squares using variables that are commonly available in firm-level microdata. Similar to the econometric framework used for the CES demand model, our approach relies on using the EDF of this transformed revenue equation to proxy for the productivity CDF.

4.2 Structural Equations Used to Estimate the Model

Starting from (11) and incorporating the zero-profit condition in (15), the revenue function can be rewritten:

$$R(z) = \frac{\mu\phi}{\tilde{z}}\left(q_1 + \sum_{k=2}^{K} q_k k(k-1)\int_{\tilde{z}}^{z}\left(\frac{\tilde{z}}{x}\right)F(x)^{k-2}F'(x)dx\right) + \Psi, \tag{26}$$

where we have used $A'(F(x)) = \sum_{k=1}^{K} q_k k(k-1)F(k)^{k-2}$ and $\Psi = \delta\mu$. Using this equation, we are able to derive the following and define the function $y(z)$:

$$y(z) \equiv log\left(\frac{R(z)-\Psi}{\Psi}\right) = log\left(\frac{\phi}{\delta\tilde{z}}\right) + log\left(q_1 + \sum_{k=2}^{K} q_k k(k-1)\int_{\tilde{z}}^{z}\left(\frac{\tilde{z}}{x}\right)F(x)^{k-2}F'(x)dx\right). \tag{27}$$

We have the following proposition:

Proposition 2. *The function $y(z)$ is monotonically increasing in z.*

Proof: See Appendix B.

An implication of Proposition 2 is that the CDF for z and y are equivalent, $F(z) = H(y)$. This result is particularly useful for productivity CDFs that satisfy the monotone-$\tilde{z}$-ratio property introduced in Section 3. Below we show how it is possible to represent the integral in equation (27) using the ratio $\tilde{z}/z$ and the model parameters, which allows us to calculate it numerically using the EDF of y.

Recall that for any distribution satisfying the monotone-$\tilde{z}$-ratio property, we can rewrite the CDF as $F(z) = G(\tilde{z}/z)$ and we have $F'(z) = -G'(\tilde{z}/z)\left(\tilde{z}/z^2\right)$. We can therefore rewrite (27):

$$y = log\left(\frac{\phi}{\delta\tilde{z}}\right) + log\left(q_1 + \sum_{k=2}^{K} q_k k(k-1)\int_{\tilde{z}}^{z}\left(\frac{\tilde{z}}{x}\right)G\left(\frac{\tilde{z}}{x}\right)^{k-2}G'\left(\frac{\tilde{z}}{x}\right)\left(\frac{-\tilde{z}}{x^2}\right)dx\right). \tag{28}$$

Applying the u-substitution integral rule with $u = \tilde{z}/x$ yields:

$$y = log\left(\frac{\phi}{\delta\tilde{z}}\right) + log\left(q_1 + \sum_{k=2}^{K} q_k k(k-1)\int_{1}^{\tilde{z}/z} uG(u)^{k-2}G'(u)\ du\right) \tag{29}$$

Equation (29) is the first of two structural equations used to estimate the model parameters.

The second structural equation corresponds to the ratio of revenue to variable costs. Using (27), we can write revenue: $R(z) = \Psi\left(1 + exp\left(y\right)\right)$. Next, note that variable costs can be written as follows:

$$C(z) = \left(\frac{\phi}{z}\right)\mu A\left(F(z)\right) = \Psi\left(\frac{\tilde{z}}{z}\right)\left(\frac{\phi}{\delta\tilde{z}}\right)\sum_{k=1}^{K} q_k k F(z)^{k-1}. \tag{30}$$

Using this final expression, we can write the revenue-cost-ratio as

$$\frac{R\left(z\right)}{C\left(z\right)} = \frac{\left(1 + exp\left(y\right)\right)}{\left(\frac{\tilde{z}}{z}\right)\left(\frac{\phi}{\delta\tilde{z}}\right)\sum_{k=1}^{K} q_k k F(z)^{k-1}}. \tag{31}$$

The parameters to be estimated include the search quote probabilities, $q_1, q_2, ..., q_K$, the parameters governing the productivity distribution, and the composite parameter $\phi/(\delta\tilde{z})$. Note that (29) and (31) depend only on these parameters and the ratio $\tilde{z}/z$. The usefulness of monotone-$\tilde{z}$-ratio property for our estimation methodology is now apparent. For any productivity distribution satisfying this property, we are able to invert the CDF and express $\tilde{z}/z$ as function of the CDF $F(z)$ or $H(y)$, by Proposition 2.[4]

4.3 Estimation Approach

Implementing our estimation strategy requires that we make parametric assumptions regarding both the distribution of firm-level productivity and the search process. With regard to the former, we follow a large number of papers in the firm-level heterogeneity literature in assuming that z follows a Pareto distribution with shape parameter γ. Under this assumption, we can write the empirical analogue of the first of our two structural equations as

$$y_i = log\left(\frac{\phi}{\delta\tilde{z}}\right) + log\left(q_1 + \sum_{k=2}^{K} q_k k(k-1)\gamma \int_{\left(1-\hat{H}(y_i)\right)^{1/\gamma}}^{1} u\left(1-u^{\gamma}\right)^{k-2} u^{\gamma-1}\, du\right) + \epsilon_{1,i}\ , \tag{32}$$

where $\hat{H}(y_i)$ is the EDF of y for firm i.

Note that y_i can be calculated with firm-level data on revenue and fixed costs. Alternatively, if the firm-level data available to the researcher includes revenue, variable costs, and profits, then fixed costs can be calculated using the definition of profits. Either way, sufficient data are needed to construct the variable y in (32). Following an approach used widely in the empirical international trade literature, we use the EDF of y, $\hat{H}(y)$ to approximate the CDF of y.[5]

[4]For example, in the case where productivity is distributed as a Pareto random variable, $\tilde{z}/z = (1 - F(z))^{1/\gamma} = (1 - H(y))^{1/\gamma}$.

[5]The standard definition of the EDF is $\hat{H}(y_i) \equiv \frac{1}{N}\sum_{j=1}^{N} I(y_j \leq y_i)$, where N is the sample size and $I(\cdot)$ is an indicator function that is 1 when the argument is true, and zero otherwise. In our application, we found that this formulation of the EDF was problematic as in some specifications $1 - \hat{H}(y_i)$ appears in the denominator or as the argument of the log function and results in $\pm\infty$ when $\hat{H}(\cdot)$ is evaluated at the maximum value in the sample. We therefore subtract $0.5/N$ from the standard definition, such that the EDF is strictly between 0 and 1 under our

With regard to the search process, in related work some researchers have assumed that the probabilities of observing different numbers of price quotes are distributed Poisson (see e.g. Baggs et al. (2018) and Mortensen (2005)). A more flexible approach is to assume a fully non-parametric distribution, allowing each q_k to vary between 0 and 1 subject only to the constraint that they collectively sum to 1. For example, Alessandria (2009), Head et al. (2010), and Head et al. (2012) all study models in which buyers observe one or two prices only.

Through experimentation with simulations, we have found that in order to account for the long right tail of the revenue distribution that is a well-documented characteristic in the literature on firm heterogeneity, some households need to be very well informed. That is, we require the maximum number of quotes observed, K, to be large. It is impractical to identify and estimate directly q_k for $k = 1, \ldots, K$ with K large. Moreover, we have we found the Poisson assumption to be overly restrictive. We therefore take the following approach. First we allow a fully non-parametric specification of q_1 and q_2, letting them vary freely between 0 and 1 subject to the constraint that $q_1 + q_2 < 1$. We then let K be infinity with the search quote probabilities for $k \geq 3$ declining at an exponential rate governed by the parameter ν. That is,

$$q_k = q\nu^{k-3} \qquad \text{for } k \geq 3,\ \nu \in (0,1) \text{ and } q \equiv (1 - q_1 - q_2)(1 - \nu). \tag{33}$$

We find this approach strikes a good balance between parsimony and flexibility.

Having thus specified the productivity and search quote distributions, we estimate the search parameters: q_1, q_2, and ν and the Pareto shape parameter γ in (32) by non-linear least squares (NLS).[6] This requires also an estimate of the composite parameter, $\phi/(\delta\tilde{z})$. We obtain an estimate for this parameter from our structural equation for the revenue-cost ratio, (31). The left-hand side of this equation can be calculated from firm-level data as the ratio of revenue to variable costs. Given estimates of q_1, q_2, ν, and γ from (32), the composite parameter is set to equate the average observed revenue-cost-ratio to the model-implied revenue-cost-ratio. We thus calculate a value of $\phi/(\delta\tilde{z})$ at each iteration within the non-linear estimation routine for (32).[7]

4.4 Monte Carlo Results

We test the properties of our estimator using a Monte Carlo exercise. We simulate 1,000 samples of size 10,000 and estimate for each sample the search parameters q_1, q_2, ν, the shape parameter, γ, and the composite parameter $log\left(\phi/\left(\delta\tilde{z}\right)\right)$ using the method described above. In generating the simulated data, we use a DGP that is similar to the estimates we report in Section 6 to test the

modified definition.

[6] Under the Pareto assumption for productivity, the integral in equation (32) is a Gaussian hyper-geometric function. This function is computationally intensive to calculate, and to do so efficiently we use numerical algorithms and code provided in Pearson et al. (2017).

[7] We also experimented with an alternative approach in which we estimated the composite parameter by least squares estimation at each iteration. In Monte Carlo experiments, however, we found that our approach described here performed better in that it produced estimates with a lower mean-squared error for each of the estimated parameters.

Table 1: Monte Carlo Results

Parameter	DGP	Mean	Bias	Std Dev.	RMSE	Min	Max
q_1	0.14980	0.14985	0.00005	0.00281	0.00281	0.13898	0.15857
q_2	0.51960	0.51932	-0.00028	0.00741	0.00741	0.48191	0.53969
ν	0.97050	0.97043	-0.00007	0.00122	0.00122	0.96651	0.97800
γ	2.21870	2.21820	-0.00050	0.02971	0.02970	1.97512	2.30879
$log(\frac{\phi}{\delta \tilde{z}})$	2.67340	2.67299	-0.00041	0.01483	0.01483	2.63161	2.75185

Notes: Monte Carlo results for estimator based on 1,000 simulations of the model, each with a sample size of 10,000. The DGP column shows the parameters used in generating the simulated data. Std Dev. denotes standard deviation and RMSE is the root-mean-squared error.

performance of our estimation methodology in the neighborhood of the parameter space that best fits the data we use.

The Monte Carlo results are reported in Table 1. For each of the estimated parameters the bias, standard deviation, and root-mean-squared error of the estimates are relatively small. The range of the estimates is also reasonably tight around the true DGP values, as can be seen from the minimum and maximum values of the estimates across the simulations. In sum, the Monte Carlo results suggest that our estimator is consistent and capable of identifying the search, shape, and composite parameters of our model.

5 Data

We use annual firm-level data from 2001–2013 obtained from the National Accounts Longitudinal Microdata File (NALMF) database maintained by Statistics Canada. The NALMF database contains variables derived from a wide range of data files, including the Business Register and administrative data from the Canada Revenue Agency (CRA).

The variables that we use from the NALMF database are originally sourced from firms' filings of CRA Form T2125, "Statement of Business or Professional Activities." We use the following firm-level variables from NALMF, with the corresponding line from Form T2125 as indicated:

1. NALMF: total_revenue; Form T2125: Gross Business or Professional Income, Line 8299
2. NALMF: total_cost_of_sales; Form T2125: Cost of Goods Sold, Line 8158
3. NALMF: net_income_9369; Form T2125: Net Income (Loss) Before Adjustments, Line 9369

The first item in this list is our measure of firm revenue, R. Our measure of variable costs is the second item, the cost of goods sold, $COGS$. The third item in the list is our measure of profits, Π. Our measure of fixed costs, Ψ, is calculated from the definition of profits, $\Pi = R - COGS - \Psi$. We estimate the model parameters and analyze markups and pass-through for the 11 three-digit NAICS sub-sectors of the retail industry that are listed in Table 2.

Table 2: Retail sub-sectors

NAICS Code	Sub-Sector	Abbreviation
441	Motor Vehicle and Parts Dealers	MOTR
442	Furniture and Home Furnishing Stores	FURN
443	Electronics and Appliance Stores	ELEC
444	Building Material and Garden Equipment and Supplies Dealers	BLDG
445	Food and Beverage Stores	FOOD
446	Health and Personal Care Stores	HLTH
447	Gasoline Stations	GASS
448	Clothing and Clothing Accessory Stores	CLTH
451	Sporting Goods, Hobby, Book and Music Stores	SPRT
452	General Merchandise Stores	GENL
453	Miscellaneous Store Retailers	MISC

We take a number of steps to clean our data prior to estimation. For our measures of firm revenue, costs of goods sold, profit, and fixed costs, we drop any observations that have non-positive or missing values. We also drop observations from the following NAICS four-digit industry groups: 4413, 4452, 4453, 4531, 4533, and 4539. These industry groups are substantively different from other industry groups in their sub-sector and therefore were excluded as they are unlikely to be in the same market.[8] We also drop outlier observations for the key variables used in our analysis. For y and our fixed cost variable, Ψ, we define outliers as observations with values above the 98th percentile, within each year and sub-sector. For the revenue-cost ratio, we define outliers as values below 1 or greater than 10.

Statistics Canada's vetting policy of analysis of business microdata stipulates that summary statistics are to be only released in the final stage of an empirical project to minimize the risk of residual disclosure. Therefore, summary statistics for the variables used in our analysis will be provided in a future draft of the paper.

6 Estimation Results

In this section, we present estimates of the parameters of the consumer search process and productivity distribution as well as the composite parameter, $log\left(\phi/\left(\delta\tilde{z}\right)\right)$, for each sub-sector in Table 2. When estimating the model, we allow the composite parameter to vary for each year in our sample, 2001–2013. We report, however, only the average value of the estimates of this parameter over the 13 years of our sample. The consumer search and productivity shape parameters are assumed to

[8] For example, the three industry groups in sub-sector 441 are *Automotive dealers* (4411), *Other motor vehicle dealers* (4412), and *Automotive parts, accessories and tire retailers* (4413). Our definition of sub-sector 441 excludes 4413 as it is unlikely that consumers search in the same manner for automotive parts and tires as they would for cars and other motor vehicles. The names of the other excluded industry groups are *Speciality food stores* (4452), *Beer, wine and liquor retailers* (4453), *Florist* (4531), *Used merchandise stores* (4533), and *Other miscellaneous store retailers* (4539).

be constant over time.[9] We calculate bootstrap standard errors for the parameter estimates.[10]

Our estimation results are presented in Table 3. The search quote parameters, q_1, q_2, and ν are qualitatively similar across each sub-sector: estimates of q_1 are considerably less than those for q_2, and ν estimates indicate a long right tail. The estimate of q_1 ranges between 0.0503 and 0.1842, reflective of a consumer search process where relatively few households contact only one firm. The estimate of q_2 has greater variation across sub-sectors, ranging from a minimum of roughly one half to a maximum of about three quarters. Note that with the estimates of q_1 being low in each sub-sector and the assumed exponential tail to the search quote distribution, for all sub-sectors the estimated modal number of contacts is two. Overall, our estimates indicate that a small fraction of households are not well informed (i.e. receive a single price quote), at least some households are very well informed (i.e. receive many price quotes), and a strong majority of households observe three or fewer quotes (see Panel C of Table 4).

For the productivity shape parameter, the estimates range from 2.5054 to 6.4586 and are similar to those found in the empirical trade literature, typically using manufacturing data, and imposing constant-elasticity-of-substitution preferences with fully informed consumers. For example, Melitz and Redding (2015) analyze a model of heterogeneous firms with a value of γ=4.25 for the Pareto shape parameter and a value of 4.0 for the elasticity of substitution parameter.[11]

In Table 4 we provide additional economic interpretation of our results by presenting the implications of our estimates for markups, relative marginal costs, and the consumer search process through the lens of our model. Using our parameter estimates and (9), we calculate model-implied gross markups for firms across the productivity distribution. The average markup ranges from about 1.10 to 1.41 across sub-sectors. In each sub-sector, the median markup is less than the mean markup, implying that the model-implied markup distribution is right skewed. The minimum and maximum markups suggest a moderate spread in the range of markups across firms of different productivity levels.

Examining the markup of the least productive firm (at $\tilde{z}$), note that in all but two sub-sectors (441 and 447), this markup is above the minimum, indicating that generally markups are *not* monotonic in revenue. Furthermore, in some sub-sectors, such as gasoline stations (447), the least productive firm's markup is quite low. Recall from (16) that the least productive firm's markup is

[9]If Statistics Canada's vetting procedure allows us to disclose the year-by-year estimates of the composite parameter, we will report these estimates in a future update to the paper. If possible, we will also report a robustness check showing the extent to which the consumer search and productivity shape parameters vary over time.

[10]Because the model is computationally intensive to estimate and we only have a limited number of hours at Statistics Canada Research Data Centre for this project, we used 100 pairs bootstrap samples when calculating the standard errors. In the final draft we will increase the number of bootstrap samples to 1000. However, we expect little change in our standard errors from changing the number of bootstrap samples, as we compared and saw little difference in the standard errors whether they were calculated from 50 or 100 samples.

[11]In that literature there is a technical condition that $\gamma > \sigma - 1$, where σ is the elasticity of substitution. This condition and the choice of σ have an important bearing on the minimum value of the Pareto shape parameter that is admissible to the model. Using firm-level sales data, it is common for researchers to estimate a value of the ratio $\gamma/(\sigma - 1)$ that is only slightly larger than 1, and only then for the right tail of the firm-level sales distribution where the Pareto assumption provides a good fit (see, for example, Head et al. (2014)). The value of the underlying Pareto productivity parameter depends on the choice of σ and the extent to which the left tail of the sales distribution is trimmed in the data prior to estimation.

Table 3: Model Estimation Results

	441 MOTR	442 FURN	443 ELEC	444 BLDG	445 FOOD	446 HLTH	447 GASS	448 CLTH	451 SPRT	452 GENL	453 MISC
q_1	0.0666	0.1531	0.0646	0.1495	0.1311	0.1047	0.0503	0.1842	0.1488	0.1137	0.1497
	(0.00149)	(0.002268)	(0.001358)	(0.002102)	(0.002589)	(0.001208)	(0.00101)	(0.003602)	(0.00274)	(0.002691)	(0.003329)
q_2	0.7757	0.6073	0.5040	0.6271	0.7641	0.7194	0.5432	0.5829	0.6161	0.6254	0.4800
	(0.002576)	(0.00335)	(0.012789)	(0.003032)	(0.011939)	(0.002335)	(0.004881)	(0.003023)	(0.003879)	(0.005064)	(0.006504)
ν	0.9742	0.9553	0.9500	0.9623	0.9659	0.9671	0.8134	0.9651	0.9654	0.9382	0.9599
	(0.000414)	(0.000846)	(0.002297)	(0.000609)	(0.076822)	(0.000565)	(0.003704)	(0.000671)	(0.000675)	(0.001551)	(0.001149)
γ	5.0880	3.6036	2.5054	3.8391	6.4586	3.2253	5.9610	3.4764	3.3591	4.1607	2.9015
	(0.076243)	(0.04041)	(0.030063)	(0.040183)	(0.105782)	(0.020294)	(0.116195)	(0.039675)	(0.037515)	(0.064525)	(0.050005)
$log\,(\phi/(\delta\tilde{z}))$	8.1485	2.8072	4.1035	3.7799	5.4322	3.1874	7.6541	2.2540	2.9580	3.8510	2.6174
	(0.036067)	(0.012636)	(0.061705)	(0.01655)	(0.033893)	(0.009538)	(0.029498)	(0.009459)	(0.014808)	(0.019767)	(0.030069)
SSR	1,138.34	425.66	369.76	732.59	1,432.61	749.47	354.66	849.29	509.02	318.31	238.32
Observations	66,155	40,540	36,615	50,365	69,120	80,505	50,535	63,995	37,325	24,105	16,875

Notes: Bootstrap standard errors are in parentheses. The composite parameter, $log\,(\phi/(\delta\tilde{z}))$, is estimated year by year and the reported value is the average.

decreasing in the spread between this firm's revenue and its fixed operating cost. Accordingly, the low markup in sub-sector 447 implies a relatively large difference between revenue and fixed cost for this firm, and a relatively high elasticity of demand equal to 218.39.

