Pass-Through and Demand Forms*

Michal Fabinger† and E. Glen Weyl‡

December 2012

Abstract

Under imperfect competition the curvature of demand is central to the rate of cost pass-through and thus to incidence and other questions of economic interest. We show that standard functional forms for demand severely and often unrealistically restrict the behavior of pass-through rates. We propose an *Adjustable pass-through* (Apt) class of demand functions that avoids these restrictions while yielding closed-form solutions to standard models. To illustrate the utility of this generalized demand form we show how to embed it in monopolistic competition. This allows us to apply it to canonical international trade models with heterogeneous firms and provide a natural case in which standard results on the competitive effects of international trade reverse.

Incidence (viz. the division between consumers and firms of gains from trade and losses from taxation of that trade) plays a central role in much economic analysis. As we show in Weyl and Fabinger (2012), it is pivotal in, among other things, the behavior of supply chains, the design of optimal procurement mechanisms and the optimal taxation of international commerce. While under perfect competition the pass-through rate is entirely determined by the elasticity of supply and demand, under imperfect competition the curvature of demand also plays a central role. Unfortunately, as we show below, standard demand forms restrict this curvature in ways that have little empirical or theoretical foundation, are hard to work with analytically. In this paper we propose a novel, simple and highly tractable class of demand functions that avoid these limitations and apply it to provide closed-form, yet flexible,

*Some results in this paper were formerly circulated as part of "Pass-through as an Economic Tool", as well as an earlier paper "Apt Demand: A Flexible, Tractable Adjustable Pass-Through Demand Function". Many individuals who contributed to those papers also aided us on the results developed her, but given that we have thanked them there for brevity we omit further acknowledgement here. We are grateful to Eric Guan and Yali Miao for their assistance in creating the Apt Demand Toolkit and to Daichi Ueda for more general research assistance.

†Department of Economics, Pennsylvania State University, University Park, PA 16802: fabinger@psu.edu.

‡Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637: weyl@uchicago.edu.

solutions to a range of canonical models of international trade. In particular, we show that in a natural case somewhat analogous to demand forms generated by income distributions, international trade is pro-competitive.[1] This behavior is opposite, for example, to what the model of Arkolakis et al. (2012) would predict in a similar context[2].

As we discuss in Section 1, pass-through rates play a key role in a wide range of problems in imperfectly competitive models, including platforms, concession auctions, price discrimination, merger analysis, entry, the welfare effects of advertising, product design and supply chains. Empirical research and theoretical considerations both suggest that properties of pass-through rates vary widely depending on the application. For example, in markets with a single-peaked distribution of valuations they will typically lie below unity and rise with prices while in market where resale is possible they will tend to lie above unity over a substantial range and decline in price. Thus, a tractable demand system allowing for all of these possibilities by varying parameters is desirable.

Yet, as we argue in Section 2, most existing functional forms used in applied analysis severely restrict both the level of pass-through and the way it changes with price. On the one hand, most standard statistical distributions used to generate demand forms are single-peaked and thus have pass-through below unity and monotonically increasing in price. While realistic in some settings, these demand forms both disallow alternative scenarios (such as that which is common with resale) and are not analytically tractable in closed-form in standard models with market power. On the other hand the demand forms that are tractable, a class identified by Bulow and Pfleiderer (1983) that includes linear, constant elasticity and negative exponential, have constant pass-through and thus rule out both of these common arrangements.

To address this disjuncture, we propose in Section 3 a class of *Adjustable pass-through* or *Apt* demand functions which are flexible regarding pass-through rates and their slope while maintaining many of the attractive features of existing demand forms. In particular the Apt class generalizes the Bulow and Pfleiderer (1983) class to allow the slope of pass-through with respect to price to vary, along with the (local-to-a-chosen-point) level, elasticity and pass-through. Unlike alternative formulations, Apt demand functions have a number of useful

[1]This effect is distinct from the pro-competitive markup dynamics of Melitz and Ottaviano (2008). In that model just increasing the population of a country would lead to lower markups due to terms in the utility function that make it non-separable. In the example we discuss in Subsection 5.3 the utility function is separable and still we obtain pro-competitive markup dynamics.

[2]Note that the model of Arkolakis et al. (2012) features supply side very different from the one considered in Subsection 5.3. For this reason one should proceed with caution when comparing the implications of the two models. Nevertheless, it is true that the assumptions about the demand lead to anti-competitive effects in one case and to pro-competitive effects in the other. We plan to discuss the relation of the two models in detail in future versions of this paper.

properties: they provide quadratic solutions to standard oligopoly pricing problems with constant returns to scale, obey the standard properties of demand theory and have closed-form surplus expressions. In fact, Apt demand is the unique class of demand functions satisfying these properties.

While the basic form of Apt demand is formulated as a demand curve facing a monopolist, it extends to a number of related contexts. As we showed in work available on request, Apt demand can be used to generate demand systems of the Apt form under monopolistic competition, when there is a multi-product monopolist or to provide distributions of random utility for discrete choice models. Section 4 focuses on extending Apt demand to the case of monopolistic competition.

To illustrate the usefulness of the Apt form, we apply it in Section 5 to models of international trade, which often assume particular functional forms of demand based on monopolistic competition, such as constant elasticity and linear. One case where such functional forms were used especially heavily was the recent "firm heterogeneity revolution". We show how to generalize two important models in that literature, Melitz (2003) and Melitz and Ottaviano (2008), to the case of Apt demand. Despite this much broader demand structure, it is possible to explicitly perform aggregation over firms with heterogeneous productivities and obtain a small number of closed-form equations for aggregate variables. This allows us to characterize, in Subsection 5.3, the effects of international trade on firm markups in a manner much more general than in the standard literature but that, unlike recent work (Arkolakis et al., 2012; Dhingra and Morrow, 2012), retains the tractability of the constant elasticity and linear forms used in previous work.

In particular we show that, while standard demand forms in the Bulow and Pfleiderer (1983) class all lead to the conclusion international trade is associated with markup dynamics that is not pro-competitive, in a natural Apt case that resembles demand curves derived from income distributions international trade lowers average markups. The intuition behind the result is natural and plausible: trade expands sales (of the best firms), leading them to reach deeper into the income distribution where demand is more elastic.

We conclude in Section 6. Detailed derivations and calculations appear in an appendix following the main text of the paper. A Mathematica toolkit developed jointly with Eric Guan and Yali Miao that graphs Apt Demand for any valid user-provided parameters is available at http://www.glenweyl.com/software.

1 Motivation

In this section we discuss why pass-through is an important parameter, how it relates to demand curvature and what existing evidence reveals about this parameter.

1.1 Importance of pass-through

In Weyl and Fabinger (2012), we analyze a wide range of applied problems with market power in which pass-through plays an important role. Rather than discussing any of these in detail here, we briefly list a few examples:

1. The local incidence of a specific (per-unit) tax under a general model of symmetric imperfect competition, $I \equiv \frac{\frac{dCS}{dt}}{\frac{dPS}{dt}}$ is $\frac{\rho}{1-(1-\theta)\rho}$, where the *conduct parameter* θ is a general measure of the degree of market power (1 is monopoly, 0 competition and $\frac{1}{n}$ in Cournot oligopoly with n symmetric firms, for example). CS and PS are consumer and producer surplus respectively.

2. The global incidence, the division of the gains from the market existing, if θ is independent of price (as it often is) $\overline{I} \equiv \frac{CS}{PS}$ is $\frac{\overline{\rho}}{1-(1-\theta)\overline{\rho}}$, where the *average pass-through rate* is the quantity-weights average of the pass-through rate at tax rates ranging from 0 to infinity: $\overline{\rho} \equiv \frac{\int_{t=0}^{\infty} Q(t)\rho(t)dt}{\int_{t=0}^{\infty} Q(t)dt}$.

3. In an auction for a monopolistic concession, knowledge about the pass-through rate and its heterogeneity across bidders summarizes the differences between bidder-efficient and socially efficient allocations of the concession.

4. In an imperfectly competitive supply chain the relationship between markups at, and the effects of changes in industrial structure of, various stages of the chain is determined by pass-through rates and their slope.

5. The slope of pass-through rates play an important role in whether third-degree price discrimination is beneficial or harmful.

1.2 Empirical evidence on curvature and pass-through

Thus the results of many applied analyses turn on simple properties of the pass-through rate, such as its comparison to unity, whether it increases or decreases in price and how it compares across different markets. These are thus crucial properties to measure in imperfectly competitive models, just as Chetty (2009) argues for more standard incident quantities.

Unfortunately, empirical studies on pass-through have thus-far strongly refuted any such prior assumptions. Barzel (1976) famously found that taxes on cigarettes were passed-through more than one-for-one to consumers. Broader studies of sales taxes have found pass-through rates ranging from below unity (Haig and Shoup, 1934), to above unity (Besley and Rosen, 1998), to typically equal to unity (Poterba, 1996), depending on the methodology used. In a more detailed industry study, Genesove and Mullin (1998) found pass-through slightly above unity even in a very competitive industry, while more macro pass-through rates of exchange rate shocks are typically found to be below unity at least in the short-term (Menon, 1995; Campa and Goldberg, 2005).

