Kyoto University,
Graduate School of Economics
Discussion Paper Series

Pass-Through with Quantity Discounts:

A New Edgeworth-Salinger Paradox

Takanori Adachi, Naoshi Doi

Discussion Paper No. E-26-002

Graduate School of Economics
Kyoto University
Yoshida-Hommachi, Sakyo-ku
Kyoto City, 606-8501, Japan

September 2026

Pass-Through with Quantity Discounts: A New Edgeworth-Salinger Paradox*

Takanori Adachi[†] Naoshi Doi[‡]

September 25, 2026

Abstract

We study cost pass-through under nonlinear pricing when a monopolistic seller offers a single good in two fixed package sizes to heterogeneous consumers. Building on the Edgeworth–Salinger paradox of taxation, we show that Edgeworth-type price responses can arise across package sizes of the same underlying product. A common increase in per-unit marginal cost may induce the seller to lower the per-unit price of the small package while raising that of the large package. Under multinomial logit demand, pass-through to the large package is analytically greater than pass-through to the small package. Numerical analysis shows that this ordering is remarkably robust to random-coefficient demand heterogeneity. Negative pass-through for the small package is quantitatively important and becomes substantially more likely as the difference between package sizes increases. Consumer sorting also matters: a more positive correlation between price sensitivity and preferences for larger packages increases the likelihood of negative pass-through for the small package and widens the pass-through gap across package sizes. Thus, nonlinear pricing can transform a common supply-side cost shock into sharply different price responses across quantities purchased, with potentially important implications for the distributional incidence of cost shocks.

Keywords: Pass-through; Quantity discounts; The Edgeworth-Salinger paradox.

JEL Classification: D42; D43; H22; L11; L13.

---

*We are grateful to Nicolas Schutz for helpful discussion. Adachi acknowledges a Grant-in-Aid for Scientific Research (C) (25K05102) from the Japan Society for the Promotion of Science, and Doi acknowledges a Grant-in-Aid for Scientific Research (C) (23K01368) from the Society. All remaining errors are solely ours.

[†]Graduate School of Management and Graduate School of Economics, Kyoto University, Japan. E-mail: adachi.takanori.8m@kyoto-u.ac.jp

[‡]Department of Economics, Otaru University of Commerce, Japan. E-mail: doi@res.otaru-uc.ac.jp

1 Introduction

Do supply-side cost shocks affect consumers differently depending on the package size they purchase? This question arises naturally in markets with quantity discounts, where the same underlying good is sold at different per-unit prices across package sizes. A common increase in production cost need not be transmitted uniformly along such a nonlinear price schedule. Since consumers also sort across package sizes according to their price sensitivity and preferences for quantity, heterogeneous pass-through across package sizes may generate heterogeneous exposure to the same supply-side shock.

This paper shows that nonlinear pricing can generate a novel and robust form of pass-through asymmetry. We consider a monopolistic seller offering a single good in two fixed package sizes, small (S) and large (L), with a quantity discount, $p_S > p_L$, where p_j denotes the per-unit price of package $j \in \{S, L\}$. A common increase in per-unit marginal cost raises the total cost of the large package by more than that of the small package. At the same time, the seller chooses the two prices jointly because their relative levels determine how consumers sort between the small package, the large package, and the outside option.

This interaction produces two striking implications. First, pass-through to the large package is systematically greater than pass-through to the small package. Under multinomial logit demand, we verify analytically that $\rho_L > \rho_S$. Our numerical analysis indicates that this ordering remains remarkably robust in a framework of random-coefficient logit demand where consumers are heterogeneous in both price sensitivity and preferences for package size. Second, pass-through to the small package can be *negative*: following an increase in marginal cost, the seller may optimally reduce p_S while increasing p_L. Thus, a negative price response can arise across package sizes of the same product even though the cost shock is common and package sizes themselves are fixed.

We call this finding a new version of the Edgeworth–Salinger paradox because it extends the Edgeworth–Salinger logic to nonlinear pricing within a single product category. Building on an insight first identified by Edgeworth (1925), Salinger (1991) shows that a cost or tax change affecting one product can generate seemingly paradoxical price responses for another product supplied by the same firm. Luco and Marshall (2020) provide empirical evidence of a closely related mechanism in the U.S. carbonated-beverage industry. More recently, Armstrong and Vickers (2023) develop a general characterization of such cross-price effects in multiproduct environments, while Vickers (2026) places these results in the broader context of multiproduct pricing and pass-through. These analyses, however, typically concern firms choosing a vector of linear prices for distinct products. We instead show that a related effect can arise across fixed package sizes of the same underlying good. In our setting, the relevant demand interdependence operates through consumer substitution and sorting across different points on a nonlinear pricing schedule, rather than through cross-price effects among distinct goods under linear pricing.

The underlying mechanism combines differential cost exposure with consumer sorting. A one-unit increase in marginal cost raises the cost of supplying the small package by one unit but the cost of supplying the large package by $\bar{x} > 1$ units. The seller therefore has an incentive to adjust the relative prices of the two packages rather than simply shifting both per-unit prices by the same amount. Because consumers substitute across package sizes when relative prices change, the optimal response may involve absorbing part of the cost shock in the small-package margin, or even lowering the small-package price, while passing a larger fraction of the shock through to the large package. The phenomenon is therefore generated by adjustment of the *entire nonlinear price schedule*, rather than by independent pricing of two unrelated products.

Our numerical analysis reveals that the difference between package sizes is a first-order determinant of this Edgeworth-type response. In the baseline random-coefficients specification, negative small-package pass-through occurs in approximately one-half of the parameter configurations considered. When the large package contains only 1.5 times as much as the small package, negative pass-through is relatively rare. When the size ratio rises to 2.5, however, it occurs in roughly 70 percent or more of the configurations in our robustness exercises. By contrast, changing the intensity in utility from package size has a substantially smaller effect. These findings suggest that the dispersion of quantities offered along the nonlinear pricing schedule is central to whether a common cost shock generates an Edgeworth-type price response.

Consumer heterogeneity provides an additional comparative-static dimension. Let $\tau = \mathrm{Corr}(\beta_i, \gamma_i)$ denote the correlation between consumers' price sensitivity and their preference for larger package sizes. In our numerical analysis, a more positive value of τ systematically increases the incidence of negative small-package pass-through and widens the average pass-through gap $\rho_L - \rho_S$. Thus, even when the qualitative ranking $\rho_L > \rho_S$ is unchanged, the joint distribution of price sensitivity and size preferences materially affects the quantitative incidence of a cost shock.

Recent work on pass-through and quantity discounts provides complementary methodological perspectives. On the pass-through side, Koh (2026) develops a tractable framework for analyzing cost pass-through with high-dimensional sets of differentiated products, highlighting the roles of substitution, demand curvature, and multiproduct ownership. On the nonlinear-pricing side, Iaria and Wang (2026) develop an empirical framework for studying quantity discounts with many products and package sizes. Our analysis connects these two perspectives by studying how a common cost shock is transmitted through a nonlinear price schedule and how consumer sorting across package sizes shapes that response.

Our analysis is also related to the literature on second-degree price discrimination and nonlinear pricing (Mussa and Rosen, 1978; Wilson, 1993; Armstrong, 2016). This literature emphasizes how firms construct menus of price–quantity combinations to screen heterogeneous consumers. Conlon (2017), for example, shows that buyer heterogeneity can affect the slope of the profit-maximizing quantity-discount schedule. Attanasio and Pastorino (2020) study the distributional consequences of nonlinear pricing when consumers differ in their tastes and ability to pay, while Bornstein and Peter (2025) highlight consumer-level misallocation generated by nonlinear pricing. We complement this literature by asking how a given nonlinear pricing schedule adjusts to a common supply-side cost shock and how this response varies across the quantities purchased by heterogeneous consumers.

Finally, the paper relates to the broader literature on pass-through and incidence under imperfect competition. Weyl and Fabinger (2013) develop pass-through as a general tool for studying the incidence of cost and tax shocks, while Adachi and Fabinger (2022) derive sufficient-statistics formulas linking pass-through, demand curvature, and welfare in imperfectly competitive markets. Our analysis complements this approach by considering a seller that chooses a nonlinear pricing schedule rather than a single price or a vector of independent linear prices. The relevant object is therefore a *pass-through schedule*: the response of per-unit prices at different points of the quantity menu to a common cost shock. Consumer sorting along that schedule determines how these price responses are distributed across buyers.

The remainder of the paper is organized as follows. Section 2 presents the model and derives the analytical characterization of pass-through. Section 3 examines the quantitative implications under multinomial and random-coefficients logit demand, with particular attention to package-size differences and consumer heterogeneity. Section 4 extends the baseline model to non-constant marginal costs and shows that, under additive marginal-cost shocks, production-cost curvature scales package-specific pass-through while preserving the qualitative asymmetry emphasized in the baseline analysis. Section 5 concludes.

2 Model

We consider a monopolistic retailer that provides a quantity menu for each consumer, $\{1, \bar{x}\}$ with the prices p_S and p_L, respectively. We assume that $\bar{x}$ is exogenous: is given to the retailer: it would be plausible to suppose that it is given to the retailer in accordance with consumers' expectations, custom, etc. We expect that in equilibrium non-linear pricing, $p_S > p_L$, realizes. The retailer commits to $\bar{x}$.[1] Note that effectively, the monopolist is a *multi-product* retailer, albeit selling a homogeneous product.

[1]Alexandrov (2014) studies pass-through when firms can endogenously adjust package sizes to show that a cost increase can induce firms to reduce both package sizes and package prices, thereby generating negative pass-through when pass-through is measured in package prices. Per-unit prices, however, increase in his framework. In our model, however, negative pass-through occurs in the *per-unit price* of the small package. Endogenous package resizing is therefore not necessary for an Edgeworth-type response in per-unit prices.