In other sub-sectors, such as clothing stores (448), the least-productive firm posts a significantly higher markup indicating a relatively high level of fixed costs and a relatively low elasticity of demand equal to 2.79. In the next section, we more fully explore the properties of markup heterogeneity across firms with different productivity levels.

Panel B in Table 4 presents the relative marginal cost for firms with productivity level corresponding to the 99.99^{th} percentile. This variable is determined by the degree of productivity dispersion across firms, and variations in estimates of γ lead to significant differences in this measure of relative marginal cost across sub-sectors.[12] For example, a clothing store with productivity at the 99.99^{th} percentile has marginal costs that are only 7 percent of those of the least productive firm while the corresponding number for grocery stores is 24 percent. In the analysis in the next section at the sub-sector level, we examine simulated samples for firms with relative marginal costs that range between the relative marginal cost given in this table for a highly productive firm and the relative marginal cost of the least productive firm.

In Panel C of Table 4, we present the implications of our estimates for the consumer search process. As noted above, we report the percentage of consumers obtaining three or fewer quotes for each sub-sector. We interpret a smaller number here as indicating a sub-sector where consumers are relatively better informed of the range of prices. For example, the estimates suggest that consumers are relatively well informed of prices at gasoline stations (447). This is interesting as the common practice of gas stations posting prices on road-facing signage suggests that at any given point in time, a motorist is likely to have seen several prices and thus be relatively well informed.

Similarly, the estimates suggest that consumers visiting electronics and appliance stores are also relatively well informed, a result that may reflect consumers searching more intensively for low prices in a sector where big-ticket items are sold. Consumers at grocery stores, on the other hand,

[12] Here we have $(\tilde{z}/z_{99.99}) = (1 - .9999)^{\frac{1}{\gamma}}$.

Table 4: Model Implied Markups, Relative Marginal Cost, and Consumer Search

	441 MOTR	442 FURN	443 ELEC	444 BLDG	445 FOOD	446 HLTH	447 GASS	448 CLTH	451 SPRT	452 GENL	453 MISC
A. Markups											
Mean	1.1328	1.3169	1.3136	1.2286	1.1156	1.3132	1.1022	1.4096	1.3207	1.2062	1.4073
Median	1.0849	1.2546	1.2136	1.1566	1.0744	1.2228	1.0671	1.3507	1.2471	1.1475	1.3430
Min	1.0036	1.1931	1.1140	1.0988	1.0262	1.1473	1.0046	1.2825	1.1808	1.0952	1.2685
Max	3.0940	5.7345	11.5433	4.8650	2.6803	6.8734	3.3377	5.7363	6.0112	4.7748	7.7653
At $\tilde{z}$	1.0036	1.3843	1.2221	1.1468	1.0313	1.3625	1.0046	1.5563	1.3421	1.1792	1.4715
B. Relative Marginal Cost (%)											
$100 \times (\tilde{z}/z_{99.99})$	16.36%	7.76%	2.53%	9.08%	24.03%	5.75%	21.33%	7.07%	6.44%	10.93%	4.18%
C. Consumer Search											
% with 3 or fewer quotes	84.64%	77.11%	59.01%	78.50%	89.88%	82.99%	66.93%	77.52%	77.30%	75.52%	64.46%

Notes: Markup statistics are based on model-implied estimated markups for firms with productivity parameters ranging between $\tilde{z}$ and $z_{99.99}$, where $z_{99.99}$ is the value of the productivity parameter at the 99.99^{th} productivity percentile. For each sub-sector, the estimated value for α used in the markup calculations equals the average across years of the estimates for that sub-sector of the anti-log of the composite parameter from equation (32). Relative Marginal Cost reports the marginal cost of the low-cost firm at the 99.99^{th} productivity percentile as a percentage of the marginal cost of the highest cost firm.

are the least likely to get more than three quotes, a result that may reflect convenience shopping by households when picking up everyday essentials. While it is somewhat surprising that our results suggest relatively less search across motor vehicle dealers, this may in fact reflect brand loyalty among car buyers to particular automotive manufacturers.

Finally, we observe a negative correlation across sub-sectors between the mean or median markup and the fraction of households which receive three or fewer quotes.[13] This aligns with intuition: markups tend to be higher in sub-sectors where households are less well informed about prices, potentially reflecting a higher degree of market power. In sum, the implications of our estimates for the consumer search process are generally consistent with basic economic intuition and the nature of products bought and sold in different retail sub-sectors.

7 Prices, Markups, and Revenues

In this section we analyze further the theoretical and empirical implications of the model using the estimated structural parameters from the previous section. Specifically, we examine the properties of equilibrium prices, markups, and revenues for a specific sub-sector, clothing and clothing accessory stores (sub-sector 448). Focusing the discussion on a single sub-sector serves to elucidate the qualitative properties of the equilibrium functions derived in Section 2, while also highlighting our quantitative findings for an important retail industry. We briefly discuss results for other sub-sectors and include more details for those industries in an appendix.

[13] The correlation coefficient for the mean markup equals -.36, while for the median markup it is -.33.

7.1 The Role of Productivity Heterogeneity

We begin by more thoroughly examining the role of productivity heterogeneity in affecting the qualitative and quantitative variation in both prices and markups across firms. We follow the same methodology as in Section 2.5 by comparing prices and markups in our model with those from a homogeneous firm model in which all firms have the same marginal cost, $\frac{\phi}{\tilde{z}}$, and where all relevant parameters are set to their estimated values for the clothing sub-sector.

In our model with firm-level cost heterogeneity, we follow the approach in Section 3.2 and present firm-level equilibrium variables as functions of $v = \frac{\tilde{z}}{z} \in [0, 1]$, i.e. of firms' relative costs.[14] Since estimated parameters alone do not determine price *levels*, we use (24) and (47) to calculate *scaled* prices in our model, $p(v)/\delta$, and in the model with homogeneous firms, $\hat{p}/\delta$, as functions of estimated parameters alone:

$$\frac{p(v)}{\delta} = \left(\frac{1}{A\big(G(v)\big)}\right)\left(1 + \alpha q_1 - \alpha \int_v^1 h(u)du\right) \qquad \frac{\hat{p}}{\delta} = \alpha + \frac{1}{A(1 - \hat{L}(\hat{p}))}. \tag{34}$$

Here $\alpha \equiv \frac{\phi}{\delta \tilde{z}}$ is the anti-log of the composite parameter from our econometric framework and $\hat{L}(\hat{p})$ is the endogenous CDF of equilibrium prices in the homogeneous firm model. The maximum scaled price in both models equals $\frac{\tilde{p}}{\delta} = \alpha + \frac{1}{q_1}$. The minimum scaled price in our model, $\frac{p(o)}{\delta} = \left(\frac{1}{A(1)}\right)\left(1 + \alpha q_1 - \alpha \int_o^1 h(u)du\right)$, is positive and less than the minimum scaled price in the homogeneous productivity model, $\alpha + \frac{1}{A(1)}$. In Appendix A, we derive analytical expressions for the probability density functions of the log of equilibrium prices for each model as functions of estimated parameters.

We can also express markups in each model as functions of estimated parameters alone:

$$mkup(v) = \frac{p(v)/\delta}{\alpha v} \qquad \widehat{mkup} = \frac{\hat{p}/\delta}{\alpha}. \tag{35}$$

In Appendix A, we present the analytical probability density function of the log of equilibrium markups for the homogeneous productivity model. It is not possible, however, to derive an analogous analytical expression for the distribution of markups in the heterogeneous firm model.

To compare prices and markups across the two models, Panels A and B of Figure 1 plot the log of the scaled price and log of the markup respectively for both models using estimates from the clothing sub-sector.[15] These figures depict those variables as a function of the percentiles of revenue for the homogeneous productivity model and as a function of the percentiles of productivity for the model with productivity heterogeneity. Thus, in both models, firms at the low end in the figures occupy low percentiles of the firm size distribution (measured by revenue) and *high* percentiles of the price distribution. The bottom two panels plot the probability density functions (PDFs) of the logs of the scaled prices (Panel C) and the histograms of the logs of the markups (Panel D) based

[14] Because we are assuming that z follows an unbounded Pareto distribution, the minimum value of v is $v = 0$.

[15] The estimated value for α used in this figure and all subsequent figures in the paper is the average across years of the estimates for the relevant sub-sector of the anti-log of the composite parameter from equation (32).

on simulations of each model.

In Figure 1, note that at lower percentiles of the firm size distribution, the pricing and markup functions are similar across the two models. In both models, since these firms are posting relatively high prices, a large share of their sales is made to households with no alternative and, hence, the firms face relatively low-demand elasticities compared to other (more productive) firms charging somewhat lower prices. The much larger fall in firm price as one moves up through the revenue distribution in the heterogeneous productivity model reflects the fact that firm-level costs are also falling. Cost heterogeneity contributes to greater price dispersion in that model relative to the homogeneous model as demonstrated by the PDFs in Panel C of Figure 1.

Importantly, the two models diverge significantly with regard to the behavior of markups across firms. In the homogeneous productivity economy, the only factor affecting the elasticity of demand as revenue rises is the search process through its effect on the relative importance of sales to consumers with no alternative. Because this share falls monotonically as price falls, the elasticity of demand *rises* and the markup *falls* monotonically with firm size, as depicted in Figure 1. This effect, emanating from incomplete information about prices, is also present in our model and has a dominant effect on markups for relatively small firms.

In the heterogeneous productivity economy, however, there is an additional force at work: the distribution of productivity. This determines, at each productivity level, the measure of firms with similar productivity who are thus choosing a relatively similar price. As productivity increases, not only does a firm's fraction of sales made to buyers with no alternative fall, but the measure of firms with similar productivity to the firm falls as well. This can be seen from the price PDF in Panel C of Figure 1 where, after a certain point, as prices decrease due to rising productivity, the measure of firms choosing similar prices falls, putting downward pressure on the elasticity of demand. This force initially mitigates and eventually overcomes the effect of the search process on the elasticity. In this model, over most of productivity distribution, the productivity effect dominates; the elasticity of demand *falls* while the markup *rises* with firm size, as depicted in Panel B of Figure 1.

7.2 Heterogeneous Markups

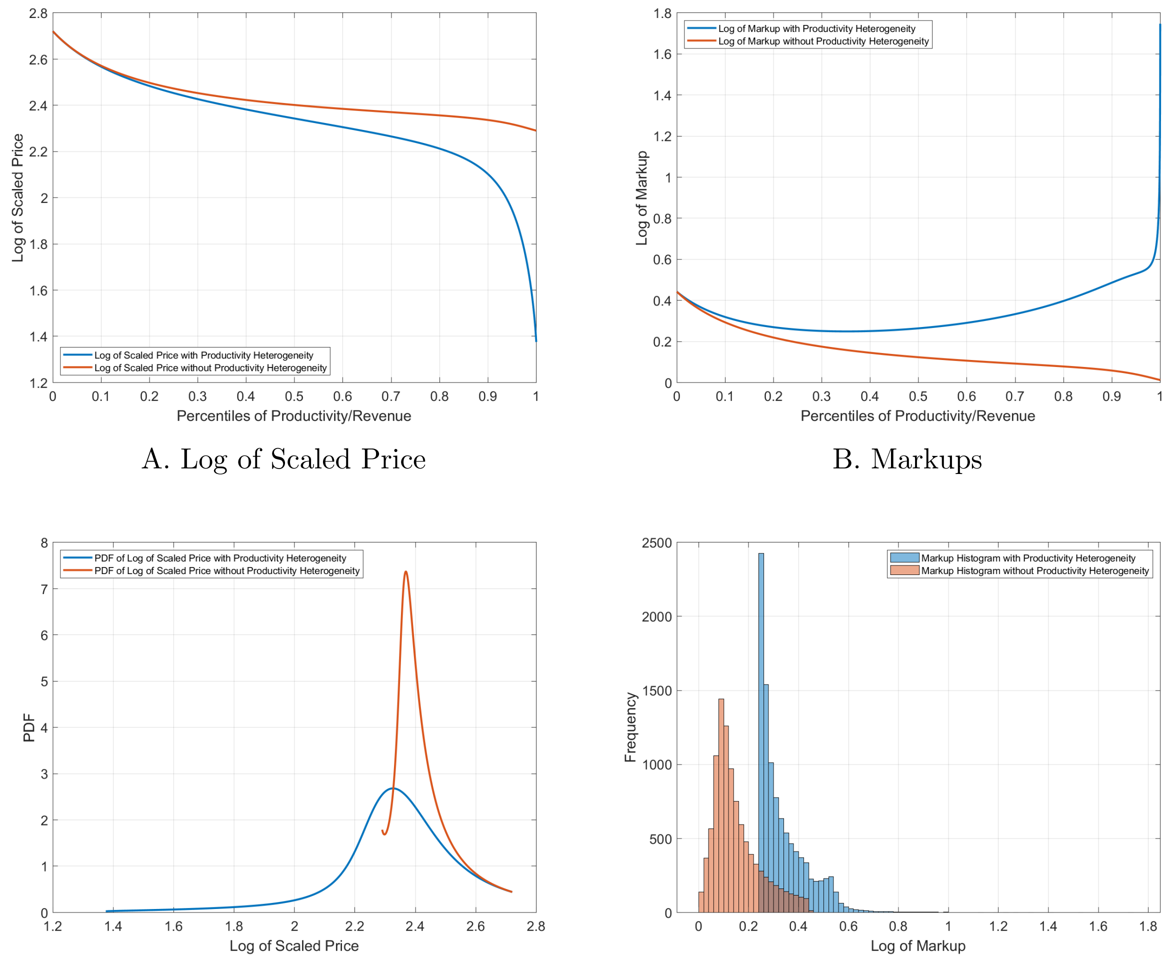
To more fully understand the role of productivity heterogeneity in affecting the behavior of firm-level markups in our model, we examine variation in the markup associated with a firm's relative productivity. The following proposition characterizes conditions under which firm-level markups do and do not vary monotonically with relative productivity:

Proposition 3. *The markup function is decreasing in relative productivity at $v = 1$ if and only if*

$$elas(1) < \left(-2G'(1)\right)\left(\frac{q_2}{q_1}\right), \tag{36}$$

where $elas(1)$ is the elasticity of demand of the least productive firm given by (17).

Figure 1: The Role of Firm Productivity Heterogeneity in Prices and Markups



A. Log of Scaled Price

B. Markups

C. PDF of the Log of Scaled Price

D. Histogram of the Log of Markup

Notes: In panels A and B, the log of the scaled price and log of the markup are plotted for each percentile of the revenue distribution for the model with homogeneous productivity, and for percentiles of productivity for the model with productivity heterogeneity. The plots use the same consumer search parameters, and the common marginal cost parameter in the homogeneous productivity model is set equal to the variable cost of the least productive firm in heterogeneous cost model , $\phi/\tilde{z}$. Scaled prices in both models are relative to the same value of the parameter δ. The plots use estimates from the clothing and clothing accessory sub-sector (448). Panels C and D display the probability density functions (PDFs) of the log of the scaled prices (Panel C) and the histograms of the log of the markups (Panel D) for each model.

Proof: See Appendix B.

Proposition 3 identifies a condition under which the markup is decreasing in productivity at the lowest level of z, $\tilde{z}$. Under this condition, the elasticity of demand facing the least productive firm is below that of its slightly more productive competitors. Recall that for this firm only, the elasticity of demand is completely determined by the ratio of the firm's expected revenue to fixed costs. In contrast, for all firms with productivity greater than $\tilde{z}$, the elasticity of demand depends on the search process and the distribution of firm productivity. Given this, the role of the parameters that determine if the least productive firms' markup is decreasing in relative productivity is intuitive. For example, if the ratio of the measure of households receiving two price quotes to that of households receiving only one is relatively high so that (36) is more likely to hold, then we expect other nearby firms to face a relatively high elasticity of demand because a low q_1 relative to q_2 confers them less market power. Similarly, if $-G'(1) = \gamma$ is relatively high, the condition is more likely to hold and we expect firms with slightly higher productivity than the least productive firm to face a relatively high elasticity of demand because there is a relatively high concentration of firms with similar productivity levels pricing close to it.

For relatively larger, more productive, firms, the markup is increasing in productivity as the effect of heterogeneous productivity on the elasticity of demand described above becomes the dominant factor. Furthermore, since the minimum price is positive, the markup approaches infinity as v approaches zero (i.e. as z approaches infinity). Combining all of these elements, we conclude that in the model with productivity heterogeneity, the markup will either have a u-shaped relationship with firm size or it will be monotonically increasing in firm size.

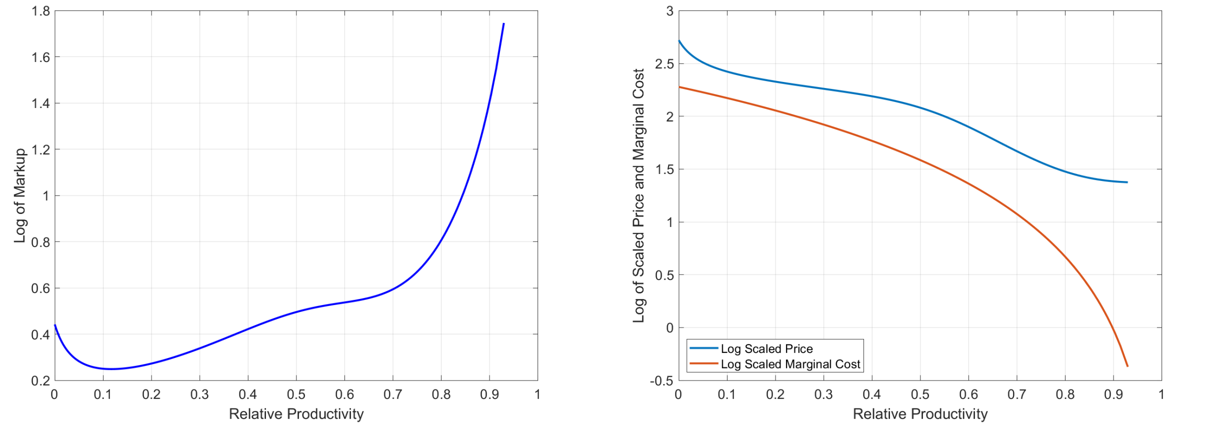
To illustrate these relationships for the clothing sub-sector, Figure 2 presents the log of markups (left panel) and a decomposition into the difference between the logs of scaled price, $p(v)/\delta$, and scaled marginal cost, $(\phi/z)/\delta = \alpha v$, (right panel). For clarity of presentation in this figure, we graph these variables as a function of a convenient measure of relative *productivity*, $(1 - v) = \frac{z-\tilde{z}}{z}$, where $(1 - v) \in [0, 1]$. This measure, when multiplied by 100, is interpreted as the percentage deviation in a firm's productivity from that of the least productive firm.