More focused studies, of a single firm, find even more widely varying results. Ashenfelter et al. (1998) and Besanko et al. (2001) found pass-through of individual firm cost shocks to that firm's price to be small, at about 25 to 60%.[3] Work with multi-product firms has found widely varying pass-through depending on the product using both accounting methodologies (Chevalier and Curhan, 1976) and detailed scanner data studies (Besanko et al., 2005); the latter study found that approximately 40% of products have pass-through rates above unity. However, few clear patterns linking product types to pass-through levels have emerged.

1.3 Pass-through and demand curvature

Pass-through under imperfect competition is determined by the elasticity of supply and demand, as well as the curvature of demand. However, when there are constant returns to scale, only the curvature of demand plays a role. For the rest of the paper we focus on this case, keeping in mind that more generally supply conditions (returns to scale) also play a role.

Let $Q(p)$ be the aggregate demand for a symmetric industry's products as a function of the symmetric price charged by the firms. All firms have constant marginal cost c. Profits are then $(p-c)Q(p)$. The first-order condition of a monopolist or a cartel would then be

$$Q(p) + (p-c)Q'(p) = 0 \iff p - c = -\frac{Q(p)}{Q'(p)} \equiv \mu(p).$$

In Weyl and Fabinger (2012) we show that a wide range of competitive environments may be represented as $p - c = \theta\mu$, where θ is the conduct parameter alluded to in Subsection 1.1 above. For example, under homogeneous products Cournot competition, $\theta = \frac{1}{n}$ where n is the number of firms and under symmetrically differentiated Bertrand competition $\theta = 1 - D$

[3] They measured pass-through as an elasticity, while we measure it as an absolute number; to obtain the absolute figure from the elasticity it must be inflated by the mark-up, which they measure. This is how we obtained this range.

where D is the aggregate diversion ratio, the fraction of sales lost from an increase in price that are recaptured by other firms, from any product to all other products. Under Cournot, θ is independent of price while under differentiated Bertrand it may depend on price. We focus on the case when θ is constant for exposition, which also includes the Delipalla and Keen (1992) symmetric conjectural variations models, many monopolistic competition models and the Cournot complementary monopoly model. Variable θ introduces additional effects driven by factors other than the aggregate demand pattern in the industry, which are at least conceptually separable from the demand curvature effects we focus on.

With constant θ, we can calculate the pass-through rate of cost increases $\rho \equiv \frac{dp}{dc}$ by implicit differentiation:

$$\rho - 1 = \theta\mu'\rho \iff \rho = \frac{1}{1-\theta\mu'}.$$

Thus the comparison of pass-through to unity is determined by the sign of μ', with $\rho > (<)1$ if $\mu' > (<)0$. μ' measures the log-curvature of demand as $(\log Q)'' = \frac{\mu'}{\mu^2}$. Note also that under monopoly, $\theta = 1$, $\mu' < 1$ is the second-order condition, namely that marginal revenue is strictly decreasing. Now $\rho' \equiv \frac{d\rho}{dc}$ can be computed by differentiation:

$$\rho' = \frac{\theta\mu''\rho}{(1-\theta\mu')^2} = \theta\mu''\rho^3.$$

The sign of ρ' is the same as that of μ'' since θ and ρ are both strictly positive. We see that ρ' is determined by the dependence of log-curvature on price. In the next section we analyze the properties of μ' and μ'' for standard demand forms. We refer to these in terms of their induced properties on ρ when there are constant returns and constant l, but when these conditions fail to hold $\rho > (<)1$ should be understood as $\mu' > (<)0$ and $\rho' > (<)0$ should be understood as $\mu'' > (<)0$.

1.4 Cases of particular interest

While empirical evidence is highly ambiguous about the magnitude and slope of pass-through rates, theoretical arguments in combination with some empirical observations at least suggest two cases of particular interest:

1. If a product has unit demand and consumer valuations are drawn from a unimodal distribution, then (constant marginal cost) monopoly pass-through is always below $\frac{1}{2}$ at prices below the mode and above $\frac{1}{2}$ at prices above the mode. This follows from the fact that in the former region demand is concave and in the latter it is convex. As we will see in the following section, many common unimodal distributions have stronger properties:

pass-through is globally increasing and below unity. One particularly natural unimodal case arises if consumer willingness-to-pay for the good is simply proportional to income, as in the Shaked and Sutton (1982) model, and income follows the typical log-normal or double-Pareto-log-normal distributions that fits observed income distributions well in many countries. The strong convexity of the Pareto distribution then strongly suggests that pass-through increases rapidly with price (and thus with the income group served), though we will see that the log-normal distribution absent a Pareto tail does not have a clearly signed slope of pass-through. Nonetheless, for many settings where unimodal distributions seem plausible, it is reasonable to expect that pass-through rates rise with price and are often below unity is reasonable.

2. Conversely, Einav et al. (2012) study online auctions where the goods being sold by an individual seller can always be resold by buyers on the auction site, at some fairly homogeneous cost (at least for the best resellers). This puts a floor on willingness-to-pay above which demand is downward sloping, creating extreme convexity of demand at low prices. Einav et al. verify this, using data from seller experiments to show that while demand is extremely convex at low prices it becomes closer to linear at higher prices. Technically Einav et al. study an "auction demand curve" rather than the demand curve for a uniform price. However the intuitions they provide for their findings suggest a mechanism that would apply more broadly and which suggests that in markets with resale or other sources of minimum "scrappage" value for the good pass-through is likely to be very high, probably above unity, at low prices and to fall as price rises.

2 Existing Demand Forms

Common functional forms in industrial organization fall into three categories:

1. Most common in theoretical work, empirical analysis of homogeneous product industries and monopolistic competition models are (special cases of) the Bulow and Pfleiderer (1983) *constant pass-through* class of demand functions defined as the set of all demand functions with $\mu'' = 0$. This has the form

$$\sigma \left(1 + \frac{\mu'}{1-\mu'} \frac{p}{m}\right)^{-\frac{1}{\mu'}}, \tag{1}$$

where μ', σ and m are parameters. This includes linear ($\mu = -1$), constant elasticity ($\sigma = \lambda m^{\epsilon}$, $\mu' = -\frac{1}{\epsilon}$, $m \to 0$, where ϵ is the elasticity and λ is a scale parameter)

and constant markup ($\mu' \to 0$) demands as special cases. These may be derived from, and are thus in a sense equivalent to, the generalized Pareto class of statistical distributions when the latter are viewed as distributions of consumer willingness-to-pay. In particular, linear demand corresponds to a uniform distribution, constant elasticity demand to a Pareto distribution and constant markup demand to a negative exponential distribution. A primary advantage of this class, and the reason why it is presumably used so frequently is its extreme tractability: it yields linear solutions to symmetric constant marginal cost oligopoly models when θ is constant.[4] While it allows arbitrary $\mu' < 1$ if the full class is analyzed, it is defined by $\mu'' = 0$ and thus has constant pass-through.

2. A common class of demand functions are those based on statistical distributions. These are more often used as building blocks for multi-product demand systems (McFadden, 1974; Berry et al., 1995) than as direct demand functions, though the log-curvature properties of these are often connected (Gabaix et al., 2010; Quint, 2012).[5]

3. The Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980) with constant expenditures has been used in many applications (Hausman, 1997).[6]

2.1 Pass-through taxonomy of demand forms

The pass-through properties of the first class are immediate. Table 1 provides a taxonomy of the properties of μ in the second and third categories. The reader should understand by a probability distribution F a demand function $Q(p) = \sigma\left(1 - F\left(\frac{p}{m}\right)\right)$ where σ and m are

[4] Monopoly is clearly a special case.

[5] One interesting statistical distribution that is often used to model income distributions is log-normal. This is not commonly used in demand analysis and has quite complex behavior of μ; we thus do not include a taxonomy of it here. However, in an appendix (to be written) we show that it has $\mu' < 0$ at low prices and $\mu' > 0$ at high prices and $\mu'' > 0$ at low prices but $\mu'' < 0$ at high prices. This may be an interesting set of predictions to explore empirically as willingness-to-pay for some products is closely related to income.

[6] The single-product version of AIDS can be written (over a particular range of prices as discussed below) as

$$D(p) = \frac{a + b\log(p)}{p} \tag{2}$$

The range of prices over which this formula can be viewed as valid depends on whether b is positive or negative. With $b > 0$, demand behaves very strangely, sloping upwards for low enough prices. We therefore only consider the (more commonly used) case when $b \leq 0$. If $b = 0$ this is just constant elasticity demand with an elasticity of 1, which violates (strict) DMR as discussed below. With $b < 0$, formula (2) is valid only for $p \leq e^{-\frac{a}{b}}$; for prices above this, demand is 0. It is this demand function that is considered in the table below as AIDS.