Essentially, we formulate market demand using a discrete-choice model with random utility (see, e,g. Galichon (2026) for a recent textbook). In this sense, a package is treated as a 'product' with the menu selection (i.e., how large and how many packages are provided) is exogenously given. We believe that this formulation is tractable and can be extended to the case of oligopoly with product differentiation.[2]

2.1 Market Demand

Each consumer $i \in \mathcal{I}$ has three options and thus chooses one of $\{0, 1, \bar{x}\}$. If consumer i chooses size $j = S$ (the "Small" size), then she consumes one unit. If she chooses $j = L$ (the "Large" size), she consumes $\bar{x} > 1$ units. She can be opt out by choosing $j = 0$.

Specifically, with some constant α and $\delta \in (0, 1)$ that captures the intensity of preference for large consumption,[3] consumer i's choice-specific utility is given by:

$$u_{ij} = \begin{cases} \alpha - \beta_i \hat{p}_j + \gamma_i (x_j - 1)^\delta + \epsilon_{ij} & \text{for } j = S \text{ or } L \\ \epsilon_{i0} & \text{for } j = 0, \end{cases}$$

where $x_j = 1$ for $j = S$ and $x_j = \bar{x}$ for $j = L$. Note also that $\hat{p}_j := p_j \cdot x_j$ is the total payment, and the two coefficients for each heterogenous consumer $i \in \mathcal{I}$ in the market are drawn according to $(\beta_i, \gamma_i)^{\mathrm{T}} \sim iid\ N((\bar{\beta}, \bar{\gamma})^{\mathrm{T}}, \Sigma)$ and the covariance matrix is given by:

$$\Sigma = \begin{pmatrix} \sigma_\beta^2 & \tau \sigma_\beta \sigma_\gamma \\ \tau \sigma_\beta \sigma_\gamma & \sigma_\gamma^2 \end{pmatrix}$$

with the correlation coefficient $\tau \in (-1, 1)$.[4] Note that the standard notation for the correlation coefficient, ρ, is retained for the pass-through parameter.

Then, if ϵ_{ij} is identically and independently distributed according to the Type I ex-

[2]See, e.g., McManus (2007) and Cohen (2008) for early studies that employed a discrete-choice model to study quantity discounts empirically. As another formulation similar to Alexandrov (2014), Luo, Perringe, and Vuong (2018) consider a continuous menu of nonlinear tariffs, but they restrict their attention to the case of homogenous products provided by a monopolist.

[3]Here, $\delta < 1$ reflects diminishing marginal utility of size, which seems a natural assumption.

[4]If each individual has observable characteristics x_i, then we could consider the characteristic-dependent means, $\bar{\beta} = \bar{\beta}(x_i)$ and $\bar{\gamma} = \bar{\gamma}(x_i)$.

treme distribution, consumer i's conditional choice probabilities are:

$$\Pr(j|i) = \begin{cases} \dfrac{\exp(\alpha - \beta_i p_S)}{1 + \exp(\alpha - \beta_i p_S) + \exp(\alpha - \beta_i p_L \bar{x} + \gamma_i (\bar{x} - 1)^\delta)} & \text{for } j = S \\ \\ \dfrac{\exp(\alpha - \beta_i p_L \bar{x} + \gamma_i (\bar{x} - 1)^\delta)}{1 + \exp(\alpha - \beta_i p_S) + \exp(\alpha + \gamma_i (\bar{x} - 1)^\delta - \beta_i p_L \bar{x})} & \text{for } j = L \\ \\ \dfrac{1}{1 + \exp(\alpha - \beta_i p_S) + \exp(\alpha - \beta_i p_L \bar{x} + \gamma_i (\bar{x} - 1)^\delta)} & \text{for } j = 0. \end{cases}$$

With the market size normalized to one, the market demand for $j = S$ (i.e., the small-size 'market') is given by

$$q_S(p_S, p_L) = \int \underbrace{\left[\frac{\exp(\alpha - \beta_i p_S)}{1 + \exp(\alpha - \beta_i p_S) + \exp(\alpha - \beta_i p_L \bar{x} + \gamma_i (\bar{x} - 1)^\delta)} \right]}_{=\Pr(j=1|i)} dG(i)$$

and the market demand for $j = L$ (i.e., the large-size 'market') is given by

$$q_L(p_S, p_L) = \int \underbrace{\left[\frac{\exp(\alpha - \beta_i p_L \bar{x} + \gamma_i (\bar{x} - 1)^\delta)}{1 + \exp(\alpha - \beta_i p_S) + \exp(\alpha - \beta_i p_L \bar{x} + \gamma_i (\bar{x} - 1)^\delta)} \right]}_{=\Pr(j=\bar{x}|i)} dG(i).$$

Consequently, the 'demand' for the outside option is: $q_0(p_S, p_L) = 1 - q_S(p_S, p_L) - q_L(p_S, p_L)$.

2.2 Retailer's Problem

Given the demand structure described above, the monopolistic retailer's profit function is given by

$$\pi(p_S, p_L; c) = (p_S - c) q_S(p_S, p_L) + (p_L - c)\{\bar{x} \cdot q_L(p_S, p_L)\},$$

where $c \geq 0$ is the marginal cost that is common between markets s and w. Then, the first-order conditions for profit maximization with respect to p_S and p_L are:

$$\begin{cases} \pi_S(p_S, p_L; c) = q_S(p_S, p_L) + (p_S - c)\dfrac{\partial q_S}{\partial p_S}(p_S, p_L) + (p_L - c)\bar{x} \cdot \dfrac{\partial q_L}{\partial p_S}(p_S, p_L) = 0 \\ \\ \pi_L(p_S, p_L; c) = (p_S - c)\dfrac{\partial q_S}{\partial p_L}(p_S, p_L) + \bar{x} \cdot \left\{ q_L(p_S, p_L) + (p_L - c)\dfrac{\partial q_L}{\partial p_L}(p_S, p_L) \right\} = 0, \end{cases} \tag{1}$$

where $\pi_j := \partial \pi / \partial p_j$ for $j = S, L$.

Note that the second-order conditions are:

$$\frac{\partial \pi_S(p_S, p_L; c)}{\partial p_S} < 0, \quad \frac{\partial \pi_L(p_S, p_L; c)}{\partial p_L} < 0,$$

and

$$\frac{\partial \pi_S(p_S, p_L; c)}{\partial p_S} \frac{\partial \pi_L(p_S, p_L; c)}{\partial p_L} - \left(\frac{\partial \pi_S(p_S, p_L; c)}{\partial p_L} \right)^2 > 0,$$

which are assumed to hold in the analysis below.

2.3 Elasticities and Curvatures

Given the demand structure, the *elasiticity matrix* in the case of monopoly is given by

$$\boldsymbol{\varepsilon} \ := \ \begin{pmatrix} \varepsilon_{SS} & \varepsilon_{SL} \\ \varepsilon_{LS} & \varepsilon_{LL} \end{pmatrix},$$

where

$$\begin{cases} \varepsilon_{SS} := -\dfrac{p_S}{q_S}\dfrac{\partial q_S}{\partial p_S} \\ \varepsilon_{LL} := -\dfrac{p_L}{q_L}\dfrac{\partial q_L}{\partial p_L} \end{cases}$$

are the *own* elasticities in the small-size and large-markets, respectively, and

$$\begin{cases} \varepsilon_{SL} := \dfrac{p_L}{q_S}\dfrac{\partial q_S}{\partial p_L} \\ \varepsilon_{LS} := \dfrac{p_S}{q_L}\dfrac{\partial q_L}{\partial p_S} \end{cases}$$

are the *cross* elasticities. Note that these definitions together imply:

$$\frac{\partial q_S}{\partial p_S} = -\frac{q_S \varepsilon_{SS}}{p_S}, \quad \frac{\partial q_L}{\partial p_L} = -\frac{q_L \varepsilon_{LL}}{p_L},$$

$$\frac{\partial q_S}{\partial p_L} = \frac{q_S \varepsilon_{SL}}{p_L}, \quad \frac{\partial q_L}{\partial p_S} = \frac{q_L \varepsilon_{LS}}{p_S}.$$

Similarly, the *curvature tensor* in the case of monopoly is given by

$$\begin{aligned} \boldsymbol{\kappa} \ &:= \ \{\underbrace{\{\kappa^m_{hk}\}_{h,k}}_{:=\kappa^m}\}^m \\ &= \ \{\kappa^m\}^m \end{aligned}$$

for $h, k = S, L$ where, for $m = S, L$,

$$\kappa^m \ := \ \begin{pmatrix} \kappa^m_{SS} & \kappa^m_{SL} \\ \kappa^m_{LS} & \kappa^m_{LL} \end{pmatrix},$$

with

$$\begin{cases} \kappa_{SS}^m = -\dfrac{\partial(\partial q_S/\partial p_S)}{\partial p_m} \cdot \dfrac{p_m}{\partial q_S/\partial p_S} = -\dfrac{p_m}{\partial q_S/\partial p_S} \dfrac{\partial^2 q_S}{\partial p_S \partial p_m} \\ \kappa_{LL}^m = -\dfrac{\partial(\partial q_L/\partial p_L)}{\partial p_m} \cdot \dfrac{p_m}{\partial q_L/\partial p_L} = -\dfrac{p_m}{\partial q_L/\partial p_L} \dfrac{\partial^2 q_L}{\partial p_L \partial p_m} \end{cases}$$

and

$$\begin{cases} \kappa_{SL}^m = -\dfrac{\partial(\partial q_S/\partial p_L)}{\partial p_m} \cdot \dfrac{p_m}{\partial q_S/\partial p_L} = -\dfrac{p_m}{\partial q_S/\partial p_L} \dfrac{\partial^2 q_S}{\partial p_L \partial p_m} \\ \kappa_{LS}^m = -\dfrac{\partial(\partial q_L/\partial p_S)}{\partial p_m} \cdot \dfrac{p_m}{\partial q_L/\partial p_S} = -\dfrac{p_m}{\partial q_L/\partial p_S} \dfrac{\partial^2 q_L}{\partial p_S \partial p_m}. \end{cases}$$

Note that the number of curvature components is the *cube* of the total number of "products" (i.e., sales methods in our case), while that of elasticities is its *square*.