The figure demonstrates that in the clothing sub-sector, for low values of relative productivity, price declines more quickly than marginal cost and thus the markup falls with relative productivity.[16] At higher relative productivity, however (here for all firms with relative cost below roughly 90% of that of the least productive firm), marginal cost declines more quickly than price, and the markup is increasing.

Figure 9 in Appendix E presents analogous graphs for all the retail sub-sectors we study. There we see that markups are increasing in productivity across most of the productivity distribution in all sub-sectors, with a negative relationship occurring for low values of relative productivity in all but two industries (motor vehicles and gasoline stations). Thus, our estimated model generally exhibits a u-shaped relationship between firm-level markups and size as measured by revenue. More broadly,

[16] From the parameter estimates for the clothing sub-sector, we can determine that the condition in Proposition 3 is satisfied: $elas(1) = 2.80 < 21.20 = \gamma(q_2/q_1)$.

Figure 2: Log Markup Decomposition for Clothing Sub-Sector (448)



A. Log of Markups

B. Log of Scaled Price and Scaled Marginal Cost

Notes: The value of the log markup in the left panel equals the difference between the log of scaled price and the log of scaled marginal cost in the right panel. Scaled price and marginal cost are defined as those variables divided by δ, and the difference between these two variables is independent of δ. Estimates are for the clothing and clothing accessory sub-sector (448). For estimates of the other sub-sectors, see Figures 9 and 10 in Appendix E.

those figures and equation (25) indicate that the precise shape of the markup function depends in a complicated way on the parameters of the price quote distribution and the productivity distribution.

Table 5 provides empirical evidence consistent with the predicted u-shaped relationship in our model between firm-level markups and firm size as measured by revenue. In the table, we present estimates for each sub-sector from panel regressions of firm-level sales-to-cost ratios regressed on the EDF of revenue and the squared EDF of revenue, where we include time and firm fixed effects. For all sub-sectors, the estimated parameters on the revenue EDF and squared EDF are highly significant and negative and positive respectively, thus indicating evidence of a u-shaped relationship.[17]

The largely positive relationship we find between markups and firm size is consistent with both empirical and theoretical research with firm heterogeneity and varying endogenous markups. For example, Autor et al. (2020), De Loecker et al. (2020), and De Loecker and Warzynski (2012) use U.S. firm-level data to provide evidence that larger firms tend to charge higher markups than smaller ones. In much of the theoretical literature, researchers rationalize this result in environments with differentiated products using either a nested constant-elasticity-of-substitution demand model (see Atkeson and Burstein (2008), for example) or the linear demand system developed by Ottaviano et al. (2002) (see Melitz and Ottaviano (2008), for example). Our framework provides an alternative explanation for accounting for the positive relationship between size and markups, where market power arises from imperfect information rather than from product differentiation.

[17]We find the same qualitative results in OLS regressions without firm-level fixed effects.

Table 5: Regression of Sales-Cost Ratio on Revenue Percentiles

	441 MOTR	442 FURN	443 ELEC	444 BLDG	445 FOOD	446 HLTH	447 GASS	448 CLTH	451 SPRT	452 GENL	453 MISC
Revenue EDF	-3.478	-2.891	-3.151	-2.471	-2.237	-2.065	-2.409	-2.145	-2.279	-2.484	-2.431
	(0.0672)	(0.0886)	(0.123)	(0.0752)	(0.0436)	(0.0618)	(0.0506)	(0.0689)	(0.0930)	(0.105)	(0.148)
Revenue EDF Squared	2.361	1.888	1.623	1.638	1.461	1.135	1.388	1.428	1.463	1.537	1.484
	(0.0662)	(0.0883)	(0.121)	(0.0771)	(0.0475)	(0.0594)	(0.0498)	(0.0704)	(0.0922)	(0.121)	(0.150)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	66,155	40,540	36,615	50,365	69,120	80,505	50,535	63,995	37,325	24,105	16,875
R-squared	0.071	0.078	0.081	0.053	0.091	0.073	0.115	0.054	0.052	0.066	0.074

Notes: Standard errors are in parenthesis. All reported coefficients are statistically significant the 1% level.

7.3 Firm-Level Revenues

We now examine briefly some properties of firm-level revenues. We use (34) to derive a measure of scaled revenue as a function of a firm's relative cost, v, and estimated parameters:

$$\frac{R(v)}{\Psi} = \left(\frac{p(v)}{\delta}\right) A(G(v)) = 1 + \alpha q_1 - \alpha \int_v^1 h(u)du. \tag{37}$$

In Appendix A, we derive an analytical expression for the probability density function of the log of equilibrium revenues as a function of estimated parameters.

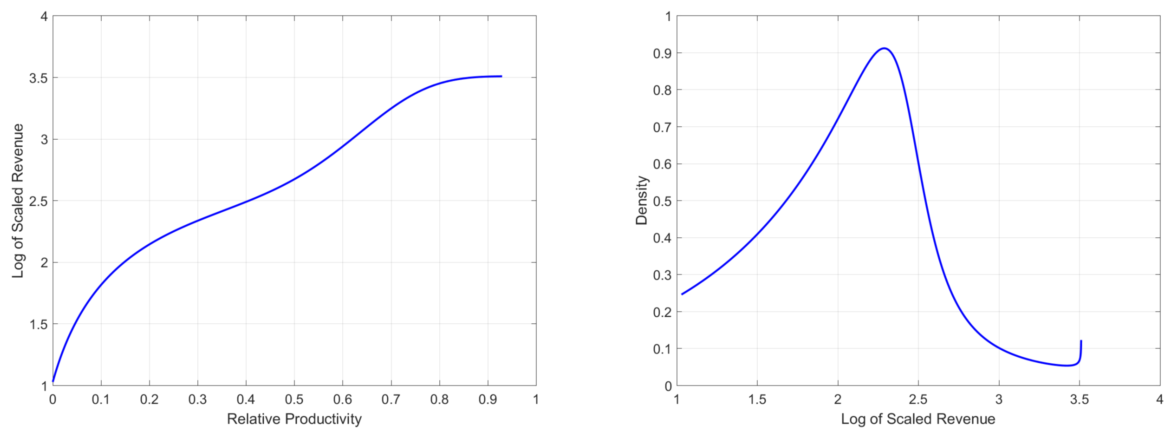
To demonstrate the properties of firm-level revenues for our estimated model, Panel A of Figure 3 presents the log of scaled revenue as a function of relative productivity, $1 - v$, using estimated parameters for the clothing sub-sector. Consistent with a robust prediction of our theoretical model, revenues are strictly increasing in relative productivity.

Panel B of the figure displays the probability density function for the log of scaled revenue implied by the estimates for this sub-sector. We see that the estimated model generates a broadly single-peaked density for the log of revenue.[18] Figure 11 indicates that this broadly single-peaked property is also apparent for estimated revenues for most other sub-sectors. As noted in the discussion on our estimation methodology in Section 4, this property is *not* an inherent feature of our theoretical model but rather depends on estimated parameters, especially those governing the search process. Notably, this feature of the revenue distribution is consistent with the actual empirical patterns we observe in all sub-sectors in our retail data and is a recurring pattern in manufacturing data.[19]

[18]The upturn in the density for the very highly productive firms occurs as these firms face very similar costs and demand conditions and so have very similar revenues.

[19]The model parameters are estimated using the distribution of a function of firm-level revenue, $y(z)$, as described above. We cannot display the actual empirical distribution of revenue given data release restrictions.

Figure 3: Log of Scaled Revenue



A. Log of Scaled Revenue D. PDF of Log of Scaled Revenue

Notes: Estimates are for the clothing and clothing accessory stores sub-sector (448). Scaled revenue is defined as implicitly divided by fixed costs, Ψ.

8 Firm-Level and Average Pass-Through

In this section, we characterize the response of firms' prices and markups to changes in four economy parameters: (i.) an individual firm's own productivity parameter, z; (ii.) the common cost parameter, ϕ; (iii.) the reservation price, $\tilde{p}$; and (iv.) the ratio of fixed costs to market tightness, δ. Movements in the first two parameters alter firm-level variable costs, changes in $\tilde{p}$ modify demand conditions, and changes in the last parameter, δ, can be interpreted as a shift in either cost or demand conditions.

We measure *pass-through* using the elasticities of prices and markups with respect to changes in economy parameters. We present the model's theoretical predictions for pass-through in a series of propositions and also calculate firm-level responses for our sub-sectors based on computed equilibria using parameter estimates from the preceding section. We focus primarily on variation in price and markup responses across firms, although at the end of the section we take up the implications of these responses for movements in average prices.

Throughout we focus on parameter changes that stimulate firm entry, i.e. reductions in cost or more favorable demand conditions.[20] We distinguish between the *short run*, in which we consider the responses of only incumbent firms without allowing for entry (i.e. no change in $\tilde{z}$ nor N), and the *long run*, in which we allow entry. In the cases we consider, the latter is characterized by a reduction in $\tilde{z}$ and thus an increase in firm heterogeneity through a decrease in the lower support of the firm productivity distribution.

[20] We focus on cost reductions and demand increases here so that there are no complications associated with firms desiring exit in the short run. Since we focus on elasticities, all of our results for incumbent firms are symmetric when changes in economy parameters lead to entry in the long run.

8.1 Short-Run Pass-Through

8.1.1 Firm-Specific Productivity Changes

We begin with a firm-specific productivity change; specifically, a reduction of $\frac{1}{z}$ for an individual retailer. We denote pass-through of this change to the price of a firm with relative marginal cost v as $SRPT^{1/z}_{p(v)}$ and note that $SRPT^{1/z}_{p(v)} = \epsilon^{v}_{p(v)}$, where the latter term is the elasticity of $p(v)$ with respect to v. We have the following proposition for price pass-through in response to a change in firm-specific productivity:

Proposition 4. *In response to a change in $\frac{1}{z}$, pass-through to a retailer's price is positive. That is, a retailer's price falls if its productivity increases.*

This proposition follows directly from Proposition 1.

To understand the forces determining the extent of pass-through in this case, note that from (25) we have the following relationships between price pass-through of a firm-specific productivity change, the markup, and the elasticity of demand for a firm with relative marginal cost v:

$$SRPT^{1/z}_{p(v)} = 1 + \epsilon^{1/z}_{mkup(v)} = 1 + \frac{\epsilon^{v}_{elas(v)}}{1 - elas(v)}. \tag{38}$$

Here $\epsilon^{1/z}_{mkup(v)}$ measures pass-through of a change in $1/z$ to the markup, $elas(v) = \frac{mkup(v)}{mkup(v)-1} > 1$ is the elasticity of demand, and $\epsilon^{v}_{elas(v)}$ measures pass-through of a change in v to the elasticity of demand. Consistent with intuition, pass-through of a change in a firm's productivity will be more than complete if and only if pass-through to the markup is positive (e.g. if an increase in productivity leads to an increase of the markup). This occurs when the elasticity of demand rises with relative cost. Clearly, positive pass-through to the markup of a firm-level increase in z occurs when the markup is increasing in relative productivity, a relationship we focused on in the previous section. Hence, price pass-through by the lowest productivity firm of an increase in their productivity will be more than complete if and only if the condition in Proposition 3 is satisfied.

Note also that (38) clarifies that whether high-productivity (large, low-price) firms exhibit greater or lesser pass-through of a firm-specific productivity shock than do lower productivity (small, high-price) firms depends on how the markup and the elasticity of demand vary with z. Specifically, it depends on the second derivative of the markup function, which, in turn, depends in a complicated way on the parameters of both the price quote and productivity distributions, as discussed in the previous section.

To gain further insight into this relationship, consider Figure 4, which depicts short-run price and markup pass-through rates of a change in a firm's productivity parameter for the clothing sub-sector.[21] First, consider the lowest-productivity (highest-cost) firm in equilibrium. This firm has $v = 1$ and a relative productivity of $1 - v = 0$. If this firm experiences a 1 percent decrease in its marginal cost due to an increase in its productivity, it reduces its price by roughly 8 percent.

[21] From (38), we note that markup pass-through equals price pass-through minus one.

That is, its pass-through is substantially more than complete. The firm's sizeable price and markup reductions are consistent with the shape of the price and markup functions depicted in Figure 2.

More generally, Figure 4 shows that small firms change their prices *more* in response to a change in their *own* cost, holding other retailers' costs constant, than do large firms. Small firms are relatively unproductive and as a result post high prices, selling primarily to buyers with no alternative. As described in Section 7, these firms face a relatively high elasticity of demand compared to much more productive firms and respond strongly to changes in their own cost. In this case, a reduction in their price made possible by an increase in productivity enables them to capture a large number of sales from competitors. Note again that this is reflected in the relatively steep decline in the price function in a neighborhood of $\tilde{z}$ (Figure 2.B.).

The responses of low-productivity firms contrast with those of their relatively high-productivity competitors. These firms sell successfully to a large number of consumers who have other alternatives, but at higher prices. They face low elasticities of demand and thus respond much less aggressively to a cost reduction, muting their price responses as they retain a large fraction of their sales in any case.

At high levels of relative productivity, there is a non-monotonic relationship between productivity and pass-through of a change in a firm's productivity. As suggested above, this non-monotonicity can be related to the shape of the markup function by noting that pass-through begins rising with relative productivity near the inflection point in the markup function around $1 - v = .4$. Figure 12 in Appendix E depicts analogous figures for the other sub-sectors, showing similar qualitative results but with notable quantitative differences due to variation in estimated parameters.

Our key finding that large retail firms exhibit relatively low levels of pass-through of a change in *own* cost while small firms fully pass-through those changes is consistent with the empirical findings from studies using manufacturing data. For example, Amiti et al. (2019), Amiti et al. (2014), and Berman et al. (2012) document similar behavior for manufacturing firms' responses to changes in their marginal costs, which are primarily driven by exchange rate shocks. More generally, our results, similar to theirs, contrast with the predictions of models that assume constant-elasticity-of-substitution demand and monopolistic competition, which predict identical pass-through rates (100 percent) for all firms.

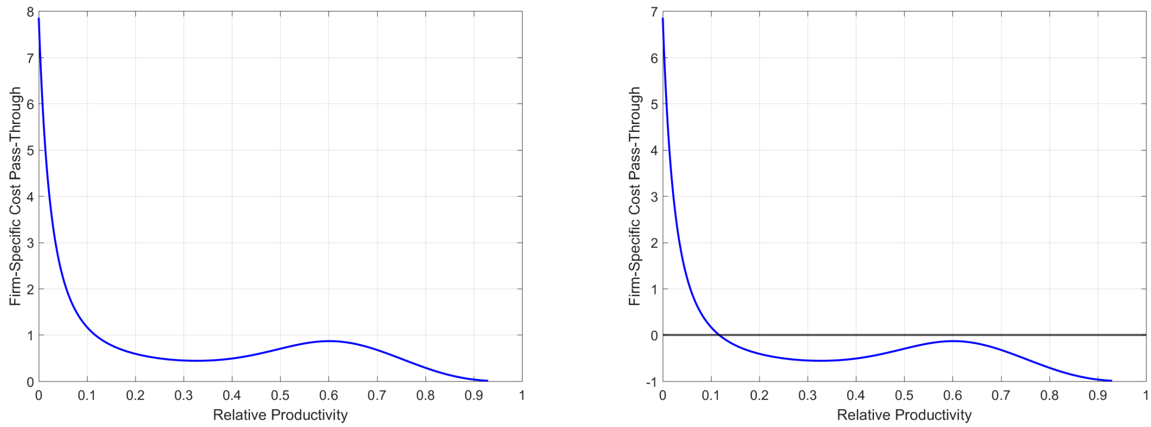
8.1.2 Common Productivity Changes

We now examine the effects of a change (again focusing on a reduction) in the common cost parameter, ϕ, starting with the following proposition regarding its effect on incumbent firms' prices:

Proposition 5. *In response to a change in ϕ, short-run pass-through to prices of incumbent retailers with $z > \tilde{z}$ is positive, incomplete, decreasing in q_1 when holding $q_1 + q_2$ constant, and increasing in a retailer's productivity parameter, z.*

Proof: See Appendix B.

Figure 4: Short-Run Pass-Through of a Firm-Specific Change in Productivity for Clothing Sub-Sector



A. Price Pass-Through B. Markup Pass-Through

Notes: Pass-through estimates are for the clothing and clothing accessory stores sub-sector (448). The calculations use the pass-through equations from Appendix D and the parameter estimates from Table 3. For estimates for the other sub-sectors, see Figure 12 in Appendix E.

Noting that the elasticity of the markup with respect to ϕ satisfies $\epsilon^{\phi}_{mkup(v)} = SRPT^{\phi}_{p(v)} - 1$, this proposition is also informative regarding the behavior of markup pass-through. In particular, pass-through of a common productivity change to markups is negative and decreasing in absolute value in a retailer's productivity parameter.

Proposition 5 establishes that all retailers, except those pricing at the maximum price, lower their prices in response to a reduction in ϕ. Each firm, however, passes through the common cost reduction only partially, resulting in an increase of their markup. To understand this, note first that a firm that was charging $\tilde{p}$ before the cost change has no incentive to lower its price. Because this firm sells exclusively to buyers who observe a single price, its profit-maximizing price cannot be below the reservation price, $\tilde{p}$. As described above, its markup is invariant to changes in ϕ because it is determined only by the fixed cost of operation and the reservation price, $\tilde{p}$.

Consider now firms initially pricing below $\tilde{p}$. All of these firms lower their price to increase the share of their sales to buyers that observe more than one price and to avoid losing sales to their less-productive competitors, which now lower their prices as well. As these firms sell mainly to buyers with few or no alternatives, there is relatively little gain to them from reducing prices in an attempt to increase sales when *all* firms' costs fall. Moreover, the lack of any price response from the highest-pricing firm mitigates their need to reduce their price to avoid *losing* sales to less-productive competitors. This continues throughout the distribution, with each more productive firm decreasing their price by less than the cost reduction but by more than those pricing above it. Incomplete pass-through here is due to the combined effect of households having incomplete information regarding prices *and* a reservation price that is unaffected by the cost change.

Proposition 5 also establishes that through this process, for those sectors characterized by more "captive" households (i.e. by a higher q_1/q_2), firms adjust their prices by relatively less in response to a common cost decrease in the short run. This aligns with the notion that more market power is typically associated with less cost pass-through. Moreover, an implication of productivity heterogeneity is that small price reductions by low-productivity firms effectively enable their larger, more productive competitors to mitigate their price reductions relative to those they would make if firms were equally productive.