	$\rho < 1$	$\rho > 1$	Price-dependent	Parameter-dependent
$\rho' < 0$			AIDS with $b < 0$	
$\rho' > 0$	Normal (Gaussian) Logistic Type I Extreme Value (Gumbel) Laplace Type III Extreme Value (Reverse Weibull) Weibull with shape $\alpha > 1$ Gamma with shape $\alpha > 1$		Type II Extreme Value (Fréchet) with shape $\alpha > 1$	
Price-dependent				
Parameter-dependent				
Does not globally satisfy declining MR		Type II Extreme Value (Fréchet) with shape $\alpha < 1$ Weibull with shape $\alpha < 1$ Gamma with shape $\alpha < 1$		

Table 1: A taxonomy of some common demand functions

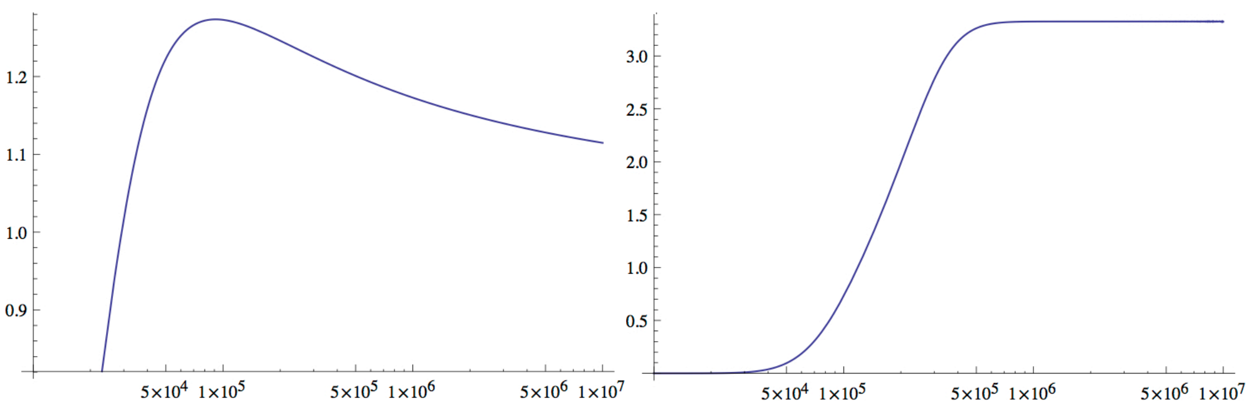

Figure 1: Constant marginal cost monopoly pass-through rates for log-normal (left) and double-Pareto-log-normal (right) distributions calibrated to the US income distribution. The x-axis has a logarithmic scale in income.

stretch parameters (Weyl and Tirole, 2012). Note that in this case

$$\mu = \frac{\sigma m \left(1 - F\left(\frac{p}{m}\right)\right)}{\sigma f\left(\frac{p}{m}\right)} = m \frac{1 - F\left(\frac{p}{m}\right)}{f\left(\frac{p}{m}\right)}.$$

Note, thus, that neither global slope nor convexity properties of μ are affected by either σ or m.

We can thus analyze the properties of relevant distributions independently of their values, as represented in the table and the following proposition.

Proposition 1. *Table 1 summarizes global properties of the listed statistical distributions generating demand functions. α is the standard shape parameter in distributions that call*

for it.

Proof. Characterization of the pass-through level (comparisons of ρ to unity) follow from classic classifications of distributions as log-concave or log-convex as in Bagnoli and Bergstrom (2005), except in the case of AIDS in which the results are novel. All other results are established in Appendix A.[7] □

An interesting statistical distributions on which we do not derive results here are the log-normal, log-logistic and double-Pareto-log-normal distribution, which typically approximate income distributions. Computational results for a log-normal distribution calibrated to the 2011 US income distribution are shown in the left panel of Figure 1. Its parameters ($\mu = 10.4$, $\sigma = .85$) were derived from 2011 US World Bank Gini coefficient of .45 and mean income of \$45989 according to the standard formula deriving σ from the Gini coefficient and expressing mean income as a function of σ and μ. They indicate that in this calibrated case, pass-through rises monotonically from below unity to above unity until the parts of the distribution where it would begin transitioning to a Pareto distribution (around \$100,000) under the double-Pareto-log-normal distribution. However, then it begins to fall while staying above unity. The right hand panel show the effect of inserting a transition to a Pareto distribution through the double-Pareto-log-normal distribution proposed by Reed (2003), with parameters ($\alpha = 1.43$, $\beta = 22.43$, $\mu = 11.27$, $\sigma = .51$) taken from his maximum likelihood estimation based on 1997 US Household Survey. In this case pass-through to levels off at a value well above one rather than falling. Thus a reasonable characterization for common income distributions is that pass-through monotonically increases from below unity to well above unity as we move up the income scale. This matches the direction of monotonicity of other single-peaked distributions, but has pass-through rates above unity for most individuals above the median, unlike with other single-peaked distributions. However, some parameter values for these distributions violate declining marginal revenue. We are still working to derive a useful analytic characterization of their behavior, which we hope to include in a future draft.

2.2 Motivation for Apt class

The taxonomy above reveals two motivations for seeking more flexible and tractable function forms. First, increasing pass-through appears to be a common feature of many if not most standard, single-peaked statistical distributions. However, such statistical distributions

[7] We do not classify the slope of pass-through for demand functions violating declining marginal revenue as this is such a common assumption that we think such forms would be unlikely to be widely used and because it is hard to classify the slope of pass-through when it is infinite over some ranges.

are difficult to work with analytically as demand functions in standard models with market power. They are tractable only computationally, which may be one reason they are not commonly employed in theoretical work or used as the basis of demand functions in international trade models of monopolistic competition. An analytically tractable functional form that matches their pass-through properties would thus be useful in settings where a single-peaked distribution is plausible but analytic results are desired.

Second, in many settings analysts may wish to be more agnostic about pass-through and its slope than allowed by existing demand forms. There is no reason why a particular class should fall into one of these four categories: different parameter values and/or prices might well lead to different pass-through rates and slopes. Table 1 allows for violations, but strikingly many commonly used distributions *do* turn out to be simply classifiable according to this taxonomy. This shows that in problems where the level and slope of pass-through are crucial, many commonly-used demand functions are too restrictive. If the distinction between $\rho > 1$ and $\rho < 1$ or between $\rho' > 0$ and $\rho' < 0$ determines an important comparative static as it does for many problems as discussed in Subsection 1.1, then the assumption that demand is of almost any of the common forms may, sometimes inadvertently, drive the conclusions of the analysis.[8] While the Bulow-Pfleiderer class allows flexibility on the level of pass-through, it clearly rules out important behavior of the slope of pass-through rates on which flexibility is desirable in many contexts.

3 Apt Demand

These restrictions can be relaxed and the tractability of increasing pass-through demand achieved by generalizing the constant pass-through demand class to allow flexibility in the slope as well as level of pass-through. The constant pass-through class has a linear μ; that is

$$\mu(p) = m\left[1 - \mu' + \mu' \frac{p}{m}\right].$$

This implies that the constant pass-through class yields linear solution to imperfectly competitive models with constant θ and constant returns. To allow the slope of pass-through to vary, a quadratic term must be added. This can be done in one of two ways. The most natural is to add a second-degree term to the expression for μ. However, this form is not very convenient because any positive coefficient on p^2 leads to infinite consumer surplus and

[8]AIDS is an exception, but even in this case the level of prices determine the properties. That is while AIDS *does* have pass-through rates that can be either side of unity it *does not* allow these to be flexible once the elasticity and level of demand have been tied down. That is, while it is first-order flexible it is not flexible on pass-through given these first-order properties. In this way it suffers from the same defects as, say, constant elasticity demand (Bulow and Pfleiderer, 1983).

thus is not derivable from a coherent utility function.