To understand these curvature concepts more deeply, note first that κ_{SS}^m measures how the own-price sensitivity of small-package demand, $\partial q_S/\partial p_S$, changes as p_m varies. Specifically, κ_{SS}^S measures the curvature of small-package demand with respect to its own price, capturing how the consumer's own-price sensitivity to the small package changes as the small-package price itself rises. By contrast, κ_{SS}^L measures how the own-price sensitivity of small-package demand is affected by changes in the *large*-package price, capturing whether an increase in the price of the large package alters the structure of own-price responsiveness in the small-package market — a distinct feature of multi-size or multil-product pricing.

Next, κ_{SL}^m measures how the sensitivity of small-package demand with respect to the price of the large-package, p_L, $\partial q_S/\partial p_L$, changes in response to a change in p_m. Specifically, κ_{SL}^S measures how an increase in the price of the small-size product affects consumers' sensitivity to the price of the large-size product. That is, it captures whether the pattern of substitution between the small and large sizes changes as the price of the small-size product rises. By contrast, κ_{SL}^L how the sensitivity of demand to the price of the large-size product varies with the price of the large-size product itself. In this sense, it represents a form of own-price curvature for the large-size product, as reflected in the demand for the small-size product, q_S.

As in the case of the elastcities, these definitions together imply the following relationships:

$$\frac{\partial^2 q_S}{\partial p_S^2} = \frac{q_S \varepsilon_{SS} \kappa_{SS}^S}{p_S^2}, \quad \frac{\partial^2 q_L}{\partial p_L^2} = \frac{q_L \varepsilon_{LL} \kappa_{LL}^L}{p_L^2},$$

$$\frac{\partial^2 q_S}{\partial p_S \partial p_L} = \frac{q_S \varepsilon_{SS} \kappa_{SS}^L}{p_S p_L}, \quad \frac{\partial^2 q_L}{\partial p_S \partial p_L} = \frac{q_L \varepsilon_{LL} \kappa_{LL}^S}{p_S p_L}$$

$$\frac{\partial^2 q_S}{\partial p_L^2} = -\frac{q_S \varepsilon_{SL} \kappa_{SL}^L}{p_L^2}, \quad \frac{\partial^2 q_L}{\partial p_S^2} = -\frac{q_L \varepsilon_{LS} \kappa_{LS}^S}{p_S^2}.$$

Note here that two of the eight curvature terms, κ_{SL}^S and κ_{LS}^L , do not appear explicitly in the six equations above because

$$\begin{aligned}
\kappa_{SL}^{S} &= -\frac{p_S}{\partial q_S/\partial p_L} \cdot \frac{\partial^2 q_S}{\partial p_L \partial p_S} \\
&= \left(-\frac{p_S}{\partial q_S/\partial p_L}\right)\left(-\frac{\partial q_S/\partial p_S}{p_L}\right)\kappa_{SS}^{L} \\
&= \left(-\frac{p_S p_L}{q_S \varepsilon_{SL}}\right)\left(\frac{q_S \varepsilon_{SS}}{p_S p_L}\right)\kappa_{SS}^{L} \\
&= -\frac{\varepsilon_{SS}}{\varepsilon_{SL}}\kappa_{SS}^{L},
\end{aligned}$$

where the symmetry, $\partial^2 q_S/\partial p_L \partial p_S = \partial^2 q_S/\partial p_S \partial p_L$ is used. Similarly, it is also verified that $\kappa_{LS}^{L} = -(\varepsilon_{LL}/\varepsilon_{LS})\kappa_{LL}^{S}$.

Appendix A illustrates the actual expressions for the first- and second-order derivatives for the random-coefficient logit demand that are used in Section 3 below. For future research, it would be interesting to deeply analyze the relationship between these two measures $\boldsymbol{\varepsilon}$ and $\boldsymbol{\kappa}$ that constitute the *demand manifold* as in the case of product differentiation (Mrázová and Neary 2017; Miravete, Seim, and Thurk 2025).

2.4 Pass-Through

Now, we define *pass-through* in the small-size and large-size markets by $\rho_S := dp_S/dc$ and $\rho_L := dp_L/dc$, respectively.

The analytical formula for pass-through is given by:

$$\begin{aligned}
\begin{pmatrix} \rho_S \\ \rho_L \end{pmatrix} &= -\begin{pmatrix} \frac{\partial \pi_S}{\partial p_S} & \frac{\partial \pi_S}{\partial p_L} \\ \frac{\partial \pi_L}{\partial p_S} & \frac{\partial \pi_L}{\partial p_L} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial \pi_S}{\partial c} \\ \frac{\partial \pi_L}{\partial c} \end{pmatrix} \\
&= \frac{1}{\underbrace{\frac{\partial \pi_S}{\partial p_S}\frac{\partial \pi_L}{\partial p_L} - \left(\frac{\partial \pi_S}{\partial p_L}\right)^2}_{:=\Delta>0\ (\because\ \text{2nd-order condition})}} \begin{pmatrix} \frac{\partial \pi_S}{\partial p_L}\frac{\partial \pi_L}{\partial c} - \frac{\partial \pi_L}{\partial p_L}\frac{\partial \pi_S}{\partial c} \\ \frac{\partial \pi_L}{\partial p_S}\frac{\partial \pi_S}{\partial c} - \frac{\partial \pi_S}{\partial p_S}\frac{\partial \pi_L}{\partial c} \end{pmatrix}
\end{aligned}$$

where

$$\begin{cases}
\frac{\partial \pi_S}{\partial p_S} = 2\frac{\partial q_S}{\partial p_S} + (p_S - c)\frac{\partial^2 q_S}{\partial p_S^2} + (p_L - c)\bar{x} \cdot \frac{\partial^2 q_L}{\partial p_S^2} \\
\frac{\partial \pi_S}{\partial p_L} = \frac{\partial q_S}{\partial p_L} + (p_S - c)\frac{\partial^2 q_S}{\partial p_S \partial p_L} + \bar{x} \cdot \left\{\frac{\partial q_L}{\partial p_S} + (p_L - c)\frac{\partial^2 q_L}{\partial p_S \partial p_L}\right\} = \frac{\partial \pi_L}{\partial p_S} \\
\frac{\partial \pi_L}{\partial p_L} = (p_S - c)\frac{\partial^2 q_S}{\partial p_L^2} + \bar{x} \cdot \left\{2\frac{\partial q_L}{\partial p_L} + (p_L - c)\frac{\partial^2 q_L}{\partial p_L^2}\right\}
\end{cases}$$

and

$$\begin{cases} \dfrac{\partial \pi_S}{\partial c} = -\dfrac{\partial q_S}{\partial p_S} - \bar{x} \cdot \dfrac{\partial q_L}{\partial p_S} \\ \dfrac{\partial \pi_L}{\partial c} = -\dfrac{\partial q_S}{\partial p_L} - \bar{x} \cdot \dfrac{\partial q_L}{\partial p_L}. \end{cases}$$

For convenience, we introduce several definitions that will be useful for characterizing pass-through. First, define the *Lerner indices* for the small and large packages as $\Lambda_S := (p_S - c)/p_S$ and $\Lambda_L := (p_L - c)/p_L$, respectively. Using the first-order conditions (1), these can be expressed as

$$\begin{cases} \Lambda_S = \dfrac{\varepsilon_{LL} + \bar{x}\left(\dfrac{p_L q_L}{p_S q_S}\right)\varepsilon_{LS}}{\varepsilon_{SS}\varepsilon_{LL} - \varepsilon_{LS}\varepsilon_{SL}}, \\ \Lambda_L = \dfrac{\varepsilon_{SS} + \dfrac{1}{\bar{x}}\left(\dfrac{p_S q_S}{p_L q_L}\right)\varepsilon_{SL}}{\varepsilon_{SS}\varepsilon_{LL} - \varepsilon_{LS}\varepsilon_{SL}}. \end{cases}$$

We next define four *weighted demand elasticities*. The weighted own-price elasticities are

$$\begin{cases} \omega_{SS} := q_S \varepsilon_{SS}, \\ \omega_{LL} := \bar{x} q_L \varepsilon_{LL}, \end{cases}$$

while the weighted cross-price elasticities are

$$\begin{cases} \omega_{SL} := q_S \varepsilon_{SL}, \\ \omega_{LS} := \bar{x} q_L \varepsilon_{LS}. \end{cases}$$

These objects have a useful interpretation in terms of quantity responses to percentage price changes. Recall that the physical package sizes are $x_S = 1$ and $x_L = \bar{x}$. Using the definitions of the elasticities above, the weighted own-price elasticities satisfy $\omega_{SS} = -\partial q_S / \partial \ln p_S$ and $\omega_{LL} = -\bar{x} \cdot \partial q_L / \partial \ln p_L$, whereas the weighted cross-price elasticities satisfy $\omega_{SL} = \partial q_S / \partial \ln p_L$ and $\omega_{LS} = \bar{x} \cdot \partial q_L / \partial \ln p_S$. More compactly, since $x_S = 1$ and $x_L = \bar{x}$,

$$\omega_{jj} = -x_j \frac{\partial q_j}{\partial \ln p_j}, \qquad \omega_{jk} = x_j \frac{\partial q_j}{\partial \ln p_k}, \quad j \neq k.$$

Thus, the ω's rescale conventional demand elasticities by market demand and physical package size. They provide a convenient measure of package-size-adjusted demand responses to percentage price changes, with own-price responses expressed in absolute value. This normalization will be particularly useful for expressing the pass-through formulas below.

The following proposition shows how the pass-through is expressed in terms of elasticities $\boldsymbol{\varepsilon}$ and $\boldsymbol{\kappa}$.