Overall, higher-pricing *smaller* firms lower their prices by *less* in response to a *common* cost reduction than do larger firms charging lower prices. Turning to implications of this result for households, more informed households are more likely to purchase from lower-pricing firms. Hence, when there is a common cost change, those households experience larger price changes than do less-informed households, who are more likely to purchase from high-cost sellers that post relatively high prices and make only small price adjustments in response to the shock. If lower-income households are better informed as a result of devoting, for example, more effort to search, an increase in industry productivity redistributes real purchasing power from richer to poorer households, while a fall in productivity does the opposite.[22] This discussion underscores the importance of studying firm-level price responses in an environment in which firms optimally choose heterogeneous levels of pass-through to gain a clearer understanding of the distributional impacts of changes in firm-level productivity.

To more clearly illustrate variation in pass-through across firms, Figure 5 depicts estimated short-run price and markup pass-through in response to a change in ϕ for our focus sub-sector, clothing. As expected, the least productive firm exhibits zero price pass-through and, therefore, its markup rises in direct proportion to the fall in ϕ. All other firms exhibit incomplete price and markup pass-through, at rates increasing in their relative productivity. Figure 13 in Appendix E depicts estimated short-run price pass-through rates for the remaining sub-sectors. These exhibit similar qualitative patterns but quantitatively different pass-through rates, principally in the lower ranges of relative productivity.

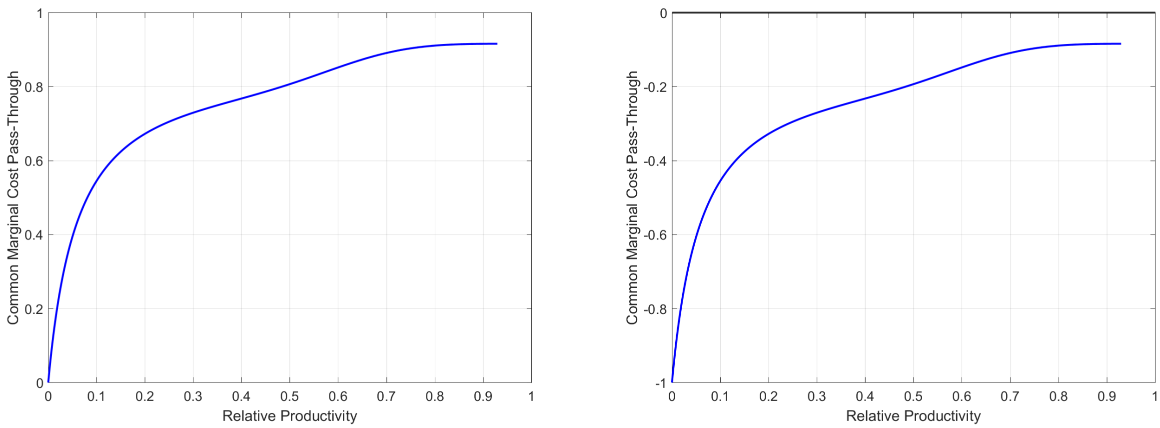
8.1.3 Reservation Price Changes

We now consider a change in households' reservation price, $\tilde{p}$. Because a change to the reservation price does not affect marginal cost, markup and price pass-through are equal for changes in $\tilde{p}$. We have the following proposition:

Proposition 6. *In response to a change in $\tilde{p}$, short-run pass-through to prices of incumbent retailers with $z > \tilde{z}$ equals $1 - SRPT^{\phi}_{p(v)}$, is positive, incomplete, increasing in q_1 when holding $q_1 + q_2$ constant, and decreasing in a retailer's productivity parameter, z.*

[22] This is broadly consistent with the results in Mnasri and Lapham (2023), who examine differences across households in the impact of exchange rate movements. In their environment, real exchange rate appreciations, which lower costs, redistribute real purchasing power from richer to poorer households while real depreciations do the opposite. Those authors also note that this result is consistent with the empirical findings of Cravino and Levchenko (2017) who document household-level impacts of the 1994 devaluation of the Mexican peso.

Figure 5: Short-Run Pass-Through of a Common Cost Change (ϕ) for the Clothing Sub-Sector



A. Price Pass-Through

B. Markup Pass-Through

Notes: Pass-through estimates are for the clothing and clothing accessory stores sub-sector (448). The calculations use the pass-through equations from Appendix D and the parameter estimates from Table 3. For estimates for the other sub-sectors, see Figure 13 in Appendix E.

Proof: See Appendix B.

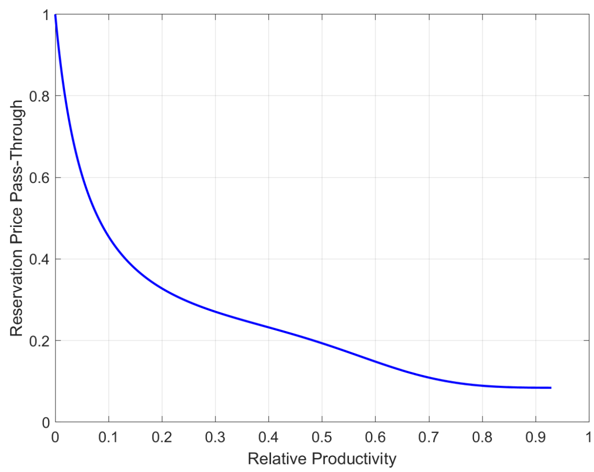
Focus now on an increase of $\tilde{p}$. We interpret this as a positive demand shock, even though the total units sold in equilibrium remains λ, the measure of households. In the absence of entry, a rise in the reservation price will cause the least productive firm to raise its price to the higher reservation price, thereby exhibiting complete pass-through. In response, all other firms also raise their price and, since their costs have not changed, their markups increase as well.

Similarly to the effect of a firm-specific productivity change, higher-pricing smaller firms adjust prices by *more* than larger firms in response to a change in $\tilde{p}$. When the least productive firm raises its price to the new reservation price, slightly more productive firms increase their prices at a relatively high rate as well, as they face little risk of being undercut in the neighborhood of the least productive firm. They mitigate their price increases somewhat, however, in an attempt to capture some sales from competitors. At higher levels of productivity, firms increasingly sell to buyers who observe multiple quotes and for whom they must compete, which serves to mitigate their price increases further. Thus, pass-through of the reservation price diminishes with relative productivity at a rate depending on both the search process and the shape of the productivity distribution.

In this case, due to the negative relationship between pass-through and relative productivity, the distributional implications for households are the opposite of those in the case of a common cost shock. More-informed households experience relatively less harm than less-informed households as prices rise. However, they benefit less when a reduction in $\tilde{p}$ puts downward pressure on prices.

Proposition 6 also establishes that prices and markups increase by more in response to an increase of $\tilde{p}$ in industries where more households observe only one price while fewer observe two

Figure 6: Short-Run Pass-Through of a Reservation Price Change ($\tilde{p}$) for the Clothing Sub-Sector



Notes: Pass-through estimates are for the clothing and clothing accessory stores sub-sector (448). The calculations use the pass-through equations from Appendix D and the parameter estimates from Table 3. For estimates for the other sub-sectors, see Figure 14 in Appendix E.

prices. This is intuitive as we expect that in markets where firms have more market power, they enjoy a larger increase in markups in response to a positive demand shock.

Figure 6 plots the short-run pass-through of an increase in the reservation price as a function of relative productivity for the clothing sub-sector. Consistent with Proposition 6 and the discussion above, the figure depicts incomplete pass-through for all firms at a rate that declines with relative productivity. In Appendix E, we present similar graphs for the other sub-sectors in Figure 14, where we observe some quantitative differences which result from differences in our estimates of the search and productivity parameters across sub-sectors.

8.1.4 Changes in the Ratio of Fixed Cost to Market Tightness

Finally, we consider a change in the parameter $\delta = \Psi/\mu$. This can reflect either a change in the fixed cost of production, Ψ, a shift in market tightness, μ, or a change in both. In all cases, however, there is no short-run impact on prices or markups of such a change (i.e. without movement in $\tilde{z}$) as neither Ψ nor μ enter the firm's pricing decision.[23]

8.2 Long-Run Pass-Through

We now consider pass-through at the firm level in the *long run*, i.e. accounting for changes in $\tilde{z}$.[24] As above, we consider changes associated with *reductions* in $\tilde{z}$ that result in entry and an increase

[23] A decrease in Ψ increases a firm's profits while an increase in μ raises both profits and revenue.

[24] Similar to short-run pass-through, long-run *markup* pass-through equals long-run *price* pass-through minus one when ϕ changes, but they are equal when $\tilde{p}$ or δ changes.

in firm heterogeneity. The following describes the long-run responses of prices to changes in the economy parameters, ϕ, $\tilde{p}$, and δ, focusing on the behavior of *incumbent* firms.[25]

Proposition 7. *(i.) In response to a change in common cost, ϕ, long-run pass-through to continuing retailers' prices is positive. (ii.) In response to a change in the reservation price, $\tilde{p}$, long-run pass-through to continuing retailers' prices can be positive, negative, or zero. (iii.) In response to a change in the ratio of fixed cost to market tightness, δ, long-run pass-through to continuing retailers' prices can be positive, negative, or zero.*

Proof: See Appendix B.

Beginning with the long-run effects of a reduction of common cost, ϕ, inspection of the pricing function in (24) indicates that the direct effect of a change in ϕ works via the term $\phi/\tilde{z}$. Since fixed operating costs and market tightness are in constant proportion, δ, however, the zero-profit condition in (15) indicates that $\tilde{z}$ falls in direct proportion to ϕ. Consequently, the ratio $\phi/\tilde{z}$ is invariant to changes in ϕ so there is no effect on incumbent firms' prices through this channel resulting from the reduction in $\tilde{z}$.

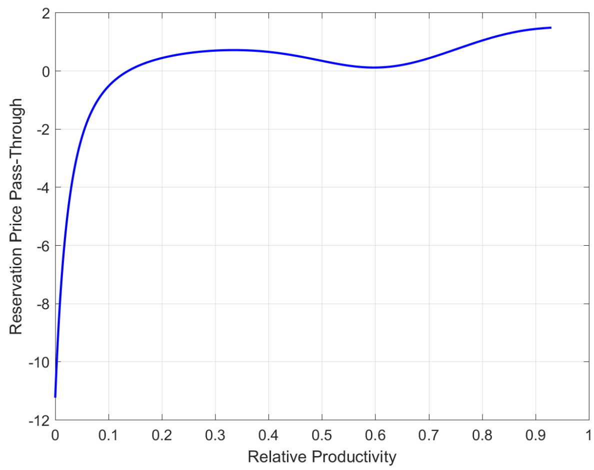
As such, in the long run an incumbent firm's price changes in response to a shift in $\tilde{z}$ only because this firm now has a different *relative* cost, v. Thus, long-run price and markup pass-through of a change in ϕ are equal to the price and markup pass-through for a firm experiencing the same change in its own relative productivity level v in the short run. Figure 4 above and Figure 12 in Appendix E depict estimated *long-run* pass-through for firms experiencing changes in ϕ as well as short-run pass-through for the equivalent firm-specific changes in v.

Turning to a change in the reservation price, we note that a rise in $\tilde{p}$ allows less productive firms that previously could not operate profitably to enter the market, thereby lowering $\tilde{z}$. This reduction in $\tilde{z}$ implies that the relative marginal cost, v, of each continuing retailer decreases, prompting them to lower their price ($\epsilon^{v}_{p(v)} > 0$ as established above). The increase in the measure of firms pricing above them, however, induces them to raise their prices. The combination of these opposing effects leads to an overall ambiguous qualitative impact on a retailer's price in response to a change in the reservation price in the long run.

To further demonstrate this ambiguity, Figure 7 depicts long-run pass-through for a change in $\tilde{p}$ for incumbent firms in the clothing sub-sector. Here both negative and positive pass-through rates are apparent. After entry, the firm that previously posted the reservation price is no longer the least productive firm and drops its price aggressively in order to take sales from its newly entered less-productive competitors. More-productive incumbents respond by cutting prices so as to avoid being undercut themselves. As productivity increases, the risk of being undercut diminishes and the positive effect of an increase in households' reservation price (as described above for the short run) strengthens. Low-productivity firms reduce prices at a decreasing rate as productivity increases and intermediate- to high-productivity firms *increase* their prices, thus exhibiting positive pass-through.

[25]By their very nature, new entrants do not change their prices, so it makes little sense to discuss their rates of pass-through. Their behavior does, however affect *average* prices (see Section 8.3). Similarly, it is not sensible to consider the long-run effect of a firm-specific productivity change, as a change in one firm's cost does not affect $\tilde{z}$ and

Figure 7: Long-Run Pass-Through of a Reservation Price Change ($\tilde{p}$) for the Clothing Sub-Sector



Notes: Pass-through estimates are for the clothing and clothing accessory stores sub-sector (448). The calculations use the pass-through equations from Appendix D and the parameter estimates from Table 3. For estimates for the other sub-sectors, see Figure 15 in Appendix E.

Lastly, consider a fall in $\delta = \frac{\Psi}{\mu} = \frac{\Psi N}{\lambda}$. For ease of exposition, we interpret this as a cost change associated with a reduction in fixed costs, Ψ, although it could also be associated with an increase in market tightness which here we associate with the exogenous measure λ as N is determined in equilibrium in the long run. While changes in δ have no effect on pricing in the short run, in the long run a reduction of fixed costs (or the increase in market tightness) increases profits conditional on z, leading to entry as reflected in a higher N and lower $\tilde{z}$.[26] The reduction of $\tilde{z}$ has similar opposing effects as in the scenario above when the reservation price changes, leading to ambiguity in the direction of long-run pass-through in response to a change in δ. These general properties are illustrated in Figure 8 for the clothing sub-sector, where we observe both positive and negative long-run pass-through of a change in δ.

8.3 Pass-Through to Average Prices

We now consider briefly the effects of these changes in economy parameters on two measures of average prices. The average *posted* price is given by

$$\bar{p}_{post} = -\int_0^1 p(v)G'(v)dv, \tag{39}$$

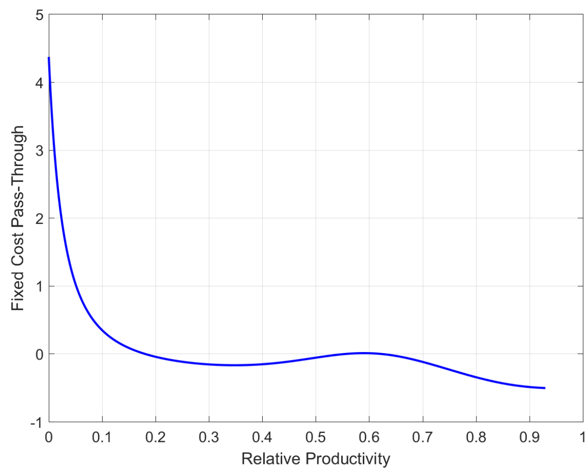
while the average *transaction* price is the average of the lowest price observed and is given by

$$\bar{p}_{tran} = -\int_0^1 p(v)A(G(v))G'(v)dv. \tag{40}$$

thus causes neither entry nor exit.

[26] From the definition of δ, we see that a fall in N must be associated with a further increase in fixed costs.

Figure 8: Long-Run Pass-Through of a Change in δ for the Clothing Sub-Sector



Notes: Pass-through estimates are for the clothing and clothing accessory stores sub-sector (448). The calculations use the pass-through equations from Appendix D and the parameter estimates from Table 3. For estimates for the other sub-sectors, see Figure 16 in Appendix E.

The following proposition summarizes the responses of average posted and transaction prices to changes in the parameters we considered above[27]:

Proposition 8. *(i.) In response to a change in ϕ, short-run pass-through to both average posted and transaction prices is positive while long-run pass-through to average prices equals zero. (ii.) In response to a change in $\tilde{p}$, short-run and long-run pass-through to both average posted and transactions prices are positive and more than complete. (iii.) In response to a change in δ, short-run pass-through to both average posted and transactions prices equals zero while long-run pass-through to average prices is negative.*

Proof: See Appendix B.

Positive short-run pass-through to both average posted and transactions prices in response to a change in ϕ follows immediately from Proposition 5 since each individual retailer lowers their price when ϕ falls. Also, since that proposition establishes that firms charging lower prices exhibit higher rates of short-run pass-through than do firms charging higher prices, it follows that short-run pass-through to the average transaction price will be higher than that to the average posted price. This occurs because more transactions occur for lower-pricing firms and so more transactions occur for firms with higher pass-through, leading to higher pass-through to the average transaction price.

In the long run, when ϕ decreases, lower productivity firms enter, causing $\tilde{z}$ to fall in direct proportion to ϕ as described above. In this case, neither average variable cost nor the average price changes.[28] The intuition here is that the upward pressure on average prices due to entry of

[27]In all cases we consider, average posted and transactions prices move (if at all) in the same direction. There are only *quantitative* differences between pass-through to average posted and transaction prices.

[28]Average variable cost as a function of v is given by $AVC(v) = -\left(\frac{\phi}{\tilde{z}}\right)\int_0^1 vG'(v)dv$. With no change in average

less-productive firms directly offsets the fall in incumbent firms' prices, and neither the distribution of prices nor the average price changes. The directly offsetting effect of intensive margin responses (surviving retailers lower their price) with extensive margin responses (relatively lower-productivity retailers enter) leading to no effect on averages is a common feature of models that assume that productivity follows a Pareto distribution and has been recognized in other contexts (see Melitz and Redding (2015), for example).[29]

Considering a change in $\tilde{p}$, as intuition and Proposition 6 suggest, an increase in the maximum price that households are willing to pay results in higher average prices in both the short and long run. Furthermore, short-run pass-through to average transaction prices is below pass-through to average posted prices in this case. This follows from Proposition 6 that the short-run pass-through of a reservation price shock is decreasing in productivity. Hence, in contrast to the scenario in which ϕ changes, more transactions occur here for firms with *lower* pass-through, leading to less pass-through to the average transaction price than to the average posted price. Finally, in the long run, for a change in $\tilde{p}$, the intensive and extensive margin responses reinforce one another, leading to positive and more than complete pass-through to both measures of average prices.

As noted above, a change in δ has no effect on any firm's price in the short run and as such no effect on either the average posted or transaction price. In the long run, however, when δ decreases, the extensive margin effect involving the entry of relatively lower-productivity firms tends to raise average prices. If the long-run pass-through of a change in δ to firm-level prices (the intensive margin effect) is negative, this reinforces the extensive margin effect, leading to negative long-run pass-through to average prices. Proposition 8 establishes that even when this is not the case and intensive margin effect is positive, the extensive margin response dominates, resulting in negative long-run pass-through to average prices following any change in δ.[30]

Tables 6 and 7 report short- and long-run quantitative pass-through estimates for both average posted and transaction prices for the sub-sectors we study.[31] In both tables, we observe variation across retail sectors in the responses of average prices to changes in economy parameters. Based on our estimates, the clothing, furniture, and miscellaneous sub-sectors show some of the lowest pass-through rates to average prices of a change in ϕ but are among the highest for a change in $\tilde{p}$ or a change in δ. In contrast, electronics, gasoline stations, and vehicle parts exhibit relatively high levels of average price pass-through in response to a change in ϕ while responding with relatively

variable cost in the long run, it follows that average prices do not change in the long run.