The only other alternative leading to quadratic solutions is to make μ quadratic in the square root of a term linear in p, i.e. quadratic in $\sqrt{\xi p-\gamma}$ where γ and ξ are constants.[9] This allows for quadratic solutions for p while also allowing desired limits on the slope of μ by appropriately choosing γ and ξ depending on the desired case and ensuring that any adjustments to the demand occur at prices below the equilibrium rather than above it, ensuring coherent surplus formulas and derivation of the demand from coherent utility functions. In particular, let

$$\mu(p)=m\left(\overline{\mu'}\left(\frac{p}{m}+\alpha\right)+\left[1-\overline{\mu'}\left(1+\alpha\right)\right]\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}\right), \tag{3}$$

where $\overline{\mu'}$, α and m are constant parameters. As in the case of constant pass-through demand, m may be interpreted as the optimal price of a zero-cost monopolist, since m is the value of the right-hand side of (3) for $p=m$. In the special case $\alpha=\frac{1-\overline{\mu'}}{\overline{\mu'}}$, $\mu\left(p\right)$ becomes

$$\mu\left(p\right)=\left(1-\overline{\mu'}\right)m+\overline{\mu'}p,$$

viz. constant pass-through demand. To obtain the demand form generally, we solve the ordinary differential equation implied by (3) and $\mu=-\frac{Q}{Q'}$, namely

$$-\frac{Q}{\frac{dQ}{dp}}=m\left(\overline{\mu'}\left(\frac{p}{m}+\alpha\right)+\left[1-\overline{\mu'}\left(1+\alpha\right)\right]\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}\right)\iff$$

$$\frac{dp}{m\left(\overline{\mu'}\left(\frac{p}{m}+\alpha\right)+\left[1-\overline{\mu'}\left(1+\alpha\right)\right]\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}\right)}=-\frac{dQ}{Q}.$$

Integrating both sides using standard techniques yields

$$\frac{2}{\overline{\mu'}}\log\left(1+\overline{\mu'}\left(1+\alpha\right)\left(\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}-1\right)\right)=-\log(Q)+k,$$

where k is an arbitrary constant. Taking the exponential of both sides and choosing k so

[9] Including cubic and quartic terms could also allow for explicit analytic solutions, but would require much more complex analysis as well as much more effort to exclude economically inconsistent demand forms.

that $Q(0) = \sigma$, we obtain

$$Q(p) = \sigma \left[\frac{1 + \overline{\mu'} (1 + \alpha) \left(\sqrt{\frac{\frac{p}{m} + \alpha}{1+\alpha}} - 1 \right)}{1 + \overline{\mu'} (1 + \alpha) \left(\sqrt{\frac{\alpha}{1+\alpha}} - 1 \right)} \right]^{-\frac{2}{\overline{\mu'}}}. \tag{4}$$

While this general form holds over the important range (always at sufficiently high prices, except when demand hits 0), at low prices demand must adjust to ensure it obeys declining marginal revenue in some cases.[10] Also, as with constant pass-through demand, not all parameter values correspond to well-defined demand functions. Finally, when $\overline{\mu'} = 0$ demand converges to an alternative, exponential form. These facts are summarized in a formal definition dealing with each case, supplied in Appendix B. However, we do not consider these special cases here, since throughout most of the text we are concerned with the properties of Apt demand over the range covered by the functional form (4). A Mathematica toolkit developed jointly with Eric Guan and Yali Miao that graphs Apt Demand for any valid user-provided parameters is available at http://www.glenweyl.com/software.

3.1 Demand theory properties

Apt demand is a coherent and "standard" demand function in the sense that it satisfies typical desirable properties as established in the following proposition:

Proposition 2. *Apt demand is*

1. *weakly positive,*
2. *strictly decreasing wherever it is strictly positive and finite, and therefore globally non-increasing,*
3. *continuous and infinitely differentiable wherever it is strictly positive and finite, except (for some parameter values) at a single point it may be only twice continuously differentiable,*
4. *and has strictly declining marginal revenue, except in special cases discussed in Definition 1 in Appendix B where it may have only weakly declining marginal revenue at sufficiently low prices.*

[10] Effectively this involves adjusting the demand curve to correspond to the "ironed" marginal revenue curve of Hotelling (1931).

Thus, except in one special case, any solution to $\theta = \frac{p-MC}{\mu}$ where θ is constant and MC is weakly increasing in quantity constitutes the unique symmetric equilibrium price in the industry and in the special case there is only a single cost at which uniqueness fails.

Even the last case, where marginal revenue may fail to be strictly declining at low prices, may actually be desirable; while we adjust the relevant demand form to avoid a region of increasing marginal revenue, such a region at low prices is precisely what Einav et al. (2012) find occurs. Thus the unvarnished form may be more useful (though somewhat less tractable) over this range in fitting the patterns of the "resale" case discussed in Subsection 1.4.

Proof. See Appendix B. □

3.2 Closed-form solutions

Consider a symmetric industry with a constant (independent of cost level and price) conduct parameter θ and constant (common) marginal cost c. In this setting, the equilibrium price can be computed in closed form if demand is Apt. While the computations require several steps, all are simple algebraic manipulations. To make this clear, we include this solution explicitly, despite the somewhat lengthy equations it requires:

$$p - c = \theta\mu(p) = \theta m \left(\overline{\mu'} \left(\frac{p}{m} + \alpha \right) + \left[1 - \overline{\mu'} (1+\alpha) \right] \sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}} \right) \iff$$

$$\left(1 - \theta\overline{\mu'}\right)(1+\alpha) \frac{\frac{p}{m}+\alpha}{1+\alpha} - \theta \left[1 - \overline{\mu'} (1+\alpha) \right] \sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}} - \alpha - \frac{c}{m} = 0.$$

Applying the quadratic formula to solve for $\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}$, we have

$$\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}} = \frac{\theta \left[1 - \overline{\mu'}(1+\alpha) \right] \pm \sqrt{\theta^2 \left[1 - \overline{\mu'} (1+\alpha) \right]^2 + \left(\alpha + \frac{c}{m} \right) \left(1 - \theta\overline{\mu'} \right) (1+\alpha)}}{2 \left(1 - \theta\overline{\mu'} \right) (1+\alpha)}.$$

The $\pm$ must correspond to the sign of $1+\alpha$, because this is necessary for the solution to exist and exhibit positive pass-through. We therefore obtain

$$\frac{\frac{p}{m}+\alpha}{1+\alpha} = \frac{\left(\theta \left[1 - \overline{\mu'}(1+\alpha) \right] + \frac{1+\alpha}{|1+\alpha|} \sqrt{\theta^2 \left[1 - \overline{\mu'} (1+\alpha) \right]^2 + \left(\alpha + \frac{c}{m} \right) \left(1 - \theta\overline{\mu'} \right) (1+\alpha)} \right)^2}{4 \left(1 - \theta\overline{\mu'} \right)^2 (1+\alpha)^2} \Rightarrow$$

$$p = \frac{m\left(\theta\left[1-\overline{\mu'}(1+\alpha)\right] + \frac{1+\alpha}{|1+\alpha|}\sqrt{\theta^2\left[1-\overline{\mu'}\left(1+\alpha\right)\right]^2 + \left(\alpha+\frac{c}{m}\right)\left(1-\theta\overline{\mu'}\right)\left(1+\alpha\right)}\right)^2}{4\left(1-\theta\overline{\mu'}\right)^2\left(1+\alpha\right)} - m\alpha. \tag{5}$$

Thus Apt demand yields closed-form, quadratic solutions to the symmetric oligopoly pricing problem with linear costs and a constant conduct parameter.

3.3 The Bulow-Pfleiderer demand form as a special case

Note that when $\alpha = \frac{1-\overline{\mu'}}{\overline{\mu'}}$, Apt demand becomes simply

$$\sigma\left(\frac{1+\overline{\mu'}\frac{1}{\overline{\mu'}}\left[\sqrt{\left(\frac{p}{m}+\frac{1-\overline{\mu'}}{\overline{\mu'}}\right)\overline{\mu'}}-1\right]}{1+\overline{\mu'}\frac{1}{\overline{\mu'}}\left[\sqrt{1-\overline{\mu'}}-1\right]}\right)^{-\frac{2}{\overline{\mu'}}} = \sigma\left(\sqrt{\frac{p}{m}\frac{\overline{\mu'}}{1-\overline{\mu'}}+1}\right)^{-\frac{2}{\overline{\mu'}}} = \sigma\left(1+\frac{\overline{\mu'}}{1-\overline{\mu'}}\frac{p}{m}\right)^{-\frac{1}{\overline{\mu'}}},$$

which is the constant pass-through form (1) where μ' is identified with $\overline{\mu'}$. The exponential demand case is also nested, as discussed in Appendix B.

3.4 Pass-through properties, flexibility and identification

In most cases, Apt demand has μ' and μ'' globally either above, below or equal to 0 for each point in the parameter space and thus, for given parameter values, can be classified according to the taxonomy in Section 2. The exceptional case is given by $\overline{\mu'} < 0 \leq \alpha$. In this case pass-through is decreasing in price: $\mu'' < 0$, and for low prices $\mu' > 0$, while for high prices $\mu' < 0$.

Proposition 3. *If $\overline{\mu'} < (>)0$ Apt demand is globally strictly log-concave (convex), except in the case when $\overline{\mu'} < 0$ and $\alpha \geq 0$ in which case $\mu' \lesseqqgtr 0$ for*

$$p \gtreqqless m\left(\frac{\left[1-\overline{\mu'}\left(1+\alpha\right)\right]^2}{4\overline{\mu'}^2\left(1+\alpha\right)} - \alpha\right).$$

If $\overline{\mu'} = 0$ and $\alpha < 0$, Apt demand is globally strictly log-concave, while if $\alpha \geq 0$ it is strictly log-convex. If $Q(p) = \sigma e^{-\frac{p}{m}}$ it is log-linear. When $1 \gtreqqless \overline{\mu'}\left(1+\alpha\right)$, Apt demand has globally $\mu'' \lesseqqgtr 0$ in the range where equation (4) applies. These properties hold weakly wherever demand is finite and strictly positive. Finally if $Q(p) = \sigma e^{-\frac{p}{m}}$ then $\mu'' = 0$ globally.