Proposition 1. *The pass-through values in markets S and L are given by:*

$$\rho_S = \frac{1}{p_S p_L^2 \cdot \Delta} \left[(p_L \cdot \Xi_L) \cdot \Omega_S + \Psi \cdot \Omega_L \right]$$

and

$$\rho_L = \frac{1}{p_S^2 p_L \cdot \Delta} \left[(p_S \cdot \Xi_S) \cdot \Omega_L + \Psi \cdot \Omega_S \right],$$

respectively, where

$$\Delta := \frac{p_S p_L \cdot (\Xi_S \cdot \Xi_L) - \Psi^2}{p_S^2 p_L^2},$$

$$\begin{cases} \Psi := p_S \omega_{SL} \left(1 - \Lambda_S \kappa_{SL}^S\right) + p_L \omega_{LS} \left(1 - \Lambda_L \kappa_{LS}^L\right) \\ \Omega_S := \omega_{SS} - \omega_{LS} \\ \Omega_L := \omega_{LL} - \omega_{SL}, \end{cases}$$

and

$$\begin{cases} \Xi_S := \omega_{SS} \left(2 - \Lambda_S \kappa_{SS}^S\right) + \left(\dfrac{p_L}{p_S}\right) \omega_{LS} \Lambda_L \kappa_{LS}^S \\ \Xi_L := \omega_{LL} \left(2 - \Lambda_L \kappa_{LL}^L\right) + \left(\dfrac{p_S}{p_L}\right) \omega_{SL} \Lambda_S \kappa_{SL}^L. \end{cases}$$

Proof. See Appendix B. $\square$

Proposition 1 provides a general decomposition of pass-through under quantity discounts in our setting. The terms Ω_S and Ω_L capture the direct effect of a common marginal-cost increase on the two first-order conditions. By contrast, Ξ_S and Ξ_L summarize the curvature of the respective pricing conditions, while Ψ captures the cross-price interaction between the two pricing conditions. Thus, even though the underlying cost shock is common to both package sizes, its transmission to the two prices need not be proportional. The firm can use one price as an instrument for adjusting demand at the other price, so that the response of one package price depends on the curvature and cross-price effects generated by the entire nonlinear price menu.

This observation also clarifies why pass-through under nonlinear pricing differs from the standard multiproduct setting. With linear prices for distinct products, cross-price effects arise because consumers substitute across products. Here, the two package sizes are different sales options for the same underlying good, and the relevant interaction is generated by the sorting of heterogeneous consumers across the menu. A change in the price of one package therefore changes not only its own demand but also the composition of consumers purchasing the other package. The terms involving Ψ in Proposition 1 capture precisely this interaction. In this sense, the cross effect is not merely a conventional product-substitution effect: it reflects the way in which the nonlinear pricing schedule allocates consumers across quantities.

Remark (MNL special case)**.** Under the standard multinomial logit demand ($\sigma_\beta = \sigma_\gamma = 0$), the pass-through rates admit closed-form expressions in terms of market shares (s_S, s_L) and Lerner indices (Λ_S, Λ_L). This provides an analytical verification of the numerical finding that pass-through is systematically larger for the large package.

It is verified that the Lerner indices are:

$$\begin{cases} \Lambda_S = \dfrac{1}{\bar{\beta} \cdot p_S(1 - q_S - q_L)} \\ \Lambda_L = \dfrac{1}{\bar{\beta}\bar{x} \cdot p_L(1 - q_S - q_L)} \end{cases}$$

and the pass-through values are simply given by:

$$\begin{cases} \rho_S = 1 - q_S - \bar{x} \cdot q_L \\ \rho_L = 1 - q_L - \dfrac{q_S}{\bar{x}} \end{cases}$$

using the expressions presented in Table 1 (see Appendix C for the expressions of the first- and second-order derivatives and the elasticities and curvatures for multinomial logit demand) .

It is immediate to see that

$$\rho_L > \rho_S \Leftrightarrow (\bar{x} - 1)q_L > \left(\frac{1}{\bar{x}} - 1\right) q_S,$$

and the latter inequality always holds because $\bar{x} > 1$. It is also verified that

$$\rho_S < 0 \quad \Longleftrightarrow \quad q_S + \bar{x}q_L > 1.$$

Thus, even the benchmark logit specification generates the possibility that the small-package price falls following a common marginal-cost increase. Consumer heterogeneity is not necessary for the basic Edgeworth-type asymmetry. Heterogeneity becomes important for determining how large the asymmetry can be, and, in particular, for generating over-shifting and its distributional implications, as shown in Section 3. □

3 Numerical Analysis

This section investigates the quantitative implications of the model and examines whether the analytical mechanisms identified above remain relevant under consumer heterogeneity. We focus on three questions. First, does the ordering $\rho_L > \rho_S$ derived under the multinomial-logit benchmark remain valid when consumers are heterogeneous in their price sensitivity and preferences for package size? Second, how frequently does nega-

Table 1: Relevant Indices for Multinomial Logit Demands

Weighted Elasticities	
ω_{SS}	$\bar{\beta} p_S q_S (1 - q_S)$
ω_{SL}	$\bar{\beta} \bar{x} p_L q_L q_S$
ω_{LL}	$\bar{\beta} \bar{x}^2 p_L q_L (1 - q_L)$
ω_{LS}	$\bar{\beta} \bar{x} p_S q_S q_L$
Lerner-Curvature Multiplications	
$\Lambda_S \kappa_{SS}^S$	$\dfrac{1 - 2q_S}{1 - q_S - q_L}$
$\Lambda_S \kappa_{SL}^L$	$\dfrac{\bar{x} p_L (1 - 2q_L)}{p_S (1 - q_S - q_L)}$
$\Lambda_L \kappa_{LL}^L$	$\dfrac{1 - 2q_L}{1 - q_S - q_L}$
$\Lambda_L \kappa_{LS}^S$	$\dfrac{p_S (1 - 2q_S)}{\bar{x} p_L (1 - q_S - q_L)}$
$\Lambda_S \kappa_{SL}^S$	$\dfrac{1 - 2q_S}{1 - q_S - q_L}$
$\Lambda_L \kappa_{LS}^L$	$\dfrac{1 - 2q_L}{1 - q_S - q_L}$
Combinations	
Ω_S	$\bar{\beta} p_S q_S (1 - q_S - \bar{x} \cdot q_L)$
Ω_L	$\bar{\beta} \bar{x}^2 p_L q_L \left[1 - q_L - \dfrac{q_S}{\bar{x}}\right]$
Ξ_S	$\bar{\beta} p_S q_S$
Ξ_L	$\bar{\beta} \bar{x}^2 p_L q_L$
Ψ	0
Δ	$\bar{\beta}^2 \bar{x}^2 q_S q_L$

tive pass-through for the small package, $\rho_S < 0$, arise? Third, how do the package-size difference and the correlation between price sensitivity and size preference affect these pass-through patterns?

Recall that we normalize the size of the small package to one, $x_S = 1$, and denote the size of the large package by $x_L = \bar{x} > 1$. In the baseline specification below, we set $(\bar{x}, \delta) = (2, 0.5)$. The parameter $\bar{x} = 2$ captures a package size ratio of 2:1 between the large and small sizes, and the parameter $\delta = 0.5$ governs the intensity of the utility gain from large package sizes. As robustness checks, we also consider other possiblities for $(\bar{x}, \delta)$ in Subsection 3.3 below.

3.1 Parameterization and Numerical Procedure

We consider the following parameter grids:

$$\alpha = 1.0, 1.5, \ldots, 5.0,$$

$$\bar{\beta} = 1.0, 1.2, \ldots, 3.0, \qquad \bar{\gamma} = 0.0, 0.1, \ldots, 1.0,$$

and

$$\sigma_\beta, \sigma_\gamma = 0, 0.05, \ldots, 0.50.$$

For the correlation between the two random coefficients, we consider $\tau \in \{-0.5, 0, 0.5\}$. The resulting baseline grid contains 395,307 ($= 9 \times 11 \times 11 \times 11 \times 11 \times 3$) parameter configurations.

The grids for α, $\bar{\beta}$, and $\bar{\gamma}$ are chosen to cover a wide range of market configurations, from low-demand, low-price-sensitivity settings to high-demand, high-price-sensitivity settings, ensuring that our numerical results are not driven by a narrow region of the parameter space. For each configuration, we numerically solve the retailer's profit-maximization problem and retain equilibria satisfying the second-order condition and non-negative profits. All 395,307 baseline configurations satisfy these validity checks.[5]

3.2 Pass-Through Ranking

Our first result is a remarkably robust ordering of pass-through across package sizes. Table 2 summarizes the baseline results separately by the correlation parameter τ. The ordering $\rho_L > \rho_S$ holds for every configuration at each value of τ. As explained in the robustness exercises reported below, this ordering is kept for substantial changes in the package-size ratio, $\bar{x}$, and the intensity for quantity preference, δ. Thus, although

[5] We use common random numbers across parameter configurations. For cases with heterogeneity in both β_i and γ_i, the underlying simulation draws are orthogonalized and standardized so that the realized correlation prior to the lower bound on β_i coincides with the target value of τ in finite samples. The lower bound on β_i binds only rarely in our simulations.

Table 2: Baseline Numerical Results by Correlation

τ	$\Pr(\rho_L > \rho_S)$	$\Pr(\rho_S < 0)$
0	1.000	0.502
0.5	1.000	0.546
-0.5	1.000	0.450

Notes: The baseline specification sets $\bar{x} = 2$ and $\delta = 0.5$. Each value of τ contains 131,769 parameter configurations.

we do not establish a general analytical theorem for the random-coefficient model, the ordering derived analytically under multinomial logit demand appears to be a highly robust numerical regularity.

Although pass-through is always larger for the large package in our numerical experiments, pass-through to the small package can be negative. Under the baseline parameterization, negative small-package pass-through occurs in approximately one-half of the parameter configurations considered.

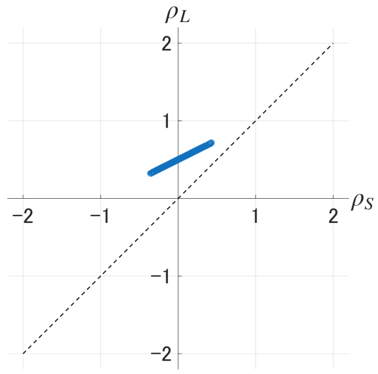
This result is the nonlinear-pricing counterpart of the Edgeworth–Salinger mechanism. A common increase in per-unit marginal cost raises the total production cost of the large package by $\bar{x}\, dc$, whereas the corresponding increase for the small package is only dc. At the same time, the two package prices jointly determine how consumers sort between the small package, the large package, and the outside option. The retailer therefore optimally adjusts the entire price schedule rather than passing the cost increase independently into each price. When the induced substitution toward the small package is sufficiently strong, the optimal response can involve a reduction in p_S even though marginal cost has increased.