[29] We have also analyzed a version of the model in which fixed costs, Ψ, are exogenous and independent of market tightness, μ. In that environment, there is an additional effect on $\tilde{z}$ as the fall in ϕ raises the measure of active firms, N, and so lowers market tightness μ. Because there is no change in fixed costs, in contrast to the baseline model, the decrease in μ lowers firm-level profits, leading to a rise in the cutoff productivity for operation, $\tilde{z}$. If this is the dominant effect, $\phi/\tilde{z}$ falls, thereby decreasing average variable cost due to the exit of less-productive firms. This decrease in average variable cost is reflected in lower average prices, resulting in positive pass-through of a change in the common cost parameter to average prices in the long run. This alternate environment is briefly analyzed in Appendix C.

[30] We also note that long-run pass-through of a change in δ equals one minus long-run pass-through of a change in $\tilde{p}$ (see Appendix D).

[31] Proposition 8 indicates that short-run pass-through to average prices of a change in δ and long-run pass-through to average prices of a change in ϕ equal zero, and so we exclude these cases from the tables.

Table 6: Short-Run Pass-Through to Average Prices

	441 MOTR	442 FURN	443 ELEC	444 BLDG	445 FOOD	446 HLTH	447 GASS	448 CLTH	451 SPRT	452 GENL	453 MISC
Change in Common Cost (ϕ)											
Posted prices	0.8547	0.6376	0.7698	0.6900	0.7688	0.7307	0.8604	0.5633	0.6492	0.7342	0.5797
Transaction prices	0.9227	0.7571	0.8845	0.7972	0.8522	0.8315	0.9420	0.6854	0.7649	0.8425	0.7234
Change in Reservation Price ($\tilde{p}$)											
Posted prices	0.1453	0.3624	0.2302	0.3100	0.2312	0.2693	0.1396	0.4367	0.3508	0.2658	0.4203
Transaction prices	0.0773	0.2429	0.1155	0.2028	0.1478	0.1685	0.0580	0.3146	0.2351	0.1575	0.2766

Table 7: Long-Run Pass-Through to Average Prices

	441 MOTR	442 FURN	443 ELEC	444 BLDG	445 FOOD	446 HLTH	447 GASS	448 CLTH	451 SPRT	452 GENL	453 MISC
Change in Reservation Price ($\tilde{p}$)											
Posted prices	1.0031	1.2451	1.1710	1.1013	1.0241	1.2649	1.0040	1.3133	1.2221	1.1316	1.2733
Transaction prices	1.0033	1.2909	1.1965	1.1170	1.0267	1.3015	1.0043	1.3812	1.2617	1.1510	1.3410
Change in Relative Fixed Cost (δ)											
Posted prices	-0.0031	-0.2451	-0.1710	-0.1013	-0.0241	-0.2649	-0.0040	-0.3133	-0.2221	-0.1316	-0.2733
Transaction prices	-0.0033	-0.2909	-0.1965	-0.1170	-0.0267	-0.3015	-0.0043	-0.3812	-0.2617	-0.1510	-0.3410

lower levels of pass-through following a change in both $\tilde{p}$ and δ.

Overall, the pass-through of cost and demand movements to average prices differs with the type of shock and between the short and long run. In the short run, common cost movements are passed through more strongly than reservation price changes, but this pattern is reversed in the long run.[32] Also, whether transactions prices or posted prices respond more strongly depends on variation in pass-through across firms. As such, our analysis indicates that modeling the responses of individual firms' prices and markups to shocks is crucial for understanding pass-through to average prices.

9 Conclusions

In this paper we develop, estimate, and study a model of consumer search in an industry comprised of heterogeneous firms competing by price-setting. Our theory is important for studying the interaction between firm heterogeneity and search frictions associated with incomplete information and their implications for market power, as reflected in firm markups. Using detailed firm-level data on Canadian retail sub-sectors, we estimate parameters of both the search process and distribution of firm-level productivity. We characterize equilibrium distributions of prices and markups and find that both search and firm heterogeneity have significant implications for market power and price dispersion. We then study the responses of firm-level and average prices to various changes in cost and demand conditions, both with and without entry and exit of firms in equilibrium.

In our theory, market power emanates from consumers' incomplete information regarding trading opportunities (i.e. prices). The estimates of our structural model suggest that in all sub-sectors, the modal number of price quotes observed equals two and in all cases more than 59% of consumers

[32] The result of zero long-run pass-through of a common cost shock depends on the class of productivity distributions we consider and is a topic for future research.

observe three or fewer prices. Only a small fraction of consumers observe only a single price, and a small but significant number are very well informed. With regard to firm productivity heterogeneity, we restrict attention to Pareto distributions, and our estimates of shape parameters are in line with others' estimates based on manufacturing data.

We characterize variation in prices and markups across firms within an industry and relate them to variation in the elasticity of demand facing different firms. This variation depends both qualitatively and quantitatively on characteristics of the search process and the extent of firm heterogeneity. Given the former, the latter plays a key role in determining both the distributions of prices and markups and their responses to changes in economy parameters. Due to the search friction, relatively unproductive firms can survive in the market despite their cost disadvantages. Importantly, their presence enables more productive firms to charge higher markups than they would in a market with homogeneous firms.

Using the estimated model, we quantify these effects and characterize heterogeneity in the responses, or pass-through, to firm-level prices and markups of firm-specific and common cost and demand movements. The extent and, in some cases, the direction of pass-through varies significantly across firms. Furthermore, the qualitative nature of the relationship between firm size and the magnitude of pass-through differs depending on the source of the shock.

Differential market power and pass-through between high- and low-productivity firms has implications not only for the efficiency of the market but also for distributional consequences of shocks across consumers. For example, when all firms face a fall in costs, low-price firms cut prices significantly more than high-price firms in the short run. Hence, consumers who engage in more search and, therefore, tend to purchase from lower-pricing firms (perhaps lower-income consumers) benefit more from the fall in costs than their less-informed counterparts. In contrast, when all firms face a fall in demand, low-price firms cut prices significantly less than high-price firms in the short run. Thus, more informed consumers benefit less from the fall in demand than their less-informed counterparts. Although we do not incorporate *ex-ante* heterogeneity across consumers in our current framework, this is a promising avenue for future research.

References

Alessandria, G., "Consumer Search, Price Dispersion and International Relative Price Fluctuations," *International Economic Review* 50 (2009), 803–29.

Amiti, M., O. Itskhoki and J. Konings, "Importers, Exporters, and Exchange Rate Disconnect," *American Economic Review* 7 (2014), 1942–78.

———, "International Shocks, Variable Markups, and Domestic Prices," *Review of Economic Studies* 86 (2019), 2356–402.

Arkolakis, C., "Market Penetration Costs and the New Consumer Margin in International Trade," *Journal of Political Economy* 118 (2010), 1151–99.

Atkeson, A. and A. Burstein, "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review* 98 (2008), 1998–2031.

Autor, D., D. Dorn, L. Katz, C. Patterson and J. Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics* 135 (2020), 645–709.

Baggs, J., L. Fung and B. Lapham, "Exchange Rates, Cross-Border Travel, and Retailers: Theory and Empirics," *Journal of International Economics* 115 (November 2018), 59–79.

Berman, N., P. Martin and T. Mayer, "How Do Different Exporters React to Exchange Rate Changes?," *Quarterly Journal of Economics* 127 (2012), 437–92.

Burdett, K. and K. Judd, "Equilibrium Price Dispersion," *Econometrica* 51 (1983), 955–69.

Burdett, K. and D. Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review* (1998), 257–73.

Cravino, J. and A. Levchenko, "The Distributional Consequences of Large Devaluations," *American Economic Review* 107 (2017), 3477–509.

Das, S., M. Roberts and J. Tybout, "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica* 75 (2007), 937–73.

De Loecker, J., J. Eeckhout and G. Unger, "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics* 135 (2020), 561–644.

De Loecker, J. and F. Warzynski, "Markups and Firm-Level Export Status," *American Economic Review* 102 (2012), 2437–71.

Head, A., B. Lapham and A. Kumar, "Market Power, Price Adjustment and Inflation," *International Economic Review* 51 (February 2010), 73–98.

Head, A., Q. Liu, G. Menzio and R. Wright, “Sticky Prices: A New Monetarist Approach,” *Journal of the European Economic Association* 10 (October 2012), 939–73.

Head, K., T. Mayer and M. Thoenig, “Welfare and Trade without Pareto,” *American Economic Review: Papers & Proceedings* 104 (2014), 310–16.

Herrenbrueck, L., “An Open-Economy Model with Money, Endogenous Search, and Heterogeneous Firms,” *Economic Inquiry* 55 (2017), 1648–70.

Hopenhayn, H., “Entry, Exit, and Firm Dynamics in Long Run Equilibrium,” *Econometrica* 60 (September 1992), 1127–50.

Kaplan, G., G. Menzio, L. Rudanko and N. Trachter, “Relative Price Dispersion, Evidence and Theory,” *American Economic Journal: Microeconomics* 11 (2019), 68–124.

Melitz, M., “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica* 71 (2003), 1695–725.

Melitz, M. and G. Ottaviano, “Market Size, Trade, and Productivity,” *The Review of Economic Studies* 75 (January 2008), 295–316.

Melitz, M. J. and S. J. Redding, “New Trade Models, New Welfare Implications,” *American Economic Review* 105 (2015), 1105–46.

Menzio, G., “Search Theory of Imperfect Competition with Decreasing Returns to Scale,” Forthcoming, *Journal of Economic Theory*, 2023.

———, “Markups: A Search-Theoretic Perspective,” Manuscript, New York University, July 2024.

Menzio, G. and N. Trachter, “Equilibrium Price Dispersion Across and Within Stores,” *Review of Economic Dynamics* 28 (2018), 205–20.

Mnasri, A. and B. Lapham, “A Competitive Search Approach to Exchange Rate Pass-Through,” *Economic Theory* 76 (2023), 153–201.

Mortensen, D., “A Comment on Price Dispersion, Inflation and Welfare,” *International Economic Review* 46 (May 2005), 573–78.

Ottaviano, G., T. Tabuchi and J.-F. Thisse, “Agglomeration and Trade Revisited,” *International Economic Review* 43 (2002), 409–36.

Pearson, J. W., S. Olver and M. A. Porter, “Numerical Methods for the Computation of the Confluent and Gauss Hypergeometric Functions,” *Numerical Algorithms* 74 (August 2017), 821–66.

Wang, L., R. Wright and L. Liu, “Sticky Prices and Costly Credit,” *International Economic Review* 61 (2020), 37–70.

Appendix A: Derivations

A.1 Derivation of the Pricing Function

The first-order condition for profit maximization given by equation (7) can be rewritten as

$$p'(z) + V(z)p(z) = \left(\frac{\phi}{z}\right) V(z), \tag{41}$$

where $V(z) \equiv \frac{A'(F(z))F'(z)}{A(F(z))}$. This is in a standard form of a linear first-order differential equation whose solution is given by

$$p(z) = \left(\frac{1}{\chi(z)}\right) \left(\int \chi(z) \left(\frac{\phi}{z}\right) V(z)dz + C\right), \tag{42}$$

where $\chi(z) = exp\left(\int V(z)dz\right)$. Given the definition of $V(z)$, we have $\chi(z) = exp\left(\ln(A(F(z)))\right) = A(F(z))$. Substituting this and the expression for $V(z)$ into equation (42) gives

$$p(z) = \left(\frac{1}{A(F(z))}\right) \left(\int \left(\frac{\phi}{z}\right) A'(F(z))F'(z)dz\right). \tag{43}$$

Recalling that an indefinite integral can be written as a definite integral plus an arbitrary constant, we can write the above as follows:

$$p(z) = \left(\frac{1}{A(F(z))}\right) \left(C + \int_{\tilde{z}}^{z} \left(\frac{\phi}{x}\right) A'(F(x))F'(x)dx\right), \tag{44}$$

where $C = \int_a^{\tilde{z}} \left(\frac{\phi}{x}\right) A'(F(x))F'(x)dx$ with a as an arbitrary constant. We impose the condition that the maximum price, $\tilde{p}$, is associated with $\tilde{z}$ to solve for C:

$$p(z) = \frac{C}{A(J(\tilde{z})} = \frac{C}{q_1} \longrightarrow C = \tilde{p}q_1. \tag{45}$$

Substituting this into equation (44) gives the pricing function:

$$p(z) = \left(\frac{1}{A(F(z))}\right) \left(\tilde{p}q_1 + \phi \int_{\tilde{z}}^{z} \frac{A'(F(x))F'(x)}{x} dx\right). \tag{46}$$

A.2 Comparison of Heterogeneous Firm Model with Homogeneous Firm Model

Prices in the model with noisy search and *ex-ante* homogeneous firms developed by Burdett and Judd (1983) are implicitly determined from the equal profit condition given by $(\tilde{p} - \phi/\tilde{z})\,\mu q_1 = (\hat{p} - \phi/\tilde{z})\,\mu A\left(1 - \hat{L}(\hat{p})\right)$, where $\hat{L}(\cdot)$ is the equilibrium price distribution in that environment. Evaluating this expression at a point in the price distribution where $1 - \hat{L}(\hat{p}) = F(z)$ and rearranging gives

$$\hat{p} = \left(\frac{q_1}{A(F(z))}\right) \left(\tilde{p} - \left(\frac{\phi}{\tilde{z}}\right)\right) + \left(\frac{\phi}{\tilde{z}}\right). \tag{47}$$

Combining this with the pricing function in our model given by equation (8), we are able to derive

$$p(v) = \hat{p} + \left(\frac{(\phi/\tilde{z})}{A(G(v))}\right) \int_{\tilde{z}}^{z} A'(F(x)F'(x) \left(\frac{\tilde{z}}{x} - 1\right) dx, \tag{48}$$

where we have used $\int_{\tilde{z}}^{z} A'(F(x))F'(x)dx = A(F(z)) - q_1$.

From (47), we have the following markup associated with $\hat{p}$:

$$\widehat{mkup} \equiv \frac{\hat{p}}{\left(\frac{\phi}{\tilde{z}}\right)} = \left(\frac{q_1}{A(F(z))}\right)(mkup(\tilde{z}) - 1) + 1. \tag{49}$$

Combining this with the markup function in our model given by equation (25) gives

$$mkup(z) = \widehat{mkup} + \left(\frac{1}{A(F(z))}\right)\left(mkup(\tilde{z})q_1\left(\frac{z}{\tilde{z}} - 1\right) + \int_{\tilde{z}}^{z} A'(F(x))F'(x)\left(\frac{z}{x} - 1\right) dx\right), \tag{50}$$

where we have again used the expression for the integral given above.

A.3. Probability Density Functions

Because price is monotonically decreasing in z in our model, we have $1 - CDF_p(\ln(p(z)/\delta)) = F(z)$. Differentiating both sides of this equation with respect to z and rearranging gives

$$PDF_p(\ln(p(z)/\delta)) = \left(\frac{-p(z)/\delta}{p'(z)/\delta}\right) F'(z). \tag{51}$$

Differentiating (8) with respect to z, we have

$$\frac{p'(z)}{\delta} = \left(\frac{A'(F(z))}{A(F(z))}\right)\left(\alpha\left(\frac{\tilde{z}}{z}\right) - \frac{p(z)}{\delta}\right) F'(z). \tag{52}$$

Substituting this into (51) and writing functions in terms of v gives

$$PDF_p\left(\ln(p(v)/\delta)\right) = \left(\frac{A(G(v))}{A'(G(v))}\right)\left(\frac{p(v)/\delta}{p(v)/\delta - \alpha v}\right). \tag{53}$$

Because revenue is monotonically increasing in z in our model, we have $CDF_R(\ln(R(z)/\Psi)) = F(z)$. Differentiating both sides of this equation with respect to z and rearranging gives

$$PDF_R(\ln(R(z)/\Psi)) = \left(\frac{R(z)/\Psi}{R'(z)/\Psi}\right) F'(z). \tag{54}$$

Differentiating (11) with respect to z, we have

$$\frac{R'(z)}{\Psi} = \alpha\left(\frac{\tilde{z}}{z}\right) A'(F(z))F'(z). \tag{55}$$

Substituting this into (54) and writing functions in terms of v gives

$$PDF_R\left(\ln(R(v)/\Psi)\right) = \left(\frac{R(v)/\Psi}{\alpha v A'(G(v))}\right). \tag{56}$$

In the homogeneous productivity economy, we can differentiate the pricing equation given by the second expression in (34) and rearrange to derive the following expression for the PDF of the log of scaled prices in that model:

$$PDF_{\hat{p}}\left(\ln(\hat{p}/\delta)\right) = \left(1 + \alpha A(1 - \hat{L}(\hat{p}))\right)\left(\frac{A(1 - \hat{L}(\hat{p}))}{A'(1 - \hat{L}(\hat{p}))}\right). \tag{57}$$

The lower support of this distribution equals $\alpha + \frac{1}{A(1)}$, and the upper support equals $\alpha + \frac{1}{q_1}$. Choosing values for $\hat{L}(\hat{p}) \in [0, 1]$ and calculating the associated $\hat{p}$ from (34) and the associated PDF values from the above equation allows us to plot the PDF for scaled prices in this model in Figure 1.

Because markups are monotonically increasing in price in the homogeneous productivity economy, we have $CDF_{\widehat{mkup}}(\widehat{mkup}(\hat{p})) = CDF_{\hat{p}}(\hat{p})$. From this and equation (35), we can derive the PDF of the log of markups in that model:

$$PDF_{\widehat{mkup}}\left(\ln(\widehat{mkup})\right) = PDF_{\hat{p}}\left(\ln(\hat{p}/\delta)\right). \tag{58}$$

The lower support of this distribution equals $1 + \frac{1}{\alpha A(1)}$, and the upper support equals $1 + \frac{1}{\alpha q_1}$.

Appendix B: Proofs of Propositions

Proof of Proposition 1

The approach in this proof follows the approach of the proof of Lemma 1 in Herrenbrueck (2017), which follows the proof in Burdett and Mortensen (1998). Let $z_1 > z_2$ and let $p_j \equiv p(z_j)$ for $\jmath \in \{1, 2\}$. To establish the proposition, it is sufficient to show that $p_1 < p_2$. Let $\tilde{\pi}(p_j, z_i) \equiv \left(p_j - \frac{\phi}{z_i}\right)\mu A\left(1 - L(p_j)\right) - \Psi$. We have the following inequalities:

$$\tilde{\pi}(p_1, z_1) \geq \tilde{\pi}(p_2, z_1) > \tilde{\pi}(p_2, z_2) \geq \tilde{\pi}(p_1, z_2). \tag{59}$$

The first and third inequalities follow from profit maximization, and the second inequality follows from $z_1 > z_2$. Subtracting the fourth term from the first term and the third term from the second term, we have

$$\tilde{\pi}(p_1, z_1) - \tilde{\pi}(p_1, z_2) \geq \tilde{\pi}(p_2, z_1) - \tilde{\pi}(p_2, z_2), \tag{60}$$

or

$$\left(p_1 - \frac{\phi}{z_1}\right)\mu A\left(1 - L(p_1)\right) - \left(p_1 - \frac{\phi}{z_2}\right)\mu A\left(1 - L(p_1)\right) \geq \left(p_2 - \frac{\phi}{z_1}\right)\mu A\left(1 - L(p_2)\right) - \left(p_2 - \frac{\phi}{z_2}\right)\mu A\left(1 - L(p_2)\right). \tag{61}$$

We can rearrange this to derive

$$\left(\frac{\phi}{z_2} - \frac{\phi}{z_1}\right) A\left(1 - L(p_1)\right) \geq \left(\frac{\phi}{z_2} - \frac{\phi}{z_1}\right) A\left(1 - L(p_2)\right), \tag{62}$$

or, since $z_1 > z_2$, we have

$$A\left(1 - L(p_1)\right) \geq A\left(1 - L(p_2)\right). \tag{63}$$

Now, because $L(p)$ is a CDF, $L(p)$ is strictly increasing in p. From the definition of $A((1 - L(p))$, we see that this function is strictly decreasing in $L(p)$. It follows, then, that $A((1 - L(p))$ is strictly decreasing in p. Hence, from equation (63), we can conclude that $p_1 \leq p_2$.