Proof. See Appendix B. □

Apt demand can fit arbitrary levels, elasticities and values of μ' locally < 1 and a wide range of local values of μ'', as described in the following proposition. Thus, in particular, it can fit the two patterns desired in Subsection 1.4.[11]

Proposition 4. *For any strictly positive collection of price $p^\star$, quantity at this price $Q(p^\star) > 0$, market power at that price $\mu(p^\star) > 0$, $\mu'(p^\star) < 1$ and $\mu''(p^\star)$ there exists an Apt demand function which achieves these values of Q, μ, μ' and μ'' at that price so long as $\mu'(p^\star) \neq 0$,*

$$\frac{\mu''(p^\star)\,\mu(p^\star)}{[\mu'(p^\star)]^2} \leq \frac{1}{8}$$

and if $\mu'(p^\star) > \frac{2}{3}$

$$\mu''(p^\star)\,\mu(p^\star) \leq [2\mu'(p^\star) - 1]\,[1 - \mu'(p^\star)]\,.$$

The only other restrictions relate to the fact that demand must be well-behaved at $p = 0$ and arise only if the observed price is sufficiently high. For example, when $\mu''(p^\star) = 0$ we know that $\mu'(p^\star)\,p^\star < \mu(p^\star)$). We are currently working to characterize this restriction and a future draft of this paper will contain a full characterization.

Proof. See Appendix C. $\square$

In order to identify Apt demand from data, one therefore needs to measure (quantities equivalent to) the level of demand and its first *three* derivatives. This may be done with exogenous variation of price sufficient to identify level and the first three derivatives of demand or with exogenous variation in cost sufficient to measure the first two derivatives, combined with a supply side model linking pass-through rates (through the level of competition) and their slope to the second and third derivatives of demand. In the latter case the first two derivatives of demand thus identified are used along with the pass-through rate and its first derivative. This latter approach has the advantage of over-identifying the model through the connection between the level of pass-through and the first two derivatives of demand. We discuss both of these identification strategies further in Appendix C.

3.5 Surplus

Consumer surplus with Apt demand also has a closed form which we omit here and include in Appendix B for the sake of brevity.

[11]However, the following proposition does restrict the magnitude of $\frac{\mu''\mu}{(\mu')^2}$ that it can match to a range that is well below the value this takes on at some point for many unimodal distributions. Hopefully matching the qualitative properties is more important in many applications than is the quantitative fit along this dimension.

4 Monopolistic competition

There are many potential extensions of the single-product demand formula (4) to more general situations. Here we focus on the case of monopolistic competition in which analytic solutions are often employed, and thus to which Apt demand seems well-suited. Consider a utility function of the form

$$U\left(\{q(\omega)\}_{\omega\in\Omega}\right) = u\left(Q_1, Q_\theta, Q_{2\theta-1}\right), \tag{6}$$

where

$$Q_1 \equiv \int_\Omega q(\omega)\, d\omega,\;\; Q_\theta \equiv \int_\Omega q^\theta(\omega)\, d\omega,\;\; Q_{2\theta-1} \equiv \int_\Omega q^{2\theta-1}(\omega)\, d\omega. \tag{7}$$

There is a continuum of varieties ω of differentiated goods, belonging to the set Ω. The amounts consumed are denoted $q(\omega)$. The parameter θ plays the role of $1-\frac{\overline{\mu'}}{2}$. Although only the dependence on Q_1, Q_θ, and $Q_{2\theta-1}$ is indicated explicitly, the function u may depend on other variables independent of $q(\omega)$, such as consumption of goods from other sectors of the economy. This specification is quite general, encompassing many distinct modeling possibilities.

The consumer first-order condition is

$$\frac{1}{p(\omega)}\frac{\delta}{\delta q(\omega)} u\left(Q_1, Q_\theta, Q_{2\theta-1}\right) = b,$$

where $b>0$ on the right hand side does not depend on ω, $\delta/\delta q(\omega)$ denotes a functional derivative, and $p(\omega)$ is the price of variety ω. Performing the differentiation gives

$$A_0 + A_1 q^{\theta-1}(\omega) + A_2 q^{2\theta-2}(\omega) = b\; p(\omega), \tag{8}$$

with

$$A_0 \equiv u_1\left(Q_1, Q_\theta, Q_{2\theta-1}\right),\;\; A_1 \equiv \theta u_2\left(Q_1, Q_\theta, Q_{2\theta-1}\right),\;\; A_2 \equiv (2\theta-1)\, u_3\left(Q_1, Q_\theta, Q_{2\theta-1}\right).$$

The first-order condition allows us to solve for $q(\omega)$:

$$q(\omega) = \left(\frac{-A_1 + \frac{1-\theta}{|1-\theta|}\sqrt{A_1^2 + 4A_2\left(b\; p(\omega) - A_0\right)}}{2A_2}\right)^{\frac{1}{\theta-1}}.$$

The sign factor in front of the square root was chosen to be $\frac{1-\theta}{|1-\theta|}$ to ensure that the demand is

downward sloping. This expression for $q(\omega)$ is precisely of the form (4) with the identification

$$
\begin{aligned}
&\theta = 1 - \tfrac{\overline{\mu'}}{2}, \\
&\tfrac{A_0}{b} = \tfrac{m}{1+\alpha}\left(1 + \alpha - \tfrac{1}{\overline{\mu'}}\right)^2 - \alpha m, \\
&\tfrac{A_1}{b} = -2\left(1+\alpha\right) m \sigma^{\frac{\overline{\mu'}}{2}} \left(\tfrac{1}{\overline{\mu'}(1+\alpha)} + \sqrt{\tfrac{\alpha}{1+\alpha}} - 1\right)\left(\tfrac{1}{\overline{\mu'}(1+\alpha)} - 1\right), \\
&\tfrac{A_2}{b} = \left(1+\alpha\right) m \sigma^{\overline{\mu'}} \left(\tfrac{1}{\overline{\mu'}(1+\alpha)} + \sqrt{\tfrac{\alpha}{1+\alpha}} - 1\right)^2.
\end{aligned}
\tag{9}
$$

While we maintain this alternative notation in this and the following section for convenience, the residual demand functions explored here are precisely the Apt demand class.

Special cases

As with single-product Apt demand, the general from (6) nests virtually all standard analytically tractable cases of demand and utility used in the literature. In particular, the choice

$$u\left(Q_1, Q_\theta, Q_{2\theta-1}\right) = A_2 Q_{\frac{\sigma-1}{\sigma}}$$

with $\theta = 1 - \frac{1}{2\sigma} \in (0,1)$ and $A_2 > 0$ independent of any $q(\omega)$ gives a constant elasticity of substitution (CES) utility function with elasticity of substitution σ.

To obtain quadratic utility of the type used in Ottaviano et al. (2002) or Melitz and Ottaviano (2008), we can choose

$$u\left(Q_1, Q_\theta, Q_{2\theta-1}\right) = \widetilde{\alpha} Q_1 - \frac{1}{2}\eta Q_1^2 + A_2 Q_2$$

with $\theta = \frac{3}{2}, \widetilde{\alpha} > 0, \eta > 0$, and $A_2 < 0$ all independent of any $q(\omega)$. In this case A_0 is a positive linear function of Q_1 with a negative slope.

In both cases mentioned here A_1 vanishes. More broadly, the residual demand curves with $A_1 = 0$ correspond[12] to the Bulow and Pfleiderer (1983) class of constant pass-through forms.

5 Application to Heterogeneous Firm Trade Models

A central challenge for the tractability of international trade models is aggregation. The progress made in recent years by Melitz (2003) and Melitz and Ottaviano (2008) in allowing such models to be solved in the presence of firm heterogeneity was largely due to the ability

[12] Note that when $\alpha = \frac{1-\overline{\mu'}}{\overline{\mu'}}$, $A_1 = 0$.

in these settings to summarize the consequences of the firm heterogeneity by a small number of aggregate equations. Here we show that this attractive aggregation property extends beyond the CES and linear classes studied by Melitz (2003) and Melitz and Ottaviano (2008) respectively to the much broader Apt class derived in the previous section.[13]

In all but the demand structure we allow, we follow Melitz (2003), as closely as possible. Consider a single country with labor endowment L, which is the only factor of production.[14] There is a continuum Ω of firms, each producing a unique variety ω of a differentiated good. The firms use linear technology: the quantity $q(\omega)$ produced with labor $l(\omega)$ is

$$q(\omega) = \frac{l(\omega) - f}{a(\omega)},$$

where f is a fixed cost common to all firms and $1/a(\omega)$ is firm-specific productivity. Denoting the common wage rate by w , the profit of the firm is

$$\pi(\omega) = (p(\omega) - wa(\omega))\, q(\omega) - wf.$$

Omitting the argument ω, the firm's first-order condition implies that its marginal cost equals its marginal revenue:

$$wa = p'q + p, \tag{10}$$

where p' is the derivative of the price p with respect to q. Using (8), we may rewrite this more explicitly as

$$bwa = A_0 + \theta A_1 q^{\theta-1} + (2\theta - 1)\, A_2 q^{2\theta-2}. \tag{11}$$

This equation combined with the consumer first-order condition (8) leads to the following formula for the quantity optimally produced by a firm with productivity parameter a:[15]

$$q(a) = \left(\frac{-\theta A_1 \pm \sqrt{\theta^2 A_1^2 + 4(2\theta - 1)\, A_2 (bwa - A_0)}}{2(2\theta - 1)\, A_2}\right)^{\frac{1}{\theta-1}}. \tag{12}$$

Entry into the industry is unrestricted, but requires a fixed cost of f_e units of labor. Ex ante all firms are the same, but after paying the sunk cost of entry, they learn the value of

[13] A discussion of concrete economic consequences of these results will be included in a future version of this work or in a separate paper.