The sufficient-statistics formula in Proposition 1 makes clear that the sign of ρ_S cannot in general be determined by the sign of the cross-price interaction term Ψ alone. In particular,

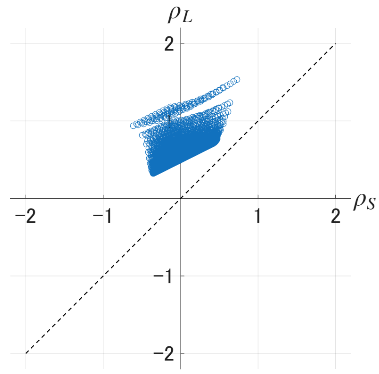
$$\rho_S = \frac{\Psi \Omega_L + (p_L \Xi_L)\Omega_S}{p_S p_L^2 \Delta}.$$

Hence negative pass-through depends on the joint contribution of the cross-price interaction and the own-price component, $\Psi\Omega_L + (p_L\Xi_L)\Omega_S$, rather than on $\Psi < 0$ as a necessary condition.

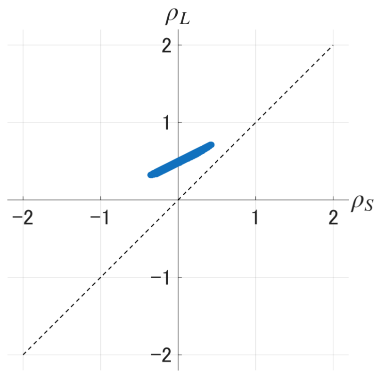
Figure 1 illustrates the geometry behind this result: how the plot region for (ρ_S, ρ_L) changes across different demand specifications. Under multinomial logit demand (panel (a)), the possible values for (ρ_S, ρ_L) are quite limited. Introducing heterogeneity in the price coefficient β_i (panel (b)) substantially expands the range of feasible pass-through values and can generate *over-shifting* for the large package (i.e., $\rho_L > 1$). However, intro-



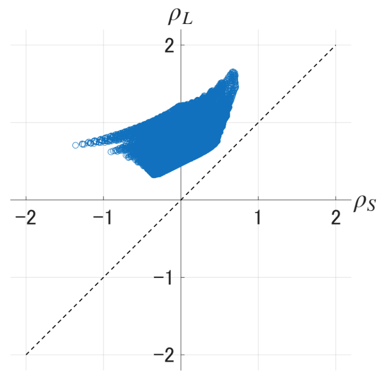
(a) Multinomial Logit Demand
($\sigma_\beta = 0$ and $\sigma_\gamma = 0$)



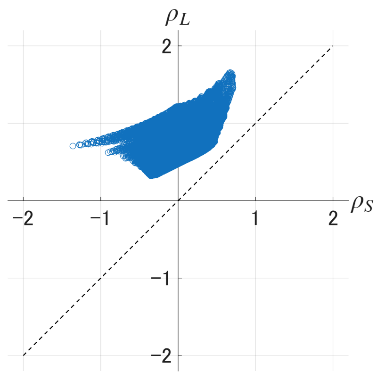
(b) Heterogeneity in β_i Only
($\sigma_\beta > 0$ and $\sigma_\gamma = 0$)



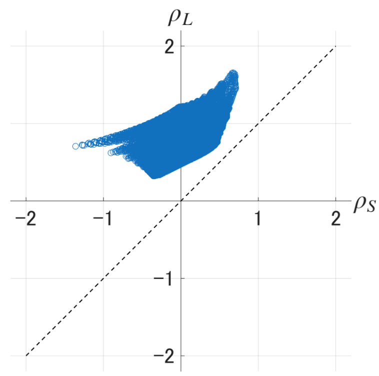
(c) Heterogeneity in γ_i Only
($\sigma_\beta = 0$ and $\sigma_\gamma > 0$)



(d) Heterogeneity in Both β_i and γ_i
($\sigma_\beta > 0$ and $\sigma_\gamma > 0$ but with $\tau = 0$)



(e) Positive Correlation $\tau = 0.5$



(f) Negative Correlation $\tau = -0.5$

Figure 1: Pass-Through under Alternative Demand Specifications.

ducing consumer heterogeneity in the size-preference coefficient γ_i alone (panel (c)) has little impact on the possible values of (ρ_S, ρ_L). Lastly, including consumer heterogeneity in both coefficients β_i and γ_i simultaneously (panel (d)) expands the feasible region relative to panel (b), although neither a positive ($\tau = 0.5$) nor negative ($\tau = -0.5$) correlation alters it much.

Table 3 shows that, regardless of the demand specification, the price offered for the large volume is approximately half that for the small volume, implying that total payments are similar across the two volumes. The market share for the small volume is approximately 27 percent and that for the large volume is approximately 36 percent, both stable across specifications. Intestingly, while the Lerner index for both packages are quite similar, the pass-through rates are not. It is observed that allowing correlation between β_i and γ_i affects the curvature values, κ^S_{SS} and κ^L_{LL}.

3.3 Robustness: Differences in the Package-Size Ratio and the Intensity for Large Consumption

For robustness checks, we repeat the analysis using a moderate grid of 29,403 configurations for each of four alternative combinations of $\bar{x}$ and δ:

$$(\bar{x}, \delta) \in \{(1.5, 0.3), (1.5, 0.7), (2.5, 0.3), (2.5, 0.7)\}.$$

Table 4 reports the share of configurations with negative small-package pass-through.

The difference in package sizes has a first-order quantitative effect. When $\bar{x} = 1.5$, negative small-package pass-through occurs in only about 1–5 percent of the parameter configurations. When $\bar{x} = 2.5$, the corresponding frequency rises to approximately 68–77 percent. By comparison, changing δ from 0.3 to 0.7 produces much smaller effects. Thus, the quantity difference between the two packages appears to be an important determinant of the Edgeworth-type response.

This pattern is economically intuitive. A larger value of $\bar{x}$ increases the difference between the effective cost shocks to the two packages. It also makes adjustment of the relative package prices more important for consumer sorting. Both forces strengthen the incentive to absorb, or even reverse, part of the cost increase in the small-package price while passing a larger fraction through to the large package.

3.4 Consumer Heterogeneity and Correlation

Consumer heterogeneity affects not only the range of pass-through values but also the incidence of negative pass-through. In particular, the correlation $\tau = \mathrm{Corr}(\beta_i, \gamma_i)$ generates a systematic pattern.

As shown in Table 2, increasing τ from -0.5 to 0.5 raises the fraction of baseline

Table 3: Mean Values

(1) Prices and Shares

	p_S	p_L	q_S	q_L
(a) Multinomial Logit	1.823 (0.837)	0.962 (0.419)	0.274 (0.063)	0.370 (0.083)
Random-Coefficient Logit				
(b) Randomness Only for β_i	1.824 (0.824)	0.968 (0.417)	0.270 (0.063)	0.361 (0.081)
(c) Randomness Only for γ_i	1.824 (0.835)	0.969 (0.421)	0.273 (0.060)	0.371 (0.078)
(d) Randomness Both for β_i and γ_i with $\tau = 0.0$	1.824 (0.823)	0.974 (0.419)	0.269 (0.060)	0.362 (0.076)
(e) Allowing the Correlation $\tau = 0.5$ between β_i and γ_i	1.839 (0.839)	0.966 (0.413)	0.266 (0.060)	0.369 (0.079)
(f) Allowing the Correlation $\tau = -0.5$ between β_i and γ_i	1.809 (0.810)	0.983 (0.425)	0.273 (0.061)	0.357 (0.074)

(2) Elasticities and Curvatures

	ε_{SS}	ε_{LL}	κ_{SS}^S	κ_{LL}^L
(a) Multinomial Logit	2.327 (0.536)	2.119 (0.376)	1.407 (0.392)	0.791 (0.450)
Random-Coefficient Logit				
(b) Randomness Only for β_i	2.265 (0.505)	2.077 (0.368)	1.499 (0.397)	0.981 (0.430)
(c) Randomness Only for γ_i	2.321 (0.515)	2.101 (0.367)	1.400 (0.345)	0.772 (0.406)
(d) Randomness Both for β_i and γ_i with $\tau = 0.0$	2.260 (0.486)	2.061 (0.359)	1.495 (0.362)	0.949 (0.386)
(e) Allowing the Correlation $\tau = 0.5$ between β_i and γ_i	2.270 (0.495)	2.055 (0.356)	1.536 (0.395)	0.886 (0.404)
(f) Allowing the Correlation $\tau = -0.5$ between β_i and γ_i	2.251 (0.478)	2.066 (0.363)	1.454 (0.335)	1.009 (0.391)

(3) Lerner and Pass-Through

	Λ_S	Λ_L	ρ_S	ρ_L
(a) Multinomial Logit	0.934 (0.028)	0.877 (0.048)	−0.013 (0.178)	0.493 (0.089)
Random-Coefficient Logit				
(b) Randomness Only for β_i	0.934 (0.028)	0.878 (0.048)	0.005 (0.192)	0.561 (0.131)
(c) Randomness Only for γ_i	0.934 (0.028)	0.878 (0.048)	−0.005 (0.171)	0.491 (0.086)
(d) Randomness Both for β_i and γ_i with $\tau = 0.0$	0.934 (0.028)	0.879 (0.047)	0.015 (0.184)	0.552 (0.120)
(e) Allowing the Correlation $\tau = 0.5$ between β_i and γ_i	0.935 (0.028)	0.878 (0.048)	−0.009 (0.195)	0.540 (0.115)
(f) Allowing the Correlation $\tau = -0.5$ between β_i and γ_i	0.934 (0.028)	0.880 (0.047)	0.035 (0.181)	0.566 (0.132)

Notes: The number of parameter configurations is: 1089 for (a), 11,979 for (b) and (c), and 131,769 for (d), (e), and (f). Standard deviations are in parentheses.