Suppose now that $p_1 = p_2$. Given that the distribution of productivity, $F(z)$, is continuous with connected support, $z_1 > z_2$ implies that there exists a positive measure of firms maximizing profits by posting the same price. This contradicts the result that $L(p)$ is a continuous distribution with connected support, established by Burdett and Judd (1983), Lemma 1, p. 959.

Proof of Proposition 2

Differentiating (27) with respect to z gives

$$y'(z) = \frac{\sum_{k=2}^{K} q_k k(k-1)\left(\frac{\tilde{z}}{z}\right) F(z)^{k-2} F'(z))}{q_1 + \sum_{k=2}^{K} q_k k(k-1) \int_{\tilde{z}}^{z} \left(\frac{\tilde{z}}{x}\right) F(x)^{k-2} F'(x) dx}. \tag{64}$$

Recalling that $F(z)$ is the CDF of z, for active firms we have that $y'(z)$ is positive.

Proof of Proposition 3

Rewrite the markup as a function of v in equation (25) as follows:

$$mkup(v) = \frac{B(v)}{vA(G(v))}, \tag{65}$$

where $B(v) \equiv mkup(1)q_1 - \int_v^1 h(u)du$ and $B'(v) = h(v)$. We can use this equation to derive the elasticity of the markup with respect to v:

$$\epsilon^v_{mkup(v)} = \left(\frac{h(v)}{A(G(v))}\right)\left(v\left(\frac{A(G(v))}{B(v)}\right) - 1\right) - 1. \tag{66}$$

Now at $v = 1$, we have $h(1) = A'(1)G'(1) = 2q_2G'(1)$, $A(G(1)) = q_1$, and $B(1) = mkup(1)q_1$, so the above elasticity for the firm with $v = 1$ equals

$$\epsilon^v_{mkup(1)} = 2\left(\frac{q_2}{q_1}\right) G'(1) \left(\frac{-1}{elas(1)}\right) - 1. \tag{67}$$

This elasticity is positive if and only if $elas(1) < -2\left(\frac{q_2}{q_1}\right) G'(1)$.

Proof of Proposition 5

Using equation (24), we can derive the elasticity of $p(v)$ with respect to ϕ, holding $\tilde{z}$ fixed:

$$SRPT^{\phi}_{p(v)} = \frac{-\left(\frac{\phi}{\tilde{z}}\right)\int_v^1 h(u)du}{\tilde{p}q_1 - \left(\frac{\phi}{\tilde{z}}\right)\int_v^1 h(u)du} = 1 - \frac{\tilde{p}q_1}{p(v)A(G(v))}. \tag{68}$$

Recalling that $h(u) < 0\ \forall u$, we see that retailer-level price pass-through to a common cost shock is positive and incomplete. Differentiating equation (68) with respect to v, we have

$$\frac{\partial SRPT^{\phi}_{p(v)}}{\partial v} = \frac{\left(\frac{\phi}{\tilde{z}}\right)\tilde{p}q_1 h(v)}{\left(p(v)A(G(v))\right)^2}. \tag{69}$$

Recalling that $h(u) < 0\ \forall\ u$, we see that this derivative is negative, implying that retailers with higher z (lower v) will exhibit higher short-run pass-through. Differentiating $ln\left(SRPT^{\phi}_{p(v)}\right)$ with respect to q_1 holding $q_1 + q_2$ fixed so that $\frac{\partial q_2}{\partial q_1} = -1$, we have

$$\frac{\partial ln\left(SRPT^{\phi}_{p(v)}\right)}{\partial q_1} \propto \int_v^1 \Big(A'(G(u)) + 2q_1 G(u)\Big)G'(u)udu. \tag{70}$$

Recalling that $G'(u) < 0\ \forall\ u$, we see that this derivative is negative, implying that lower pass-through of a common cost shock is associated with higher q_1 when $q_1 + q_2$ is held constant.

Proof of Proposition 6

Using equation (24), we can derive the elasticity of $p(v)$ with respect to $\tilde{p}$, holding $\tilde{z}$ fixed:

$$SRPT^{\tilde{p}}_{p(v)} = \frac{\tilde{p}q_1}{\tilde{p}q_1 - \left(\frac{\phi}{\tilde{z}}\right)\int_v^1 h(u)du} = \frac{\tilde{p}q_1}{p(v)A(G(v))}. \tag{71}$$

From (68), we see that $SRPT^{\tilde{p}}_{p(v)} = 1 - SRPT^{\phi}_{p(v)}$. Hence, the remaining results of the proposition follow from Proposition 5.

Proof of Proposition 7

(i.) Using equations (15) and (24), we can derive the elasticity of $p(v)$ with respect to ϕ, taking into account the effect on $\tilde{z}$ and the effect of a change in $\tilde{z}$ on $p(v)$ as our measure of long-run pass-through of a change in ϕ.

$$LRPT^{\phi}_{p(v)} = SRPT^{\phi}_{p(v)} + \left(\epsilon^{v}_{p(v)}\epsilon^{\tilde{z}}_{v} + \epsilon^{\tilde{z}}_{p(v)}\right)\epsilon^{\phi}_{\tilde{z}}, \tag{72}$$

where $\epsilon^{x}_{y(x)}$ denotes the elasticity of $y(x)$ with respect to x. Since $v = \frac{\tilde{z}}{z}$, we have $\epsilon^{\tilde{z}}_{v} = 1$, and from (14), we have $\epsilon^{\phi}_{\tilde{z}} = 1$. On inspection of (24), we have $\epsilon^{\tilde{z}}_{p(v)} = -SRPT^{\phi}_{p(v)}$, so the above expression can be written as

$$LRPT^{\phi}_{p(v)} = \epsilon^{v}_{p(v)} = \left[\frac{h(v)}{p(v)A(G(v))}\right]\left[\left(\frac{\phi}{z}\right) - p(v)\right]. \tag{73}$$

Recalling that $h(u) < 0\ \forall\ u$, the first term must be negative. Because all retailers price above their

marginal cost, the second term is also negative. Hence, $LRPT^{\phi}_{p(v)} > 0$.
(ii.) Using equations (15) and (24), we can derive the elasticity of $p(v)$ with respect to $\tilde{p}$, taking into account the effect on $\tilde{z}$ and the effect of a change in $\tilde{z}$ on $p(v)$ as our measure of long-run pass-through of a change in $\tilde{p}$:

$$LRPT^{\tilde{p}}_{p(v)} = SRPT^{\tilde{p}}_{p(v)} + \left(\epsilon^{v}_{p(v)}\epsilon^{\tilde{z}}_{v} + \epsilon^{\tilde{z}}_{p(v)}\right)\epsilon^{\tilde{p}}_{\tilde{z}}, \tag{74}$$

where $\epsilon^{x}_{y(x)}$ denotes the elasticity of $y(x)$ with respect to x. Since $v = \frac{\tilde{z}}{z}$, we have $\epsilon^{\tilde{z}}_{v} = 1$ and from (14), we have $\epsilon^{\tilde{p}}_{\tilde{z}} = -mkup(1)$. From Proposition 6, we have $SRPT^{\tilde{p}}_{p(v)} = 1 - SRTP^{\phi}_{p(v)}$ and from the proof of part (i.) of this proposition, we have $\epsilon^{v}_{p(v)} = LRPT^{\phi}_{p(v)}$. On inspection of (24), we have $\epsilon^{\tilde{z}}_{p(v)} = -SRPT^{\phi}_{p(v)}$, so the above expression can be written as

$$LRPT^{\tilde{p}}_{p(v)} = 1 + (mkup(1) - 1)\, SRPT^{\phi}_{p(v)} - mkup(1)LRPT^{\phi}_{p(v)}. \tag{75}$$

From Proposition 5, we have $SRPT^{\phi}_{p(v)} > 0$ and from part (i.) of this proposition, we have $LRPT^{\phi}_{p(v)} > 0$. Hence, because $mkup(1) > 1$, the sum of the last two terms in the above expression is negative. Thus, we cannot unambiguously determine the sign of long-run pass-through to retailer-level prices in response to a change in $\tilde{p}$, and numerical simulations presented in Section 8 demonstrate that this pass-through can be of either sign.
(iii.) Using equations (15) and (24), we can derive the elasticity of $p(v)$ with respect to δ, taking into account the effect on $\tilde{z}$ and the effect of a change in $\tilde{z}$ on $p(v)$ as our measure long-run pass-through of a change in δ:

$$LRPT^{\delta}_{p(v)} = \left(\epsilon^{v}_{p(v)}\epsilon^{\tilde{z}}_{v} + \epsilon^{\tilde{z}}_{p(v)}\right)\epsilon^{\delta}_{\tilde{z}}, \tag{76}$$

where $\epsilon^{x}_{y(x)}$ denotes the elasticity of $y(x)$ with respect to x. Since $v = \frac{\tilde{z}}{z}$, we have $\epsilon^{\tilde{z}}_{v} = 1$ and from (14), we have $\epsilon^{\delta}_{\tilde{z}} = -mkup(1) - 1$. From the proof of part (i.) of this proposition, we have $\epsilon^{v}_{p(v)} = LRPT^{\phi}_{p(v)}$. On inspection of (24), we have $\epsilon^{\tilde{z}}_{p(v)} = -SRPT^{\phi}_{p(v)}$, so the above expression can be written as

$$LRPT^{\tilde{p}}_{p(v)} = (mkup(1) - 1)\left(LRPT^{\phi}_{p(v)} - SRPT^{\phi}_{p(v)}\right). \tag{77}$$

We cannot determine the sign of the second term in this expression. Thus, we cannot unambiguously determine the sign of long-run pass-through to retailer-level prices in response to a change in δ, and numerical simulations presented in Section 8 demonstrate that this pass-through can be of either sign.

Proof of Proposition 8

(i.) Differentiating equations (39) and (40) with respect to ϕ, holding $\tilde{z}$ constant, gives

$$SRPT^{\phi}_{\bar{p}_{post}} = \frac{-\int_0^1 SRPT^{\phi}_{p(v)} p(v) G'(v) dv}{\bar{p}_{post}}, \tag{78}$$

$$SRPT^{\phi}_{\bar{p}_{tran}} = \frac{-\int_0^1 SRPT^{\phi}_{p(v)} p(v) A(G(v)) G'(v) dv}{\bar{p}_{tran}}. \tag{79}$$

From Proposition 5, we have $SRPT^{\phi}_{p(v)}, > 0$ and recalling that $G'(v) < 0$, we can conclude that both of these expressions are positive. Recalling that $\phi/\tilde{z}$ does not change when ϕ changes, inspection of (24) and (39) gives us the result that long-run pass-through of a common cost shock to average prices is zero.

(ii.) Differentiating the expressions in equation (39) with respect to $\tilde{p}$ holding $\tilde{z}$ constant gives

$$SRPT^{\tilde{p}}_{\bar{p}_{post}} = \frac{-\int_0^1 SRPT^{\tilde{p}}_{p(v)} p(v) G'(v) dv}{\bar{p}_{post}}, \tag{80}$$

$$SRPT^{\tilde{p}}_{\bar{p}_{tran}} = \frac{-\int_0^1 SRPT^{\tilde{p}}_{p(v)} p(v) A(G(v)) G'(v) dv}{\bar{p}_{tran}}. \tag{81}$$

From Proposition 6, we have $SRPT^{\tilde{p}}_{p(v)} > 0$, and recalling that $G'(v) < 0$, we can conclude that both of these expressions are positive. Differentiating the average posted price in equation (39) with respect to $\tilde{p}$ allowing $\tilde{z}$ to change gives

$$LRPT^{\tilde{p}}_{\bar{p}_{post}} = \frac{-\int_0^1 \left(SRPT^{\tilde{p}}_{p(v)} + \epsilon^{\tilde{z}}_{p(v)} \epsilon^{\tilde{p}}_{\tilde{z}} \right) p(v) G'(v) dv}{\bar{p}_{post}}. \tag{82}$$

From Proposition 6, we have $SRPT^{\tilde{p}}_{p(v)} = 1 - SRPT^{\phi}_{p(v)}$ and from the proof of Proposition 7, we have $\epsilon^{\tilde{z}}_{p(v)} = -SRPT^{\phi}_{p(v)}$ and $\epsilon^{\tilde{p}}_{\tilde{z}} = -mkup(1)$. Substituting this into the above expression and rearranging gives

$$LRPT^{\tilde{p}}_{\bar{p}_{post}} = 1 + (mkup(1) - 1)\, SRPT^{\phi}_{\bar{p}_{post}}. \tag{83}$$

From Proposition 5 and since $mkup(1) > 1$, we can conclude that this expression is positive and greater than one. Using similar methods, we can derive

$$LRPT^{\tilde{p}}_{\bar{p}_{tran}} = 1 + (mkup(1) - 1)\, SRPT^{\phi}_{\bar{p}_{tran}}, \tag{84}$$

and, using the same reasoning, conclude that this expression is positive and greater than one.

(iii.) Differentiating the average posted price in equation (39) with respect to δ allowing $\tilde{z}$ to change gives

$$LRPT^{\delta}_{\bar{p}_{post}} = \frac{-\int_0^1 \epsilon^{\tilde{z}}_{p(v)} \epsilon^{\delta}_{\tilde{z}} p(v) G'(v) dv}{\bar{p}_{post}}. \tag{85}$$

From the proof of Proposition 7, we have $\epsilon^{\tilde{z}}_{p(v)} = -SRPT^{\phi}_{p(v)}$ and $\epsilon^{\delta}_{\tilde{z}} = mkup(1) - 1$. Substituting this into the above expression and rearranging gives

$$LRPT^{\delta}_{\bar{p}_{post}} = (1 - mkup(1))\, SRPT^{\phi}_{\bar{p}_{post}}. \tag{86}$$

Because $mkup(1) > 1$, we can conclude that this expression is negative. Using similar methods, we can derive

$$LRPT^{\delta}_{\bar{p}_{tran}} = (1 - mkup(1))\, SRPT^{\phi}_{\bar{p}_{tran}}, \tag{87}$$

and, using the same reasoning, conclude that this expression is negative.
(iv.) Comparing (83) with (86) and comparing (84) with (87), we see that

$$LRPT^{\delta}_{\bar{p}_{post}} = 1 - LRPT^{\tilde{p}}_{\bar{p}_{post}} \qquad\qquad LRPT^{\delta}_{\bar{p}_{tran}} = 1 - LRPT^{\tilde{p}}_{\bar{p}_{tran}}. \tag{88}$$

Appendix C: Alternate Model Environment

Using equations (24) and (39), we can derive the following expression for long-run pass-through of a ϕ shock to the average posted price:

$$LRPT^{\phi}_{\bar{p}_{post}} = \frac{\left(\epsilon^{\phi}_{\tilde{z}} - 1\right) \int_0^1 SRPT^{\phi}_{p(v)} p(v) G'(v) dv}{\bar{p}_{post}}, \tag{89}$$

where $\epsilon^{\phi}_{\tilde{z}}$ is the elasticity of $\tilde{z}$ with respect to ϕ. Recalling that $SRPT^{\phi}_{p(v)} > 0$ and $G'(v) < 0$, we see that the sign of $LRPT^{\phi}_{\bar{p}_{post}}$ is determined by the sign of $1 - \epsilon^{\phi}_{\tilde{z}}$. This is intuitive: whether or not the average posted price rises, falls, or is unchanged in the long run when ϕ changes depends on what happens to average cost, and average cost is monotonically increasing in $\phi/\tilde{z}$ as shown in footnote 28. In the model where fixed costs are proportional to market tightness, $\epsilon^{\phi}_{\tilde{z}} = 1$ and long-run pass-through of a ϕ shock to average prices equals zero.

Consider an environment in which Ψ is exogenous and unrelated to μ, implying $\delta \equiv \frac{\Psi}{\mu}$ is now endogenous and varies inversely with μ. We may use the free-entry condition given by (19) to write δ as a function of ϕ and $\tilde{z}$:

$$\delta(\phi, \tilde{z}) = \frac{\tilde{p} q_1 (1 - J(\tilde{z})) + \phi \left(\int_{\tilde{z}}^{z_H} \left(\int_{\tilde{z}}^{z} \left(\frac{A'(F(x))F'(x)}{x} \right) dx - \frac{A(F(z))}{z} \right) J'(z) dz \right)}{\frac{f_e}{\Psi} + (1 - J(\tilde{z}))}. \tag{90}$$

We may use the zero-profit condition given by equation (14) to derive the following expression, which implicitly determines the elasticity of $\tilde{z}$ with respect to ϕ:

$$\epsilon^{\phi}_{\tilde{z}} = 1 + \left(\frac{1}{\alpha q_1}\right) \left(\epsilon^{\phi}_{\delta} + \epsilon^{\tilde{z}}_{\delta} \epsilon^{\phi}_{\tilde{z}}\right), \tag{91}$$

where $\alpha \equiv \frac{\phi}{\delta \tilde{z}}$. Rearranging this gives

$$\epsilon^{\phi}_{\tilde{z}} = \frac{\alpha q_1 + \epsilon^{\phi}_{\delta}}{\alpha q_1 - \epsilon^{\tilde{z}}_{\delta}}. \tag{92}$$

From equation (90), we can derive the following elasticities:

$$\epsilon_{\delta}^{\phi} = \frac{\frac{f_e}{\Psi} - \alpha q_1 \left(1 - J(\tilde{z})\right)}{D} \qquad \epsilon_{\delta}^{\tilde{z}} = \frac{-2\alpha q_2 G'(1) \left(1 - J(\tilde{z})\right)}{D}, \tag{93}$$

where $D \equiv \frac{f_e}{\Psi} + (1 - J(\tilde{z})) > 0$. Substituting these into equation (92) gives

$$\epsilon_{\tilde{z}}^{\phi} = \left(\frac{1 + \alpha q_1}{\alpha q_1}\right) \left(\frac{\frac{f_e}{\Psi}}{\frac{f_E}{\Psi} + \left(1 - 2\left(\frac{q_2}{q_1}\right) G'(1)\right)(1 - J(\tilde{z})}\right). \tag{94}$$

From the above equation, we note that $\epsilon_{\tilde{z}}^{\phi} < 1$ if and only if

$$\frac{f_e}{\Psi} < \alpha q_1 \left(1 - 2\left(\frac{q_2}{q_1}\right) G'(1)\right)(1 - J(\tilde{z})). \tag{95}$$

Substituting in α and recalling the definition of $G(v)$, the condition can be written as[33]

$$f_e < (\tilde{p}\mu q_1 - \Psi)\left(1 + 2\left(\frac{q_2}{q_1}\right) F'(\tilde{z})\tilde{z}\right)(1 - J(\tilde{z})). \tag{96}$$

Hence, we can conclude that if entry costs are below this level, then long-run pass-through to the average posted price of a common cost shock will be positive when the fixed costs of operation are exogenous. The intuition is that if entry costs are not too high, then the extensive margin response of a fall in the measure of active firms in response to a rise in ϕ will be sufficiently large so that the resulting rise in μ will induce a large enough fall in $\tilde{z}$ so that the average variable cost rises, thereby raising average prices.