[14] Generalization to the case of multiple countries is straightforward, as in the Melitz (2003) model. Alternatively, effects of trade liberalization may be inferred from the single-country model by looking at the effects of changing the labor endowment L.

[15] The argument a of the optimal quantity function $q(a)$ should not be confused with the suppressed variety identifier ω.

their productivity parameter a, which has a common distribution function $g(a)$, with the corresponding cumulative distribution function denoted $G(a)$. Then the firm may choose to exit the industry at no additional cost, or stay in the industry and produce in each period. In each period, producing firms are forced to exit with an exogenous probability δ. These events are interpreted as adverse firm-specific shocks.

Because of the unrestricted entry into the industry, firms make ex ante just enough profit[16] to cover the entry cost:

$$\frac{E\pi}{\delta} = wf_e.$$

More explicitly, the unrestricted entry condition is

$$\int_{a \le a_c} \left(\left(p(a) - wa \right) q(a) - wf \right) g(a)\, da = \delta w f_e, \tag{13}$$

where $p(a)$ and $q(a)$ are the price and quantity optimal for a firm with productivity parameter a, and cutoff value a_c is the highest value of a for which firms find it optimal to stay in the industry.

It is common in international trade models to assume that the productivity parameter a follows a Pareto distribution truncated to an interval (a_L, a_H):

$$G(a) = \frac{a^k - a_L^k}{a_H^k - a_L^k}.$$

We will maintain this assumption here.

5.1 Integral evaluation

For future convenience define $I_\kappa(q(a_1), q(a_2))$ as

$$I_\kappa(q(a_1), q(a_2)) \equiv \int_{a_1}^{a_2} q^\kappa(a)\, dG(a),$$

where the function $q(a)$ is given by (12). This integral can be evaluated explicitly, with the result

$$I_\kappa(q_1, q_2) = \frac{k(bw)^{-k}}{a_H^k - a_L^k} \left(\theta A_1 \tilde{I}_{\frac{\kappa}{\theta-1}+1}(q_1, q_2) + 2(2\theta - 1) A_2 \tilde{I}_{\frac{\kappa}{\theta-1}+2}(q_1, q_2) \right),$$

[16] Here we assume, as usual in this type of models, that there is no time discounting. This assumption is made for simplicity and can be easily lifted.

where the $\tilde{I}$s on the right hand side are given by

$$\tilde{I}_\alpha(q_1, q_2) = \left[\frac{x^\alpha}{\alpha} \left(\frac{A_0 + \theta A_1 x + (2\theta - 1) A_2 x^2}{(1 - O_+ x)(1 - O_- x)} \right)^{k-1} F_1(\alpha, 1-k, 1-k, \alpha; O_- x, O_+ x) \right]_{x = q_1^{\theta - 1}}^{x = q_2^{\theta - 1}}$$

with

$$O_\pm = -\frac{2(2\theta - 1)}{\theta A_1 \mp \sqrt{\theta^2 A_1^2 - 4(2\theta - 1) A_0 A_2}},$$

where F_1 is the Appell hypergeometric function (analytic in the relevant range).

5.2 Equations for aggregate variables

5.2.1 Zero cutoff profit condition

For a firm indifferent between exiting and staying in the industry, we have

$$(p_c - wa)\, q_c = wf.$$

Here the subscript 'c' stands for 'cutoff'. Using the consumer first-order condition (8), this is

$$A_0 q_c + A_1 q_c^\theta + A_2 q_c^{2\theta - 1} = bw\,(f - a_c q_c)\,.$$

Recalling (11), simple algebraic manipulation yields the following form of the zero cutoff profit condition:

$$bw = \frac{1 - \theta}{f} \left(A_1 q_c^\theta + 2 A_2 q_c^{2\theta - 1} \right). \tag{14}$$

We see that the zero cutoff profit condition allows us to express bw in terms of q_c.

5.2.2 Quantity aggregation

If the total mass of firms entering each period is $M\delta$, the quantities Q_1, Q_θ, and $Q_{2\theta-1}$ defined in (7) can be expressed as

$$Q_1 = M I_1(q_c, q(a_L))\,, \quad Q_\theta = M I_\theta(q_c, q(a_L))\,, \; Q_{2\theta - 1} = M I_{2\theta - 1}(q_c, q(a_L))\,. \tag{15}$$

The value of $q(a_L)$ is given by equation (12), where we can substitute for bw from (14), if desired. This means that $q(a_L)$ depends only on A_0, A_1, A_2, and q_c, and on known parameters.

5.2.3 Unrestricted entry condition

The unrestricted entry condition (13) may be written as

$$A_1 I_{\theta-1}\left(q_c, q\left(a_L\right)\right) + 2A_2 I_{2\theta-1}\left(q_c, q\left(a_L\right)\right) = \frac{bw}{\theta-1}\left(\delta f_e + f I_0\left(q_c, q\left(a_L\right)\right)\right).$$

Here, again, we can substitute for bw from (14):

$$\frac{A_1 I_{\theta-1}\left(q_c, q\left(a_L\right)\right) + 2A_2 I_{2\theta-1}\left(q_c, q\left(a_L\right)\right)}{A_1 q_c^{\theta} + 2A_2 q_c^{2\theta-1}} = \frac{\delta f_e}{f} + I_0\left(q_c, q\left(a_L\right)\right). \tag{16}$$

with $q\left(a_L\right)$ determined by A_0, A_1, A_2, and q_c as above.

5.2.4 Closing the model

Suppose there is no other sector in the economy besides the differentiated good sector we just described. If A_0, A_1, A_2 are given constants, then equation (16) itself determines q_c. This is the core equation of the model, and in most cases needs to be solved numerically. Once q_c is known, all aggregate variables of interest may be computed using the equations above. If A_0, A_1, A_2 depend on Q_1, Q_θ, and $Q_{2\theta-1}$, then determination of q_c requires not only (16), but also (15) and a separate equation for M. The aggregate version $\int_\Omega p\left(\omega\right) q\left(\omega\right) d\omega = wL$ of the unrestricted entry condition provides such equation. It may be rewritten as

$$A_0 I_1\left(q_c, q\left(a_L\right)\right) + A_1 I_\theta\left(q_c, q\left(a_L\right)\right) + A_2 I_{2\theta-1}\left(q_c, q\left(a_L\right)\right) = bw\frac{L}{M}.$$

Solving this set of equations numerically gives the value of q_c. Again, with the knowledge of q_c all aggregate variables of interest may be evaluated using the relations above.

Adding other sectors to the economy does not present a problem. The same equations hold, except that L needs to be replaced by L_1, the labor used by the differentiated good sector:

$$A_0 I_1\left(q_c, q\left(a_L\right)\right) + A_1 I_\theta\left(q_c, q\left(a_L\right)\right) + A_2 I_{2\theta-1}\left(q_c, q\left(a_L\right)\right) = bw\frac{L_1}{M}.$$

Of course, we also need an additional condition linking L_1 to variables in the other sectors. Deriving such condition is straightforward. For example, if there is one outside good 'o' produced only with labor in such a way that output Q_o requires $L_o\left(Q_o\right)$ units of labor, then the following conditions close the model:

$$Q_o \frac{dU}{dQ_o} = bw L_o\left(Q_o\right), \quad L_o\left(Q_o\right) + L_1 = L.$$

Generalization to multiple outside goods of this kind is immediate.

Thus with the general Apt demand, it is possible, because of the power form this demand function takes, to aggregate in Melitz-type trade models in analytic form. That is, the problem with an infinite number of heterogeneous firms may be reduced to a small number of analytic equations for the aggregate variables. Using these insights in empirical work is likely to be valuable, unless, of course, CES demand (or linear demand) approximates real-world situations closely. We may pursue such an empirical estimation in a future draft of this paper or future work, but for the time being we confine ourselves to a theoretical investigation of how the impacts of trade liberalization on markups depend on the demand structure. This helps us show how simple analytical insights emerge from the combination of tractability and generality that Apt demand affords.

5.3 Example: the "competitive" effects of trade liberalization

We consider one of the simplest versions of the proposed general model and examine its behavior in the case of liberalization of trade between two countries. The modeling choice we make here provides a natural benchmark for the numerous more complicated instances of the general model.