Table 4: Negative Small-Package Pass-Through: Robustness to Package Size and Curvature

		$\Pr(\rho_S < 0)$		
$\bar{x}$	δ	$\tau = -0.5$	$\tau = 0$	$\tau = 0.5$
1.5	0.3	0.028	0.031	0.051
1.5	0.7	0.012	0.013	0.022
2.5	0.3	0.677	0.718	0.751
2.5	0.7	0.690	0.737	0.773

*Notes:*Each $(\bar{x}, \delta, \tau)$ cell contains 9,801 parameter configurations. In all cells, $\Pr(\rho_L > \rho_S) = 1$.

configurations with $\rho_S < 0$ from 0.450 to 0.546. At the same time, Table 3 shows that average pass-through falls for both packages, but it falls more strongly for the small package, so that the average pass-through gap increases: $\bar{\rho}_L - \bar{\rho}_S = 0.531, 0.537, 0.549$ for $\tau = -0.5, 0, 0.5$, respectively.

The same pattern appears in every robustness specification reported in Table 4. For each combination of $(\bar{x}, \delta)$,

$$\Pr(\rho_S < 0 \mid \tau = -0.5) < \Pr(\rho_S < 0 \mid \tau = 0) < \Pr(\rho_S < 0 \mid \tau = 0.5).$$

Thus, a more positive correlation between price sensitivity and preferences for large package sizes systematically increases the likelihood of negative pass-through to the small package.

The correlation parameter also shifts the cross-price interaction term Ψ. In the baseline grid, its averege is $\bar{\Psi} = 0.0159,\ 0.0075,\ -0.0015$ for $\tau = -0.5, 0, 0.5$, respectively. The same monotonic decline appears in all four robustness exercises. Hence Ψ is generally small relative to the overall scale of the pass-through expressions, but it is not simply a zero-centered residual term. Rather, consumer sorting systematically shifts the interaction between the two pricing first-order conditions.

These findings also have a potential distributional interpretation. When $\tau > 0$, consumers with stronger preferences for large packages are also more price sensitive. Since the large package systematically exhibits higher pass-through, this sorting pattern may generate heterogeneous incidence of cost shocks across consumers. If higher price sensitivity is associated with lower household income, the mechanism could contribute to *regressive* incidence. Conversely, when the correlation is negative ($\tau < 0$)—that is, when consumers who prefer large packages are less price-elastic, as might be the case if bulk purchasing is associated with higher income and lower price sensitivity—the distributional consequences are reversed. In this case, cost shocks are *progressive*, and the welfare burden falls disproportionately on higher-income households.

The sign of τ is therefore an empirically important parameter for distributional analysis. We emphasize, however, that income is not explicitly modeled here, and a formal assessment of regressivity requires both an explicit welfare calculation and empirical information on the joint distribution of price sensitivity, size preferences, and household income.

4 Extension: Non-constant Marginal Costs

So far, we have assumed a constant per-unit marginal cost, $c \geq 0$. This assumption allows us to isolate the demand-side mechanism behind asymmetric pass-through across package sizes. In many applications, however, the marginal cost of supplying the product may depend on total physical sales. This section shows that the framework extends naturally to such environments and clarifies how production-cost curvature affects the pass-through results.

Let total physical quantity sold be

$$Q(p_S, p_L) = q_S(p_S, p_L) + \bar{x}\, q_L(p_S, p_L),$$

where q_S and q_L denote the market shares of the small and large packages, respectively. Let total production cost be $C(Q; t)$, where $t \in \mathbb{R}$ is an exogenous supply-side cost shifter, and define marginal cost as

$$mc(Q; t) := \frac{\partial C(Q; t)}{\partial Q}.$$

The constant-marginal-cost model studied above corresponds to the special case: $C(Q; c) = c \cdot Q$.

The seller's profit is now:

$$\pi(p_S, p_L; t) = p_S q_S(p_S, p_L) + p_L \bar{x}\, s_L(p_S, p_L) - C(Q(p_S, p_L); t).$$

Define revenue by

$$R(p_S, p_L) = p_S q_S(p_S, p_L) + \bar{x} p_L q_L(p_S, p_L).$$

The first-order condition with respect to price p_j, $j \in \{S, L\}$, can then be written compactly as

$$F_j := \frac{\partial \pi}{\partial p_j} = \frac{\partial R}{\partial p_j} - mc(Q; t) \frac{\partial Q}{\partial p_j} = 0.$$

Equivalently, the two first-order conditions are:

$$\begin{cases} \dfrac{\partial \pi}{\partial p_S} = q_S + (p_S - mc)q_{S,S} + \bar{x} \cdot (p_L - mc)q_{L,S} = 0 \\ \\ \dfrac{\partial \pi}{\partial p_L} = (p_S - mc)q_{S,L} + \bar{x} \cdot [q_L + (p_L - mc)q_{L,L}] = 0, \end{cases}$$

where $q_{j,k} := \partial q_j / \partial p_k$. Thus, the first-order conditions take exactly the same form as under constant marginal cost, except that the constant c is replaced by the endogenous marginal cost $mc(Q;t)$.

4.1 Pass-Through with Cost Curvature

Let

$$\mathbf{p} = \begin{pmatrix} p_S \\ p_L \end{pmatrix}, \qquad Q_{\mathbf{p}} = \begin{pmatrix} Q_S \\ Q_L \end{pmatrix},$$

where $Q_j := \partial Q / \partial p_j$. Differentiating the pricing first-order conditions with respect to prices gives

$$H := \frac{\partial \mathbf{F}}{\partial \mathbf{p}'} = R_{\mathbf{p}p} - mc \cdot Q_{\mathbf{p}p} - mc_Q Q_{\mathbf{p}} Q'_{\mathbf{p}},$$

where

$$mc_Q := \frac{\partial mc(Q;t)}{\partial Q}.$$

Relative to the constant-marginal-cost case, non-constant marginal cost therefore adds the rank-one term $-mc_Q Q_{\mathbf{p}} Q'_{\mathbf{p}}$ to the Hessian of profit.

We focus on an additive marginal-cost shock, $mc(Q;\theta) = mc_0(Q) + t$, so that

$$mc_t := \frac{\partial mc(Q;t)}{\partial t} = 1.$$

Totally differentiating the first-order conditions with respect to t yields

$$H \frac{d\mathbf{p}}{dt} - Q_{\mathbf{p}} = 0,$$

implying the the vector of pass-through is

$$\boldsymbol{\rho} := \frac{d\mathbf{p}}{dt} = H^{-1} Q_{\mathbf{p}}.$$

It is useful to express this result in terms of positive-sign objects. Define

$$A_0 := -\left(R_{\mathbf{p}p} - mc \cdot Q_{\mathbf{p}p}\right), \qquad \mathbf{g} := -Q_{\mathbf{p}}.$$

Then

$$-H = A_0 + mc_Q \mathbf{g}\mathbf{g}',$$

and therefore

$$\boldsymbol{\rho} = \left(A_0 + mc_Q \mathbf{g}\mathbf{g}'\right)^{-1} \mathbf{g}.$$

The matrix A_0 is precisely the local pricing matrix that would arise if marginal cost were held fixed at its equilibrium value mc. The additional term $m_Q \mathbf{g}\mathbf{g}'$ captures the effect of production-cost curvature.

Proposition 2 (Pass-through under non-constant marginal cost)**.** *Suppose that the equilibrium is interior, A_0 is nonsingular, and $1 + mc_Q \mathbf{g}' A_0^{-1} \mathbf{g} > 0$. Under the additive marginal-cost shock $mc(Q;t) = mc_0(Q) + t$, pass-through is given by*

$$\boldsymbol{\rho} = \lambda\, \boldsymbol{\rho}^0,$$

where $\boldsymbol{\rho}^0 := A_0^{-1}\mathbf{g}$ is the pass-through vector associated with locally constant marginal cost, and

$$\lambda = \frac{1}{1 + mc_Q \mathbf{g}' \boldsymbol{\rho}^0}.$$

Remark. The result follows immediately from the Sherman–Morrison formula. It shows that production-cost curvature enters pass-through through a particularly simple channel. Conditional on the equilibrium prices and the level of marginal cost, the curvature term rescales the two package-specific pass-through coefficients by the same factor λ. □

If A_0 is positive definite, then $\mathbf{g}' A_0^{-1} \mathbf{g} > 0$. Consequently, increasing marginal cost, $mc_Q > 0$, implies $0 < \lambda < 1$, so that production-cost curvature attenuates both pass-through coefficients relative to the corresponding locally constant-marginal-cost benchmark. Conversely, decreasing marginal cost, $mc_Q < 0$, amplifies pass-through, provided that the regularity condition in Proposition 2 continues to hold.

Importantly, because the curvature correction multiplies both components of the pass-through vector by the same positive factor, it does not directly alter their ordering:

$$\rho_L - \rho_S = \lambda \left(\rho_L^0 - \rho_S^0\right).$$

Thus, conditional on the equilibrium point, any strict ranking between small- and large-package pass-through in the corresponding locally linear cost problem is preserved under non-constant marginal cost.

4.2 The Multinomial Logit Benchmark

The result is particularly transparent under multinomial logit demand. Recall that under locally constant marginal cost,

$$\begin{cases} \rho_S^0 = 1 - q_S - \bar{x} q_L \\ \rho_L^0 = 1 - q_L - \dfrac{q_S}{\bar{x}}. \end{cases}$$

Proposition 2 therefore implies

$$\begin{cases} \rho_S = \lambda \left(1 - q_S - \bar{x}\, q_L\right) \\ \rho_L = \lambda \left(1 - q_L - \dfrac{q_S}{\bar{x}}\right). \end{cases}$$

Their difference is

$$\rho_L - \rho_S = \lambda(\bar{x} - 1)\left(q_L + \frac{q_S}{\bar{x}}\right).$$

Hence, whenever $\lambda > 0$ and $\bar{x} > 1$,

$$\rho_L > \rho_S.$$

The analytical ranking obtained under constant marginal cost therefore continues to hold under general production-cost curvature for an additive marginal-cost shock.

Likewise, since $\lambda > 0$,

$$\rho_S < 0 \quad \Longleftrightarrow \quad q_S + \bar{x} q_L > 1.$$

Thus, non-constant marginal cost changes the magnitude of pass-through but does not directly change the condition determining whether small-package pass-through is negative.