Appendix D: Equilibrium Variables Written as Functions of Estimated Parameters

In this appendix, we present expressions for scaled price, scaled revenue, markups, and price pass-through rates as functions of the following estimated parameters alone: q_1, q_2, ν, γ, and $\alpha \equiv \frac{\phi}{\delta\tilde{z}}$. We first note that $G(v)$ and $G'(v)$ depend only on the shape parameter of the Pareto distribution, γ. Second, $A(G(v))$ and $A'(G(v))$ depend only on the parameters of the price quote distribution, q_1, q_2, ν, and the shape parameter of the Pareto distribution, γ.

Using (24), we may write the ratio of $p(v)/\delta$ as a function of estimated parameters alone, and we refer to this ratio as scaled price:

$$\frac{p(v)}{\delta} = \left(\frac{1}{A\big(G(v)\big)}\right)\left(1 + \alpha q_1 - \alpha \int_v^1 h(u)du\right), \tag{97}$$

[33] If productivity follows a Pareto distribution with shape parameter equal to γ, this condition becomes $f_e < (\tilde{p}\mu q_1 - \Psi)\left(1 + 2\gamma\left(\frac{q_2}{q_1}\right)\right)\left(\frac{z_L}{\tilde{z}}\right)^{\gamma}$.

where recall that $h(u) = uA'(G(u))G'(u)$. We may also write $R(v)/\Psi$, scaled revenue, as a function of estimated parameters alone:

$$\frac{R(v)}{\Psi} = \left(\frac{p(v)}{\delta}\right) A(G(v)). \tag{98}$$

From equation (97), we can derive the markup as a function of estimated parameters alone:

$$mkup(v) = \left(\frac{1}{\alpha v A\big(G(v)\big)}\right)\left(1 + \alpha q_1 - \alpha \int_v^1 h(u)du\right). \tag{99}$$

From equations (39) and (40), we have

$$\frac{\bar{p}_{post}}{\delta} = -\int_0^1 \left(\frac{p(v)}{\delta}\right) G'(v)dv \qquad \frac{\bar{p}_{tran}}{\delta} = -\int_0^1 \left(\frac{p(v)}{\delta}\right) A(G(v))G'(v)dv. \tag{100}$$

Using the expressions in Appendix B, we can derive the following pass-through functions as function of estimated parameters alone:

$$SRPT^{1/z}_{p(v)} = LRPT^{\phi}_{p(v)} = \frac{h(v)\left(\alpha v - \left(\frac{p(v)}{\delta}\right)\right)}{\left(\frac{p(v)}{\delta}\right) A(G(v))}, \tag{101}$$

$$SRPT^{\phi}_{p(v)} = \frac{-\alpha \int_v^1 h(u)du}{\left(\frac{p(v)}{\delta}\right) A(G(v))}, \tag{102}$$

$$SRPT^{\tilde{p}}_{p(v)} = \frac{1 + \alpha q_1}{\left(\frac{p(v)}{\delta}\right) A(G(v))}, \tag{103}$$

$$SRPT^{\delta}_{p(v)} = 0, \tag{104}$$

$$LRPT^{\tilde{p}}_{p(v)} = SRPT^{\tilde{p}}_{p(v)} - \left(\frac{1 + \alpha q_1}{\alpha q_1}\right)\left(LRPT^{\phi}_{p(v)} + \frac{\alpha \int_v^1 h(u)du}{\left(\frac{p(v)}{\delta}\right) A(G(v))}\right), \tag{105}$$

$$LRPT^{\delta}_{p(v)} = \left(\frac{1}{\alpha q_1}\right)\left(LRPT^{\phi}_{p(v)} + \frac{\alpha \int_v^1 h(u)du}{\left(\frac{p(v)}{\delta}\right) A(G(v))}\right), \tag{106}$$

$$SRPT^{\phi}_{\bar{p}_{post}} = \frac{\alpha \int_0^1 \left(\frac{h(v)}{A(G(v))}\right) G'(v)dv}{\left(\frac{\bar{p}_{post}}{\delta}\right)}, \tag{107}$$

$$SRPT^{\phi}_{\bar{p}_{tran}} = \frac{\alpha \int_0^1 h(v)G'(v)dv}{\left(\frac{\bar{p}_{tran}}{\delta}\right)}, \tag{108}$$

$$SRPT^{\tilde{p}}_{\bar{p}_{post}} = \frac{(1+\alpha q_1)\int_0^1 \left(\frac{1}{A(G(v))}\right) G'(v)dv}{\left(\frac{\bar{p}_{post}}{\delta}\right)}, \tag{109}$$

$$SRPT^{\tilde{p}}_{\bar{p}_{tran}} = \frac{(1+\alpha q_1)\int_0^1 G'(v)dv}{\left(\frac{\bar{p}_{tran}}{\delta}\right)}, \tag{110}$$

$$LRPT^{\tilde{p}}_{\bar{p}_{post}} = \frac{\left(\frac{1+\alpha q_1}{\alpha q_1}\right)\int_0^1 \left(\frac{p(v)}{\delta} - \frac{1}{A(G(v))}\right) G'(v)dv}{\left(\frac{\bar{p}_{post}}{\delta}\right)}, \tag{111}$$

$$LRPT^{\tilde{p}}_{\bar{p}_{tran}} = \frac{\left(\frac{1+\alpha q_1}{\alpha q_1}\right)\int_0^1 \left(\left(\frac{p(v)}{\delta}\right) A(G(v)) - 1\right) G'(v)dv}{\left(\frac{\bar{p}_{tran}}{\delta}\right)}, \tag{112}$$

$$LRPT^{\delta}_{\bar{p}_{post}} = \frac{\left(\frac{1}{\alpha q_1}\right)\int_0^1 \left(\frac{p(v)}{\delta} - \frac{(1+\alpha q_1)}{A(G(v))}\right) G'(v)dv}{\left(\frac{\bar{p}_{post}}{\delta}\right)}, \tag{113}$$

$$LRPT^{\delta}_{\bar{p}_{tran}} = \frac{\left(\frac{1}{\alpha q_1}\right)\int_0^1 \left(\left(\frac{p(v)}{\delta}\right) A(G(v)) - (1+\alpha q_1)\right) G'(v)dv}{\left(\frac{\bar{p}_{tran}}{\delta}\right)}. \tag{114}$$

Appendix E: Sub-Sector-Specific Simulation Results

Figure 9: Log of Markups

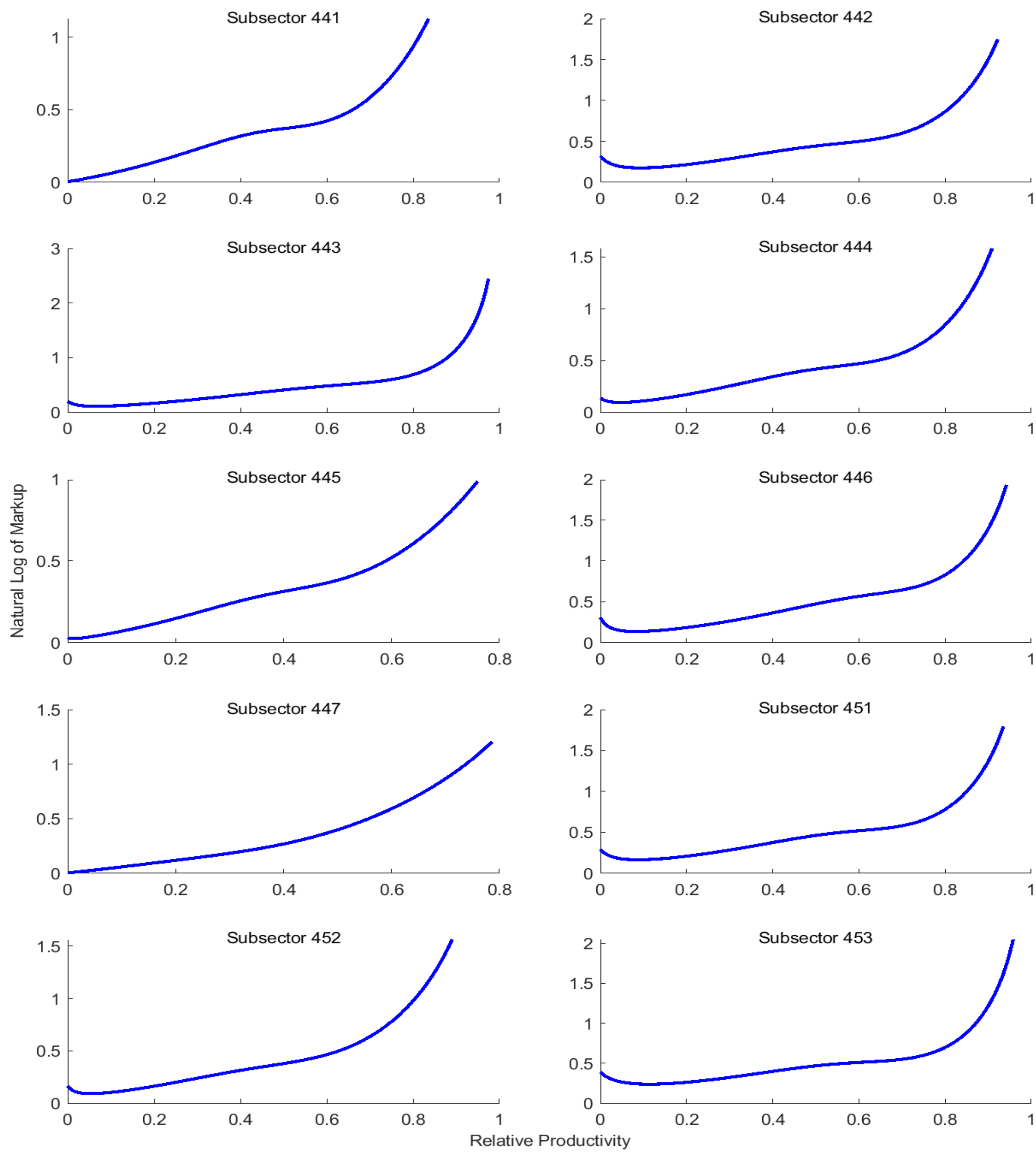
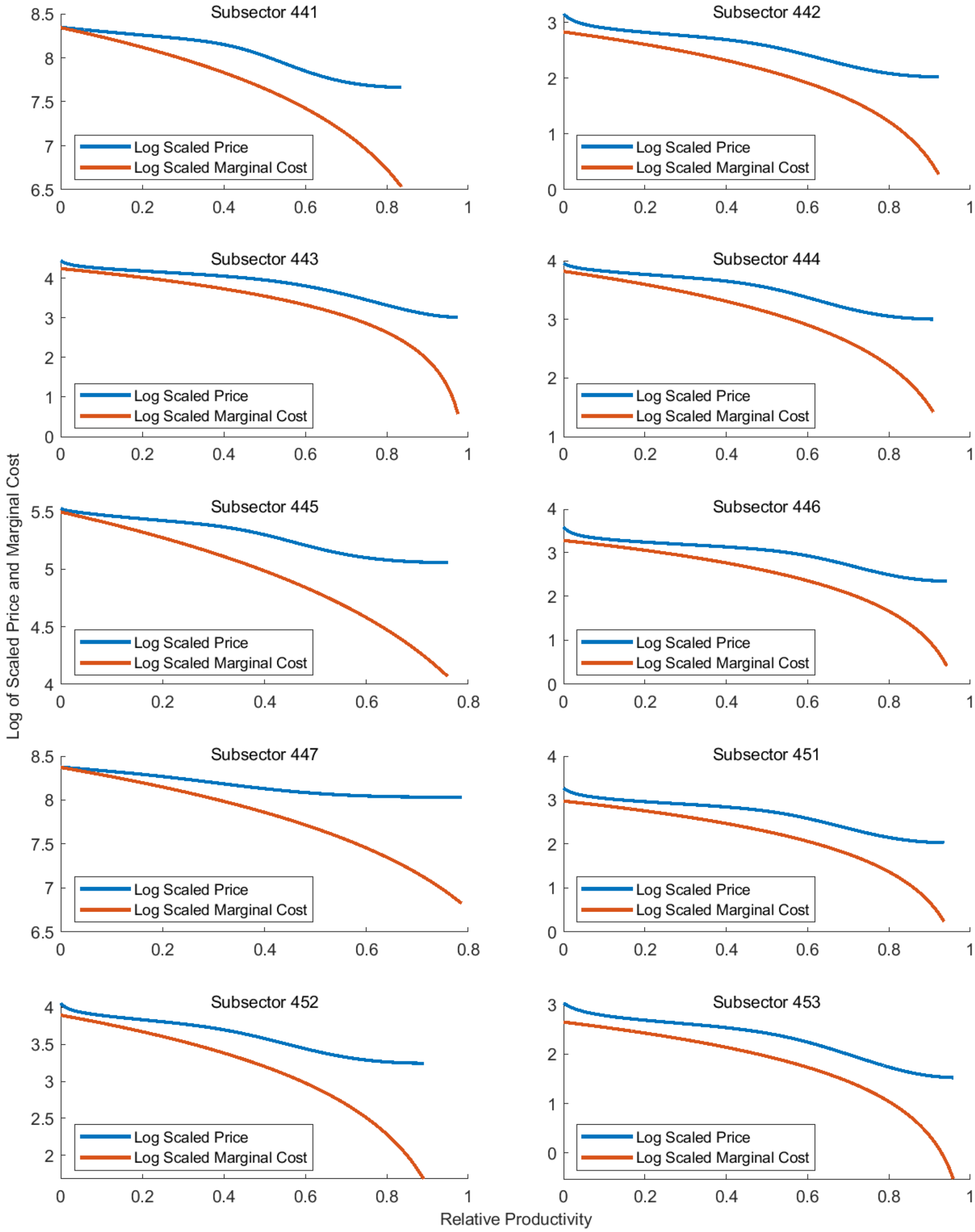
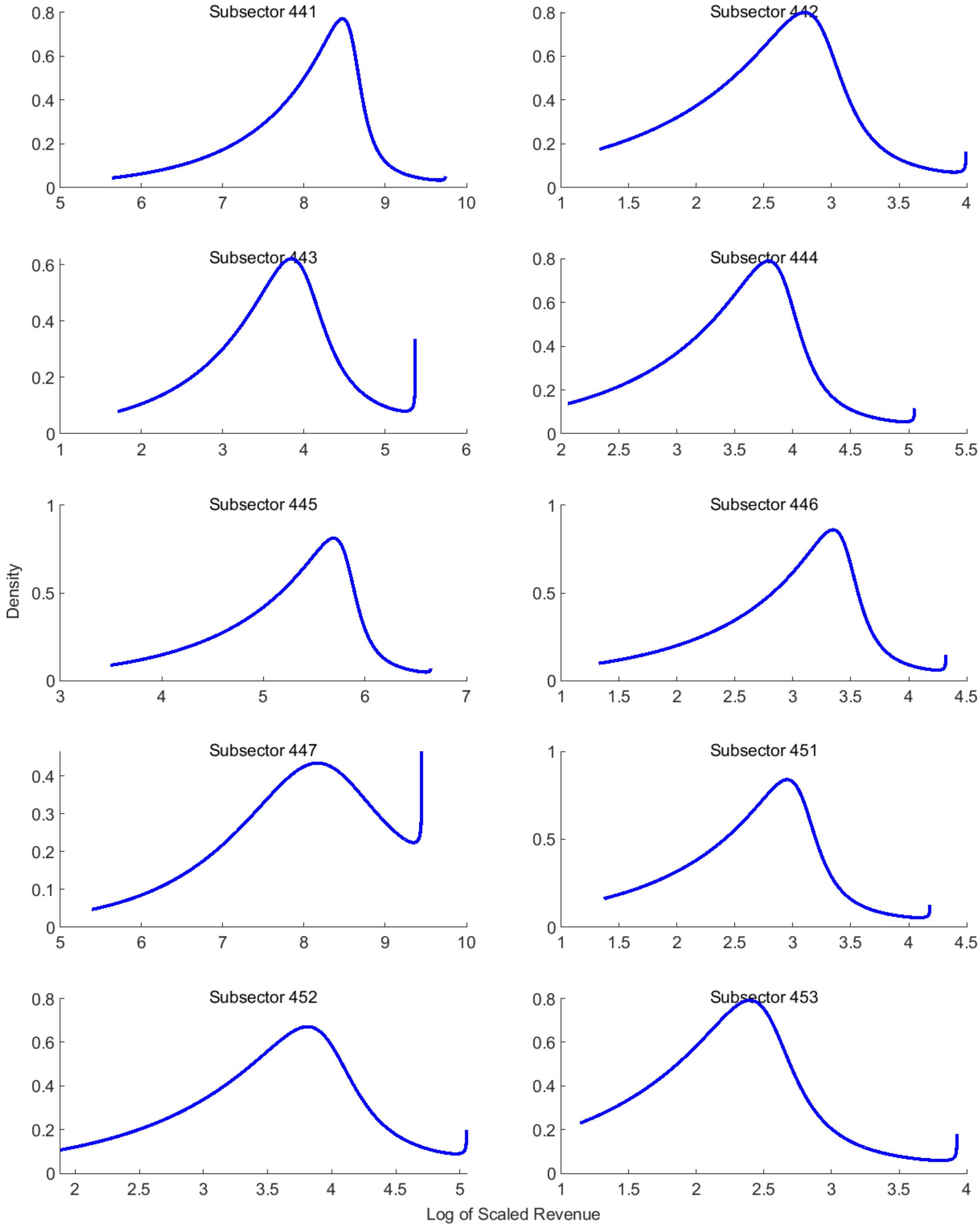