In this benchmark model, there are two symmetric countries, each with labor endowment L. In autarky each of the countries is identical to the economy described in previous subsections, without any outside good. After trade liberalization consumers will be allowed to purchase varieties of the differentiated good from foreign producers. Trade between the countries is then free, except for a fixed cost f_x that each exporting firm needs to pay in order to access the foreign market.

We choose the utility function to be a fixed linear combination of Q_1, Q_θ, and $Q_{2\theta-1}$, which implies that A_0, A_1, and A_2 are constant.[17] This includes the Melitz (2003) model, but not the Melitz and Ottaviano (2008) model, as a special case. Both before and after trade liberalization, the demand equation (8) holds: $A_0 + A_1 q^{\theta-1}(\omega) + A_2 q^{2\theta-2}(\omega) = b\, p(\omega)$, with b independent of ω. Here $p(\omega)$ is the price the firm charges in a particular country and $q(\omega)$ is the demand in that country. Note that because of the symmetry between the two countries, any exporting firm will charge the same prices at home and abroad and it will sell the same amounts in both countries. For this reason we do not have to introduce an additional index distinguishing the two markets. We will use the freedom to normalize prices to set $b = 1$, as this normalization greatly simplifies the discussion. The demand equation

[17] The variables Q_1, Q_θ, and $Q_{2\theta-1}$ are still defined by the integrals (7), but this time the integrals involve also varieties produced abroad.

then takes the form

$$A_0 + A_1 q^{\theta-1}(\omega) + A_2 q^{2\theta-2}(\omega) = p(\omega). \tag{17}$$

Consider the impact of trade on the (relative-to-price) average markups $\frac{p-mc}{p}$ that firms charge to their customers.[18] The overall distribution of markups will be influenced by two different forces: the composition of producing firms will change, and some firms will sell not only in their domestic market, but also abroad. A natural question to ask is whether international trade has a disciplining effect, in the sense of lowering markups that firms charge to their domestic customers.

As in previous subsections, let us focus on the case of Pareto distributed firm productivity, and thus Pareto distributed marginal costs aw. If the Pareto parameter is fixed, the precise distribution depends only on the values of the two ends of its support, i.e. on the marginal cost $mc_L \equiv a_L w$ of the most productive firm and the marginal cost $mc_c \equiv a_c w$ of the least productive producing firm. As in Melitz (2003), trade liberalization will increase mc_L and decrease mc_c, truncating the distribution of marginal costs at both ends; for brevity we omit a demonstration of this here, but include it in Appendix D. The value of a_L is often set to zero; if we do this in the present model, trade liberalization will truncate only the high end of the marginal cost distribution.[19] As a result, liberalization lowers the distribution of costs and thus the distribution of prices, raises the quantity sold by firms. By the Lerner formula, proportional markups are the inverse of the price elasticity of demand. Thus if price elasticity is constant, as with CES, markups are invariant to liberalization as in Melitz (2003). If price elasticities decline in quantities, markups will rise with liberalization. This would be the case, for example, under the demand assumptions of Arkolakis et al. (2012).[20] The Apt form allows us to explore how broadly this conclusion holds.

To do this note that, taking the log-derivative of equation (17) and inverting yields

$$\epsilon(q) = \frac{1}{1-\theta} \frac{A_0 + A_1 q^{\theta-1} + A_2 q^{2\theta-2}}{A_1 q^{\theta-1} + 2A_2 q^{2\theta-2}}. \tag{18}$$

First, consider the Bulow-Pfleiderer class, where $A_1 = 0$. Then expression (18) becomes,

[18]The discussion that follows is largely independent of the precise weight function used in calculating the average, as long as it depends only on equilibrium prices and quantities.

[19]This is often a reasonable approximation because the high end of the productivity distribution contributes little to economic quantities of interest.

[20]In Arkolakis et al. (2012), this assumption is referred to as log-concavity of demand. Note, however, that this differs from the standard definition of log-concavity (the concavity of the logarithm of demand in the linear price) and is more closely connected to Marshall (1890)'s Second Law of Demand.

using the relations (9),

$$\epsilon(q) = \frac{1}{1-\theta}\frac{A_0 + A_2 q^{2(\theta-1)}}{2A_2 q^{2(\theta-1)}} = \frac{1 + \frac{A_0}{A_2} q^{\overline{\mu'}}}{\overline{\mu'}}.$$

In the Bulow-Pfleiderer case, $\alpha = \frac{1-\overline{\mu'}}{\overline{\mu'}}$, the expressions (9) for A_0 and A_2 reduce to $A_0 = -bm\frac{1-\overline{\mu'}}{\overline{\mu'}}$ and $A_2 = bm\sigma^{\overline{\mu'}}\frac{1-\overline{\mu'}}{\overline{\mu'}}$. Thus the sign of $\frac{A_0}{A_2}$ is always negative.[21] This implies that ϵ declines in q regardless of the sign of $\overline{\mu'}$ so long as $\epsilon > 0$. This shows that the entire Bulow-Pfleiderer class yields[22] a result in the spirit of Arkolakis et al. (2012).

However, consider the Apt demand case when $\overline{\mu'} > 0$, so that pass-through is above unity, and $\alpha(1+\alpha) > 1$, so that pass-through is increasing in price. This corresponds to the case for a demand curve generated by standard income distributions as discussed in Subsection 2.1 above.

Then, if A_0 is small (because, for example, m is small; it is zero under CES) or if q is sufficiently small, the contribution from the A_0 term to expression (18) is small.[23] Thus that expression becomes approximately

$$\frac{1}{1-\theta}\left(1 - \frac{A_2 q^{2(\theta-1)}}{A_1 q^{\theta-1} + 2A_2 q^{2(\theta-1)}}\right) = \frac{2}{\overline{\mu'}}\left(1 - \frac{1}{2 + \frac{A_1}{A_2} q^{\frac{\overline{\mu'}}{2}}}\right). \tag{19}$$

A_1 and A_2 have the same sign in this case as A_2 is clearly positive because $1+\alpha > 0$ and A_1 is positive because $\frac{1}{\overline{\mu'}(1+\alpha)} - 1 < 0$. Therefore

$$\frac{1}{\overline{\mu'}(1+\alpha)} + \sqrt{\frac{\alpha}{1+\alpha}} - 1 > \frac{1}{1+\alpha} + \sqrt{\frac{\alpha}{1+\alpha}} - 1 = \sqrt{\frac{\alpha}{1+\alpha}} - \frac{\alpha}{1+\alpha} > 0,$$

as $\frac{\alpha}{1+\alpha} < 1$. Thus the expression (19) is clearly increasing in q. Note that in the same context the behavior of the model of Arkolakis et al. (2012) would be opposite. Thus, under the Apt demand form, when pass-through is above unity and increasing in price, trade is "pro-competitive" in the sense of reducing relative markups when either we are close to the CES case or initial quantity is small.

Using the identification of this case with demand functions generated by income distributions, the intuition behind this result is quite clear: trade broadens the sales of luxury goods leading goods to appeal to those with closer to middle incomes, who are in more elastic parts of the income distribution. This seems like a fairly plausible case, though we have verified

[21] Unless $A_0 = 0$ or $A_2 \to \infty$, or unless we allow for nonpositive $\sigma^{\overline{\mu'}}$.

[22] Obviously the limiting case of CES does so only weakly.

[23] In Appendix D we show that A_0 can be fairly large and this result still holds.

that with double-Pareto-log-normal or log-normal distributions calibrated to the US income distribution as in Subsection 2.1 the average markup behavior is anti-competitive, though in the former case at high incomes the behavior is very close to CES. In the future we hope to investigate if a plausible model of monopolistic competition based on income distributions generates pro-competitive effects of trade. Initial calculations suggest that Type I Extreme Value idiosyncratic preferences coupled with a double-Pareto-log-normal income distribution similar to what is commonly used industrial organization demand systems (Berry et al., 1995) would make trade liberalization pro-competitive at high income cutoffs for purchasing the good.[24]

Nonetheless, it seems unlikely that such a conjecture would have arisen in the absence of a parameterized class, like Apt demand, that allows the easy, closed-form exploration of these possibilities. Abstract exploration of general demand functions, as Dhingra and Morrow (2012) undertake, leaves unclear whether the requisite local conditions they identify are satisfied by any well-behaved demand function. For example, trying to plug local values from Dhingra and Morrow that yield pro-competitive markup dynamics of international trade into the Bulow-Pfleiderer form runs into inconsistencies with permissible parameter ranges arising from non-local restrictions necessary for coherence (viz. that the optimal price for a monopolist with zero cost exists). Thus Dhingra and Morrow's analysis, while suggestive, is incomplete, as well as providing limited intuition about the economic circumstances in which such results are plausible. However, once the right tool is available, the pro-competitive effects of trade may arise from natural economic intuition about income distributions.

6 Conclusion

This paper shows that most common demand forms bias incidence analysis, are not analytically tractable or both in imperfectly competitive models. We proposed an alternative demand form that is both flexible and tractable. Applied to international trade models, this demand form allows for straightforward aggregation. This allows a characterization of the competitive effects of international trade that is both sharper and more general than those existing literature, illuminating a plausible case in which international trade is pro-competitive.