These results help clarify the role of the constant-marginal-cost assumption in our baseline analysis. Constant marginal cost is not essential for the qualitative nature for our Edgeworth-Salinger paradox. Rather, it removes the supply-curvature term and thereby isolates the role of nonlinear pricing and demand interdependence. With non-constant marginal cost, the same demand-side mechanism remains operative, while production-cost curvature scales the strength of the price response.

4.3 Discussion

The extension also highlights a useful distinction between demand-side and supply-side curvature. In the baseline model, pass-through is governed by demand elasticities, demand curvature, and the interaction between the two pricing first-order conditions. With non-constant marginal cost, an additional sufficient statistic, mc_Q, captures the curvature of the production technology.

When marginal cost is increasing, the endogenous increase in production cost associated with an expansion in quantity dampens the response of the nonlinear price schedule to an exogenous cost shift. When marginal cost is decreasing, the reverse feedback amplifies the response. Nevertheless, for the additive cost shocks considered here, this supply-side feedback operates through a common multiplicative adjustment to the package-specific pass-through. The central asymmetry emphasized in the paper—larger pass-through for the large package and the possibility of negative pass-through for the small package—therefore does not rely on constant marginal cost.

More generally, if a supply shock changes not only the level but also the slope of marginal cost, so that $mc_{Qt} := \partial^2 mc/\partial Q \partial t \neq 0$, or if production costs depend separately on the quantities of the two package sizes, the common rescaling result need no longer hold. In such environments, supply-side curvature can itself generate additional asymmetry across package sizes. We leave these richer production technologies for future work.

5 Concluding Remarks

This paper studies cost pass-through when a monopolist sells the same underlying product in multiple package sizes and offers quantity discounts. In contrast to the standard setting of each product carrying a single unit price, nonlinear pricing implies that a common per-unit cost shock generates different effective cost changes across package sizes and induces the firm to adjust the entire price schedule.

We find a strong asymmetry in pass-through across package sizes. Under multinomial logit demand, pass-through to the large package is analytically greater than pass-through to the small package. Our numerical analysis shows that this ordering is remarkably robust to consumer heterogeneity. It holds in every one of the 395,307 parameter configurations in our baseline random-coefficient specification and in all 117,612 additional configurations considered in our robustness exercises. At the same time, pass-through to the small package can be negative. In the baseline grid, this occurs in approximately one-half of the parameter configurations.

Interestingly, the prevalence of negative pass-through is strongly related to the difference between package sizes. When the large package contains only 1.5 times as much as the small package, negative pass-through for the small package is relatively rare. When the size ratio rises to 2.5, however, negative pass-through occurs in roughly 70 percent or more of the configurations considered. This pattern highlights a mechanism specific to nonlinear pricing: the larger the difference in cost exposure across package sizes, the stronger the firm's incentive to adjust relative package prices in response to consumer substitution and sorting.

Consumer heterogeneity further shapes this mechanism. A more positive correlation between price sensitivity and preferences for larger package sizes systematically increases the frequency of negative pass-through for the small package and widens the average pass-through gap between the two package sizes. The same correlation shifts the cross-price interaction term Ψ downward. These results indicate that the joint distribution of price sensitivity and package-size preferences matters for the incidence of cost shocks, even when the qualitative ordering $\rho_L > \rho_S$ remains unchanged.

These findings provide a nonlinear-pricing counterpart to the Edgeworth–Salinger paradox. A firm's price response to a cost shock need not be positive for every package of the affected product: an increase in a common per-unit marginal cost can induce the firm to lower the per-unit price of a smaller package while raising that of a larger one. Analyses based on a single unit price may therefore miss an important margin of price adjustment when firms use quantity discounts.

The model also suggests a potential distributional channel. If consumers with stronger preferences for larger packages differ systematically in their price sensitivity, heterogeneous pass-through across package sizes can translate a common cost shock into heterogeneous exposure across consumers. A formal welfare assessment, however, requires richer information than the present model provides, particularly on the relationship between income, price sensitivity, and package-size preferences. Estimating these relationships using household-level scanner data is therefore a natural empirical extension of the analysis.

Finally, the present paper considers a monopolist offering two fixed package sizes. Extending the analysis to oligopolistic menu competition is an important direction for future research. With multiple firms and multiple package sizes, the dimensionality of the pricing problem expands rapidly. Recent work by Koh (2026) develops tractable tools for characterizing pass-through in high-dimensional product spaces, while Iaria and Wang (2026) provide an empirical framework for studying quantity discounts with many products. Combining these approaches with the mechanism studied here may make it possible to analyze cost pass-through, package-size competition, and consumer sorting in differentiated-product oligopoly.

Appendix

Appendix A: First- and Second-Order Derivatives for Random-Coefficient Logit Demand

1st-order	
$\frac{\partial q_S}{\partial p_S}$	$-\int_{i\in\mathcal{I}}\beta_i \mathrm{Pr}(j=1\vert i)[1-\mathrm{Pr}(j=1\vert i)]dG(i)$
$\frac{\partial q_S}{\partial p_L}$	$\bar{x}\int_{i\in\mathcal{I}}\beta_i \mathrm{Pr}(j=1\vert i)\mathrm{Pr}(j=\bar{x}\vert i)dG(i)$
$\frac{\partial q_L}{\partial p_L}$	$-\bar{x}\int_{i\in\mathcal{I}}\beta_i \mathrm{Pr}(j=\bar{x}\vert i)[1-\mathrm{Pr}(j=\bar{x}\vert i)]dG(i)$
$\frac{\partial q_L}{\partial p_S}$	$\int_{i\in\mathcal{I}}\beta_i \mathrm{Pr}(j=1\vert i)\mathrm{Pr}(j=\bar{x}\vert i)dG(i)$
2nd-order	
$\frac{\partial^2 q_S}{\partial p_S^2}$	$\int_{i\in\mathcal{I}}\beta_i^2 \mathrm{Pr}(j=1\vert i)[1-\mathrm{Pr}(j=1\vert i)][1-2\mathrm{Pr}(j=1\vert i)]dG(i)$
$\frac{\partial^2 q_S}{\partial p_L^2}$	$-\bar{x}^2\int_{i\in\mathcal{I}}\beta_i^2 \mathrm{Pr}(j=1\vert i)\mathrm{Pr}(j=\bar{x}\vert i)[1-2\mathrm{Pr}(j=\bar{x}\vert i)]dG(i)$
$\frac{\partial^2 q_L}{\partial p_L^2}$	$\bar{x}^2\int_{i\in\mathcal{I}}\beta_i^2 \mathrm{Pr}(j=\bar{x}\vert i)[1-\mathrm{Pr}(j=\bar{x}\vert i)][1-2\mathrm{Pr}(j=\bar{x}\vert i)]dG(i)$
$\frac{\partial^2 q_L}{\partial p_S^2}$	$-\int_{i\in\mathcal{I}}\beta_i^2 \mathrm{Pr}(j=1\vert i)\mathrm{Pr}(j=\bar{x}\vert i)[1-2\mathrm{Pr}(j=1\vert i)]dG(i)$
$\frac{\partial^2 q_S}{\partial p_S \partial p_L}$	$-\bar{x}\int_{i\in\mathcal{I}}\beta_i^2 \mathrm{Pr}(j=1\vert i)\mathrm{Pr}(j=\bar{x}\vert i)[1-2\mathrm{Pr}(j=1\vert i)]dG(i)$
$\frac{\partial^2 q_L}{\partial p_S \partial p_L}$	$-\bar{x}\int_{i\in\mathcal{I}}\beta_i^2 \mathrm{Pr}(j=1\vert i)\mathrm{Pr}(j=\bar{x}\vert i)[1-2\mathrm{Pr}(j=\bar{x}\vert i)]dG(i)$

Appendix B: Proof of Proposition 1

First, note that

$$\rho_S = \frac{1}{\Delta}\left(\frac{\partial \pi_S}{\partial p_L}\frac{\partial \pi_L}{\partial c} - \frac{\partial \pi_L}{\partial p_L}\frac{\partial \pi_S}{\partial c}\right) = \frac{1}{\Delta}\left(\frac{\partial \pi_S}{\partial p_L}\frac{\Omega_L}{p_L} - \frac{\partial \pi_L}{\partial p_L}\frac{\Omega_S}{p_S}\right)$$

and

$$\rho_L = \frac{1}{\Delta}\left(\frac{\partial \pi_L}{\partial p_S}\frac{\partial \pi_S}{\partial c} - \frac{\partial \pi_S}{\partial p_S}\frac{\partial \pi_L}{\partial c}\right) = \frac{1}{\Delta}\left(\frac{\partial \pi_L}{\partial p_S}\frac{\Omega_S}{p_S} - \frac{\partial \pi_S}{\partial p_S}\frac{\Omega_L}{p_L}\right),$$

because

$$\begin{aligned}
\frac{\partial \pi_S}{\partial c} &= -\frac{\partial q_S}{\partial p_S} - \bar{x}\cdot\frac{\partial q_L}{\partial p_S} = \frac{1}{p_S}(\underbrace{q_S\varepsilon_{SS} - \bar{x}q_L\varepsilon_{LS}}_{:=\Omega_S}) = \frac{\Omega_S}{p_S} \\
\frac{\partial \pi_L}{\partial c} &= -\frac{\partial q_S}{\partial p_L} - \bar{x}\cdot\frac{\partial q_L}{\partial p_L} = \frac{1}{p_L}(\underbrace{-q_S\varepsilon_{SL} + \bar{x}q_L\varepsilon_{LL}}_{:=\Omega_L}) = \frac{\Omega_L}{p_L}
\end{aligned}$$

are used from the definitions of $\boldsymbol{\varepsilon}$. In addition, we also use the definitions of $\boldsymbol{\kappa}$ to obtain:

$$\begin{aligned}
\frac{\partial \pi_S}{\partial p_S} &= 2\frac{\partial q_S}{\partial p_S} + (p_S - c)\frac{\partial^2 q_S}{\partial p_S^2} + (p_L - c)\bar{x}\cdot\frac{\partial^2 q_L}{\partial p_S^2} \\
&= -2\frac{q_S\varepsilon_{SS}}{p_s} + (p_S - c)\frac{q_S\varepsilon_{SS}\kappa_{SS}^S}{p_S^2} + (p_L - c)\bar{x}\left(-\frac{q_L\varepsilon_{LS}\kappa_{LS}^S}{p_S^2}\right) \\
&= \frac{-1}{p_S}\underbrace{\left[(q_S\varepsilon_{SS})\left(2 - \Lambda_S\kappa_{SS}^S\right) + \frac{p_L(\bar{x}q_L\varepsilon_{LS})}{p_S}\Lambda_L\kappa_{LS}^S\right]}_{:=\Xi_S} \\
&= -\frac{\Xi_S}{p_S},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \pi_S}{\partial p_L} &= \frac{\partial q_S}{\partial p_L} + (p_S - c)\frac{\partial^2 q_S}{\partial p_S \partial p_L} + \bar{x}\cdot\left\{\frac{\partial q_L}{\partial p_S} + (p_L - c)\frac{\partial^2 q_L}{\partial p_S \partial p_L}\right\} \\
&= \frac{q_S\varepsilon_{SL}}{p_L} + \Lambda_S\frac{q_S\varepsilon_{SS}\kappa_{SS}^L}{p_L} + \frac{\bar{x}q_L\varepsilon_{LS}}{p_S} + \Lambda_L\frac{\bar{x}q_L\varepsilon_{LL}\kappa_{LL}^S}{p_S} \\
&= \frac{1}{p_S p_L}[p_S(q_S\varepsilon_{SL}) + (p_S q_S)(\underbrace{\varepsilon_{SS}\kappa_{SS}^L}_{=-\varepsilon_{SL}\kappa_{SL}^S})\Lambda_S + p_L(\bar{x}q_L\varepsilon_{LS}) + (p_L\bar{x}q_L)(\underbrace{\varepsilon_{LL}\kappa_{LL}^S}_{=-\varepsilon_{LS}\kappa_{LS}^L})\Lambda_L] \\
&= \frac{1}{p_S p_L}\underbrace{\left[p_S\omega_{SL}\left(1 - \Lambda_S\kappa_{SL}^S\right) + p_L\omega_{LS}\left(1 - \Lambda_L\kappa_{LS}^L\right)\right]}_{=\Psi} \\
&= \frac{\Psi}{p_S p_L} = \frac{\partial \pi_L}{\partial p_S},
\end{aligned}$$

where $(\varepsilon_{SS}/\varepsilon_{SL})\kappa^L_{SS} = -\kappa^S_{SL}$ and $(\varepsilon_{LL}/\varepsilon_{LS})\kappa^S_{LL} = -\kappa^L_{LS}$ are used, and

$$\begin{aligned}
\frac{\partial \pi_L}{\partial p_L} &= (p_S - c)\frac{\partial^2 q_S}{\partial p_L^2} + \bar{x} \cdot \left\{ 2\frac{\partial q_L}{\partial p_L} + (p_L - c)\frac{\partial^2 q_L}{\partial p_L^2} \right\} \\
&= -2\frac{\bar{x} q_L \varepsilon_{LL}}{p_s} + (p_L - c)\frac{\bar{x} q_L \varepsilon_{LL} \kappa^L_{LL}}{p_L^2} + (p_S - c)\left(-\frac{q_S \varepsilon_{SL} \kappa^L_{SL}}{p_L^2} \right) \\
&= \frac{-1}{p_L} \underbrace{\left[(\bar{x} q_L \varepsilon_{LL}) \left(2 - \Lambda_L \kappa^L_{LL} \right) + \frac{p_S(q_S \varepsilon_{SL})}{p_L} \Lambda_S \kappa^L_{SL} \right]}_{:=\Xi_L} \\
&= -\frac{\Xi_L}{p_L}.
\end{aligned}$$

Substituting these into the numerator of ρ_s and simplifying the price terms yields:

$$\rho_S = \frac{1}{p_S p_L^2 \cdot \Delta} \left[\Psi \cdot \Omega_L + (p_L \cdot \Xi_L) \cdot \Omega_S \right].$$

Similarly, substituting into the numerator of ρ_w yields:

$$\rho_L = \frac{1}{p_S^2 p_L \cdot \Delta} \left[\Psi \cdot \Omega_S + (p_S \cdot \Xi_S) \cdot \Omega_L \right].$$

Hence, the desired expressions are obtained.

Appendix C: First- and Second-Order Derivatives and the Elasticities and Curvatures for Multinomial Logit Demands

First- and Second-Order Derivatives

1st-order		
	$\frac{\partial q_S}{\partial p_S}$	$-\bar{\beta} q_S(1-q_S)$
	$\frac{\partial q_S}{\partial p_L}$	$\bar{\beta}\bar{x} q_S q_L$
	$\frac{\partial q_L}{\partial p_L}$	$-\bar{\beta}\bar{x} q_L(1-q_L)$
	$\frac{\partial q_L}{\partial p_S}$	$\bar{\beta} q_S q_L$
2nd-order		
	$\frac{\partial^2 q_S}{\partial p_S^2}$	$\bar{\beta}^2 q_S(1-q_S)(1-2q_S)$
	$\frac{\partial^2 q_S}{\partial p_L^2}$	$-\bar{\beta}^2\bar{x}^2 q_S q_L(1-2q_L)$
	$\frac{\partial^2 q_L}{\partial p_L^2}$	$\bar{\beta}^2\bar{x}^2 q_L(1-q_L)(1-2q_L)$
	$\frac{\partial^2 q_L}{\partial p_S^2}$	$-\bar{\beta}^2 q_S q_L(1-2q_S)$
	$\frac{\partial^2 q_S}{\partial p_S \partial p_L}$	$-\bar{\beta}^2\bar{x} q_S q_L(1-2q_S)$
	$\frac{\partial^2 q_L}{\partial p_S \partial p_L}$	$-\bar{\beta}^2\bar{x} q_S q_L(1-2q_L)$

Elasticities and Curvatures

Elasticities	
ε_{SS}	$\bar{\beta} p_S(1 - q_S)$
ε_{SL}	$\bar{\beta}\bar{x} p_L q_L$
ε_{LL}	$\bar{\beta}\bar{x} p_L(1 - q_L)$
ε_{LS}	$\bar{\beta} p_S q_S$
Curvatures	
κ^S_{SS}	$\bar{\beta} p_S(1 - 2q_S)$
κ^L_{SL}	$\bar{\beta}\bar{x} p_L(1 - 2q_L)$
κ^L_{LL}	$\bar{\beta}\bar{x} p_L(1 - 2q_L)$
κ^S_{LS}	$\bar{\beta} p_S(1 - 2q_S)$
κ^L_{SS}	$-\bar{\beta}\bar{x} p_L q_L \frac{1 - 2q_S}{1 - q_S}$
κ^S_{LL}	$-\bar{\beta} p_S q_S \frac{1 - 2q_L}{1 - q_L}$
κ^S_{SL}	$\bar{\beta} p_S(1 - 2q_S)$
κ^L_{LS}	$\bar{\beta}\bar{x} p_L(1 - 2q_L)$

References

Adachi, T., & Fabinger, M. (2022). Pass-through, welfare, and incidence under imperfect competition. *Journal of Public Economics*, 211, 104589.

Alexandrov, A. (2014). Pass-through rates when firms can vary package sizes. *Journal of Competition Law & Economics*, 10(3), 611–619.

Armstrong, M. (2016). Nonlinear pricing. *Annual Review of Economics*, 8, 583–614.

Armstrong, M., & Vickers, J. (2023). Multiproduct cost pass-through: Edgeworth's paradox revisited. *Journal of Political Economy*, 151(10), 2645–2665.

Attanasio, O., & Pastorino, E. (2020). Nonlinear pricing in village economies," *Econometrica*, 88(1), 207—263.

Bornstein, G., and Peter, A. (2025). Nonlinear pricing and misallocation. *American Economic Review*, 115(11), 3853–3908.

Cohen, A. (2008). Package size and price discrimination in the paper towel market. *International Journal of Industrial Organization*, 26(2), 502–516.

Conlon, J. R. (2017). Does buyer heterogeneity steepen or flatten quantity discounts? *RAND Journal of Economics*, 48(4), 1027—1043.

Edgeworth, F. (1925). The theory of pure monopoly. *Papers Relating to Political Economy*, Vol. 1, Macmillan.

Galichon, A. (2026). *Discrete choice models: Mathematical methods, econometrics, and data science*. Princeton University Press.

Iaria, I., & Wang, A. (2026). An empirical model of quantity discounts with large choice sets. *Journal of the European Economic Association*, Forthcoming.

Koh, P. S. (2026). Demand curvature and pass-through in differentiated oligopoly. Unpublished.

Luco, F., & Marshall, G. (2020). The competitive impact of vertical integration by multiproduct firms. *American Economic Review*, 110(7), 2041–2064.

Luo, Y., Perrigne, I., & Vuong, Q. (2018). Structural analysis of nonlinear pricing. *Journal of Political Economy*, 38(2), 512–532.

McManus, B. (2008). Nonlinear pricing in an oligopoly market: The case of specialty coffee. *RAND Journal of Economics*, 38(2), 512–532.

Miravete, E. J, Seim, K., & Thurk, J. (2025). Targeted pricing and the shape of demand. Unpublished.

Mrázová, M. & Neary, J. P. (2017). Not so demanding: Demand structure and firm behavior *American Economic Review*, 107(12), 3835–3874.

Mussa, M. & Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, 18(2), 301–317.

Salinger, M. A. (1991). Vertical mergers in multi-product industries and Edgeworth's paradox of taxation. *Journal of Industrial Economics*, 39(5), 545–556.

Weyl, E. G., & Fabinger, M. (2013). Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy*, 121(3), 528–583.

Vickers, J. (2026). A classical IO triangle: Cournot, Edgeworth, and Ramsey. *Review of Industrial Organization*, 69(1), 1–12.

Wilson, R. B. (1993). *Nonlinear pricing.* Oxford University Press.