Figure 10: Log of Scaled Price and Marginal Cost



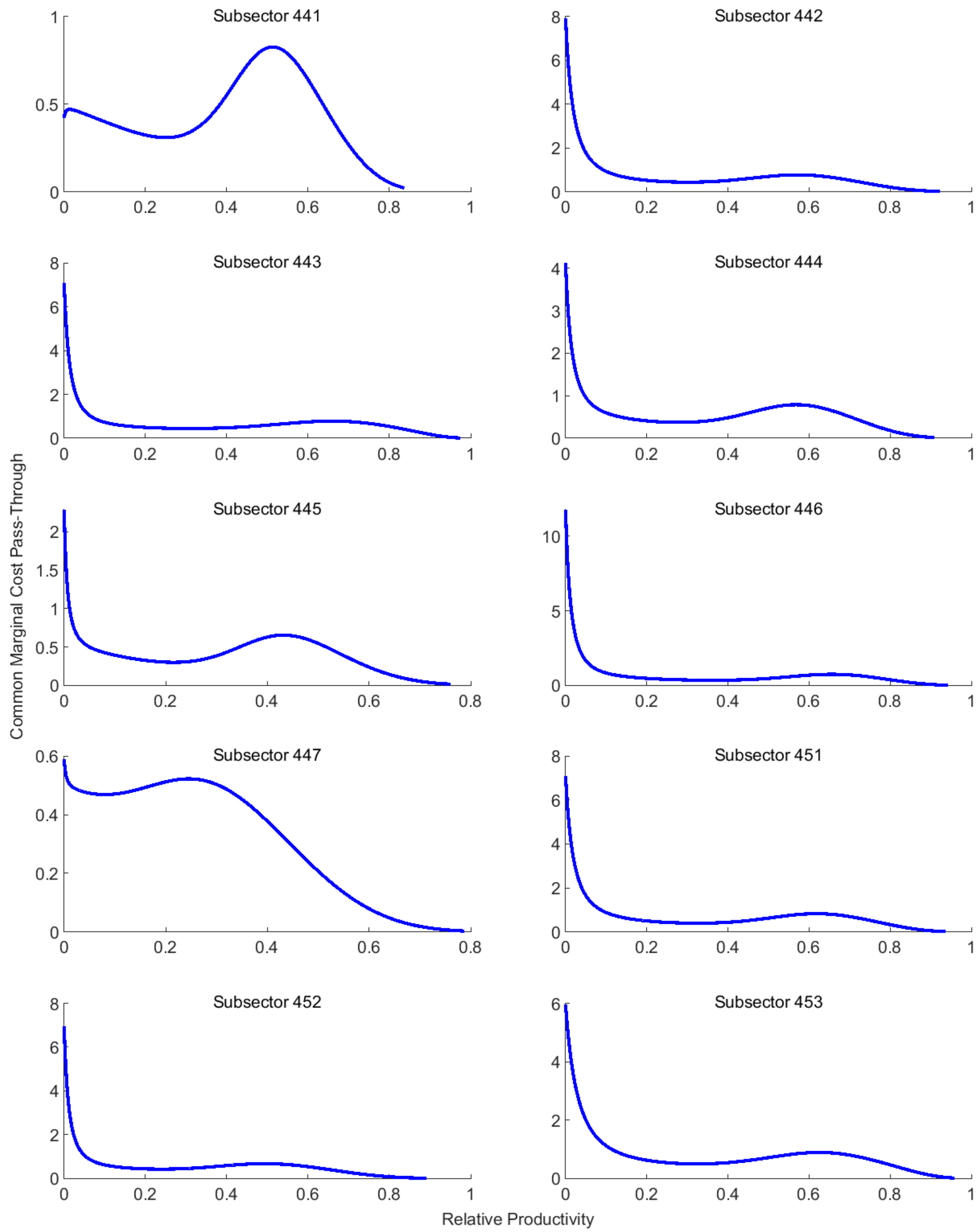
Notes: Scaled price and marginal cost are defined as those variables divided by δ, and the difference between these two variables is independent of δ.

Figure 11: Log of Scaled Revenue



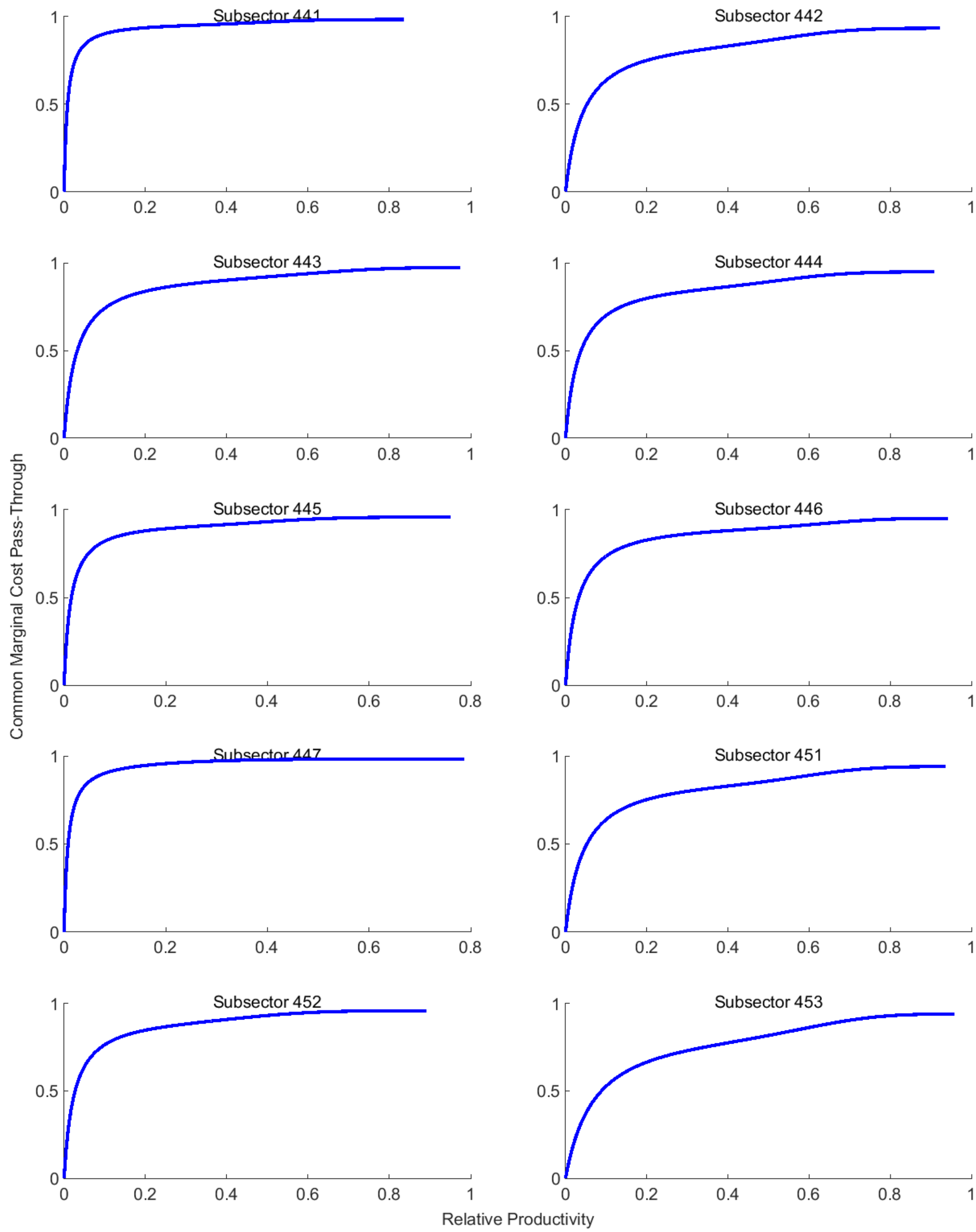
Notes: Scaled revenue is defined as implicitly divided by fixed costs, Ψ.

Figure 12: Short-Run Price Pass-Through of a Firm-Specific Change in Productivity



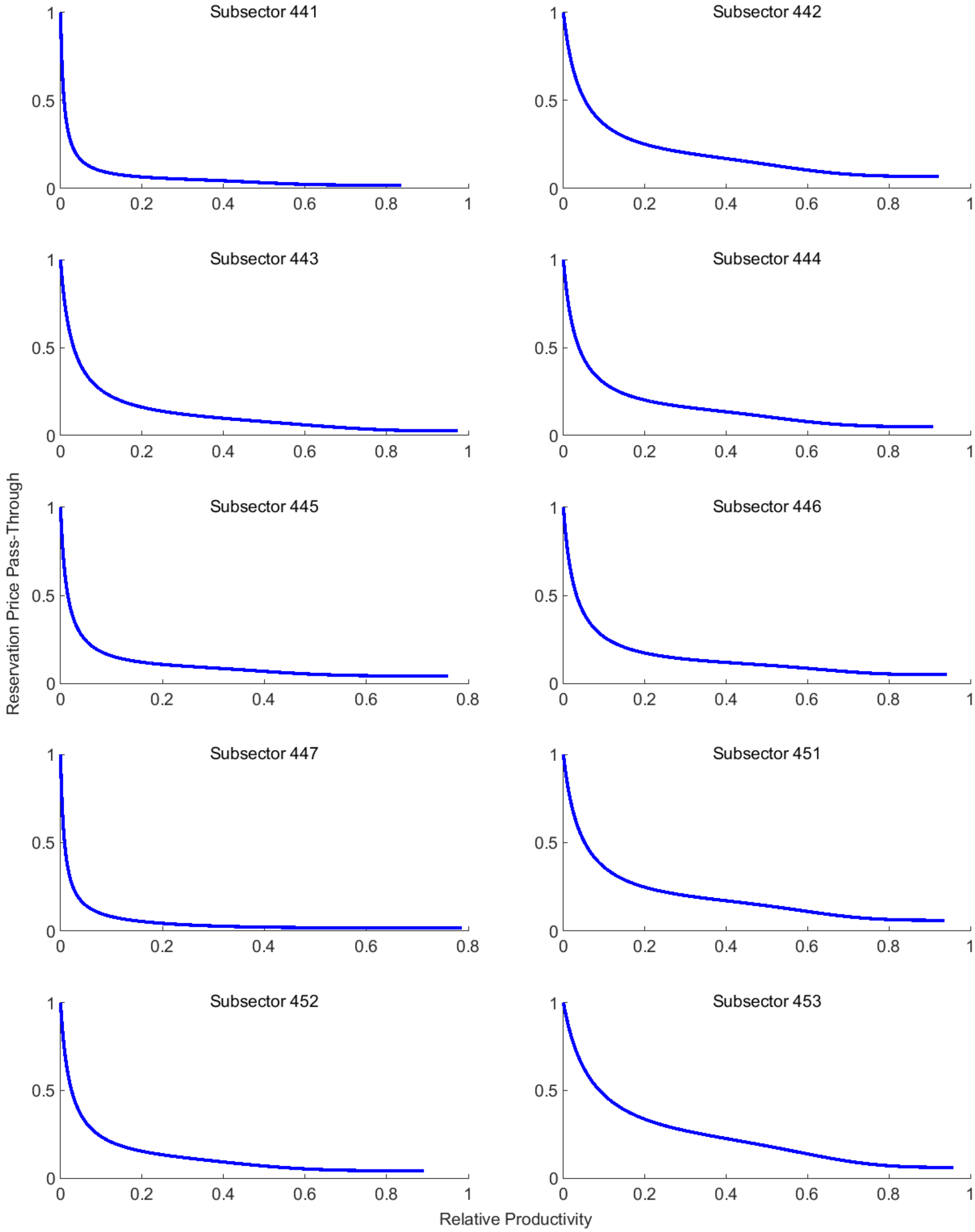
Notes: The calculations use the pass-through equations from Appendix D and the parameter estimates from Table 3.

Figure 13: Short-Run Price Pass-Through of Common Cost Change (ϕ)



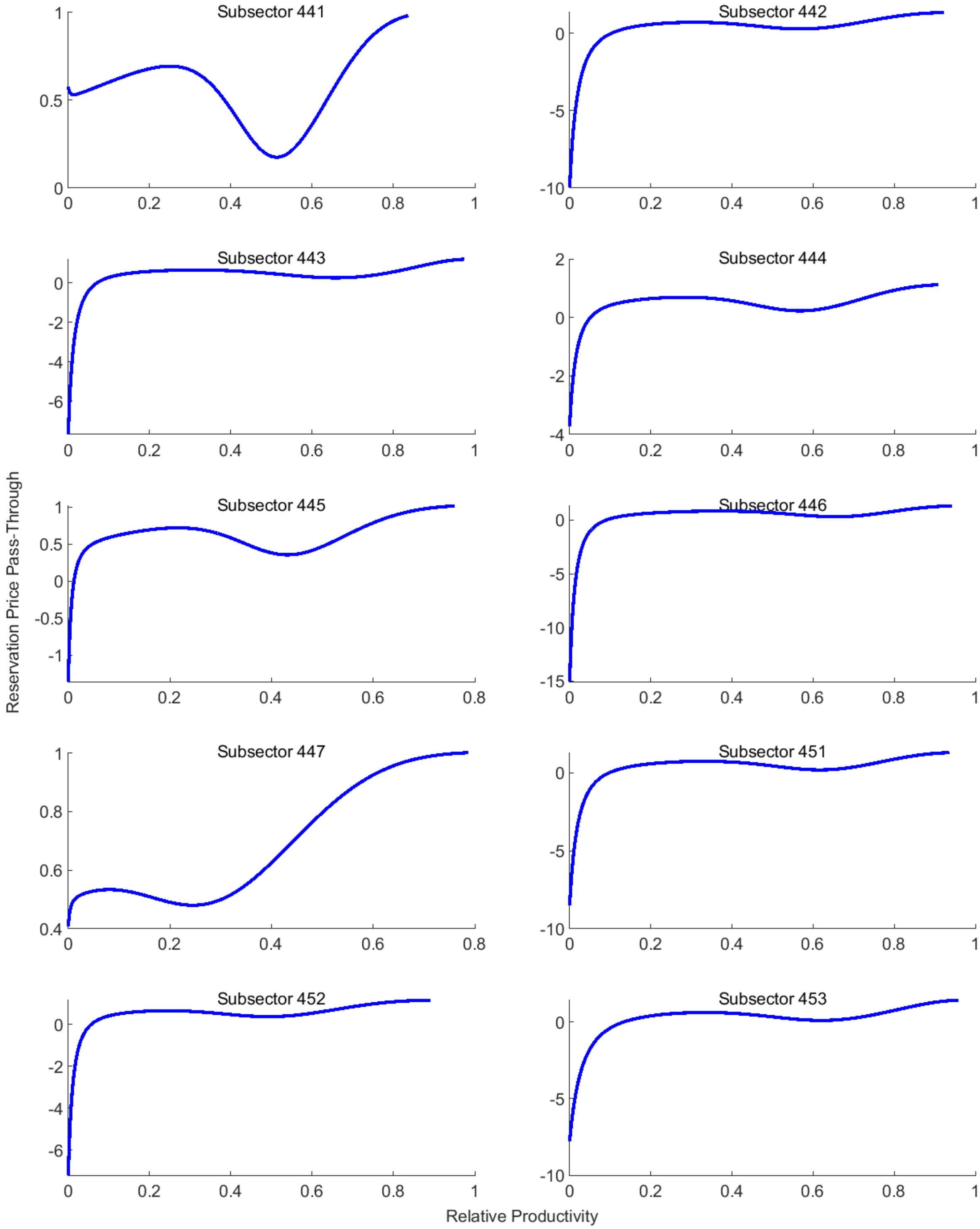
Notes: The calculations use the pass-through equations from Appendix D and the parameter estimates from Table 3.

Figure 14: Short-Run Pass-Through of a Reservation Price Change ($\tilde{p}$)



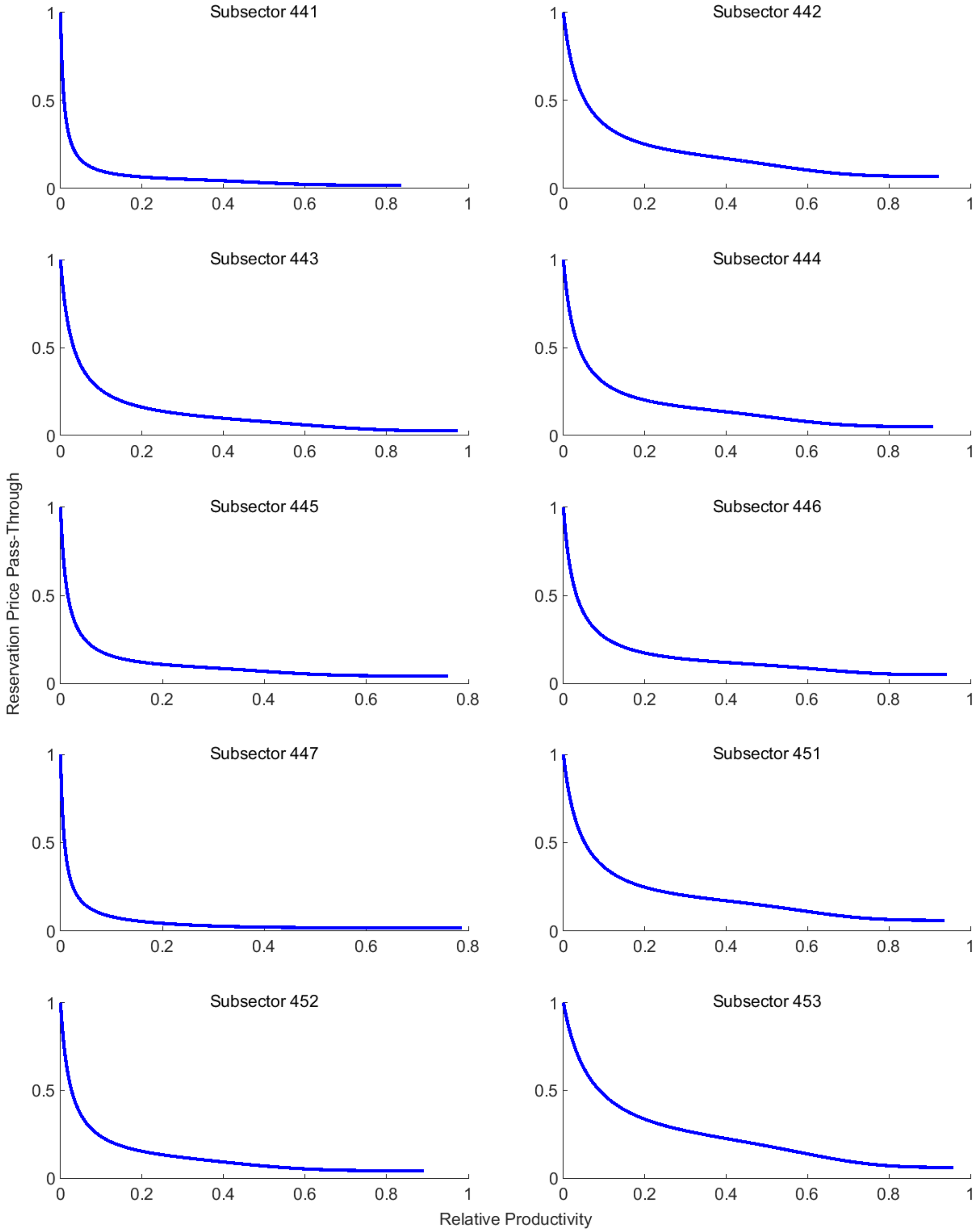
Notes: The calculations use the pass-through equations from Appendix D and the parameter estimates from Table 3.

Figure 15: Long-Run Pass-Through of a Reservation Price Change ($\tilde{p}$)



Notes: The calculations use the pass-through equations from Appendix D and the parameter estimates from Table 3.

Figure 16: Long-Run Price Pass-Through of Change in δ



Notes: The calculations use the pass-through equations from Appendix D and the parameter estimates from Table 3.