Our demand form suggests several directions for future research, either or own or by others. First, while we developed an extension of the basic form to monopolistic competition

[24]We have verified examples with discrete income distributions and Gumbel idiosyncratic preferences, but our simulations in the Berry et al. (1995)-like case are not accurate enough as of yet to be sufficiently confident they are correct. However, extensive evidence points towards elasticities falling in price over many ranges (Nevo, 2010).

here, it could also be extended to form a multi-product demand system either directly or through being used as a statistical distribution through the inverse of the process through which demand forms are derived from statistical distributions. The latter could be useful to allow for a class of statistical distributions with flexible hazard rate properties, which might, for example, model idiosyncratic valuation components in a discrete choice demand system. Second, while we studied one application of our form (to the positive competitive effects of international trade) many others inside and outside of international trade seem natural, the simplest of which would be the normative impacts of international trade that Dhingra and Morrow focus on. Finally, we proposed a particular functional form here rather than a means of parameterizing a general demand form. These typically have dual relationship: Weyl and Tirole (2012) derived the stretch parameterization they studied from transformations that preserve pass-through rates. It would thus be natural to derive, based on our work, transformations that allow variation in pass-through rates and/or their slopes but preserve some higher-order properties of demand.

References

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare, "The Elusive Pro-Competitive Effects of Trade," 2012. https://dl.dropbox.com/u/2734209/ACDR_paper.pdf.

Ashenfelter, Orley, David Ashmore, Jonathan B. Baker, and Signe-Mary McKernan, "Identifying the Firm-Specific Cost Pass-Through Rate," 1998. http://www.ftc.gov/be/workpapers/wp217.pdf.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? Intra-national Trade Costs and World Price Pass-Through in South Asia and Sub-Saharan Africa," 2012. This paper is under preparation. Email David Atkin at david.atkin@yale.edu for a draft.

Bagnoli, Mark and Ted Bergstrom, "Log-Concave Probability and its Applications," *Economic Theory*, 2005, *26* (2), 445–469.

Baker, Jonathan B. and Timothy F. Bresnahan, "Estimating the Residual Demand Curve Facing a Single Firm," *International Journal of Industrial Organization*, 1988, *6* (3), 283–300.

Barzel, Yoram, "An Alternative Approach to the Analysis of Taxation," *Journal of Political Economy*, 1976, *84* (6), 1177–1197.

Berry, Stephen, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

Besanko, David, David Dranove, and Mark Shanley, "Exploiting a Cost Advantage and Coping with a Cost Disadvantage," *Management Science*, 2001, *47* (2), 221–235.

_ , Jean-Pierre Dubé, and Sachin Gupta, "Own-Brand and Cross-Brand Retail Pass-Through," *Marketing Science*, 2005, *24* (1), 123–137.

Besley, Timothy J. and Harvey Rosen, "Sales Taxes and Prices: An Empirical Analysis," *National Tax Journal*, 1998, *52* (2), 157–178.

Bulow, Jeremy I. and Paul Pfleiderer, "A Note on the Effect of Cost Changes on Prices," *Journal of Political Economy*, 1983, *91* (1), 182–185.

Campa, José Manuel and Linda S. Goldberg, "Exchange Rate Pass-Through into Import Prices," *The Review of Economics and Statistics*, 2005, *87* (4), 679–690.

Chetty, Raj, "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 2009, *1*, 451–488.

Chevalier, Michael and Ronald C. Curhan, "Retail Promotions as a Function of Trade Promotions: A Descriptive Analysis," *Sloan Management Review*, 1976, *18* (3), 19–32.

Deaton, Angus and John Muellbauer, "An Almost Ideal Demand System," *American Economic Review*, 1980, *70* (3), 312–326.

Delipalla, Sofia and Michael Keen, "The Comparison Between Ad Valorem and Specific Taxation under Imperfect Competition," *Journal of Public Economics*, 1992, *49* (3), 351–367.

Dhingra, Swati and John Morrow, "Monopolistic Competition and Optimum Product Diversity Under Firm Heterogeneity," 2012. http://www.sdhingra.com/selection3rdGain.pdf.

Einav, Liran, Am Finkelstein, and Mark R. Cullen, "Estimating Welfare in Insurance Markets Using Variation in Prices," *Quarterly Journal of Economics*, 2010, *125* (3), 877–921.

_ , Theresa Kuchler, Jonathan Levin, and Neel Sundaresan, "Learning from Seller Experiments in Online Markets," 2012. http://www.stanford.edu/∼leinav/Seller_Experiments.pdf.

Gabaix, Xavier, David Laibson, Deyuan Li, Hongyi Li, Sidney Resnick, and Caspar G. de Vries, "Extreme Value Theory and the Equilibrium Prices in Large Economies," 2010. http://pages.stern.nyu.edu/∼xgabaix/papers/CompetitionEVT.pdf.

Genesove, David and Wallace P. Mullin, "Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890-1914," *RAND Journal of Economics*, 1998, *29* (2), 355–377.

Haig, Robert M. and Carl Shoup, *The Sales Tax in the American States*, New York: Columbia University Press, 1934.

Hausman, Jerry A., "Valuation of New Goods under Perfect and Imperfect Competition," in Timothy F. Bresnahan and Robert J. Gordon, eds., *The Economics of New Goods*, Vol. 58 of National Bureau of Economic Research*, Studies in Income and Wealth*, Chicago and London: University of Chicago Press, 1997, pp. 209–37.

Hotelling, Harold, "The Economics of Exhaustible Resources," *Journal of Political Economy*, 1931, *39* (2), 137–175.

Marshall, Alfred, *Principles of Economics*, New York: Macmillan, 1890.

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," in Paul Zarembka, ed., *Fronteirs in Econometrics*, New York: Academic Press, 1974, pp. 105–142.

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

_ **and Gianmarco I. P. Ottaviano**, "Market Size, Trade, and Productivity," *Review of Economic Studies*, 2008, *75* (1), 295–316.

Menon, Jayant, "Exchange Rate Pass-Through," *Journal of Economic Surveys*, 1995, *9* (2), 197–231.

Miller, Nathan H., Marc Remer, and Gloria Sheu, "Using Cost Pass-Through to Calibrate Demand," 2012. http://www.nathanhmiller.org/PTR-calibration-12Apr24.pdf.

Natalini, Pierpaolo and Biagio Palumbo, "Inequalities for the Incomplete Gamma Function," *Mathematical Inequalities and Applications*, 2000, *3* (1), 69–77.

Nevo, Aviv, "Empriical Models of Consumer Behavior," 2010. http://www.nber.org/papers/w16511.pdf.

Ottaviano, Gianmarco I. P., Takatoshi Tabuchi, and Jacques-François Thisse, "Agglomeration and Trade Revisited," *International Economic Review*, 2002, *43* (2), 409–436.

Poterba, James M., "Retail Price Reactions to Changes in State and Local Sales Taxes," *National Tax Journal*, 1996, *49* (2), 165–176.

Quint, Daniel, "Imperfect Competition with Complements and Substitutes," 2012. http://www.ssc.wisc.edu/∼dquint/papers/quint-complements-substitutes.pdf.

Reed, William J., "The Pareto Law of Incomes – An Explanation and An Extension," *Physica A*, 2003, *319* (1), 469–486.

Shaked, Avner and John Sutton, "Relaxing Price Competition Through Product Differentitation," *Review of Economic Studies*, 1982, *49* (1), 3–13.

Weyl, E. Glen, "Slutsky meets Marschak: the First-Order Identification of Multiproduct Monopoly," 2009. http://www.glenweyl.com.

_ **and Jean Tirole**, "Market Power Screens Willingness-to-Pay," 2012. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1659449.

_ **and Michal Fabinger**, "Pass-through as an Economic Tool," 2012. http://www.glenweyl.com.

Appendix

A Taxonomy of Functional Forms

Proof of Proposition 1. Note that our discussion of stretch parameters in the text implies we can ignore the scale parameter of distributions, normalizing this to 1 for any distributions which has one. A similar argument applies to position parameter: because this only shifts the values where properties apply by a constant, it cannot affect global curvature or higher-order properties. This is useful because many of the probability distributions we consider below have scale and position parameters that this fact allows us to neglect. We will denote this normalization by *Up to Scale and Position* (USP).

We begin by considering the first part of the proof, that for any shape parameter $\alpha < 1$ the Fréchet, Weibull and Gamma distributions with shape α violate DMR at some price. We show this for each distribution in turn:

1. Type II Extreme Value (Fréchet) distribution: USP, this distribution is $F(x) = e^{-x^{-\alpha}}$ with domain $x > 0$. Simple algebra shows that

$$\mu'(x) = \frac{\left(e^{x^{-\alpha}} - 1\right)x^{\alpha}(1+\alpha) - e^{x^{-\alpha}}\alpha}{\alpha}.$$

As $x \to \infty$ and therefore $x^{-\alpha} \to 0$ (as shape is always positive), $e^{x^{-\alpha}}$ is well-approximated by its first-order approximation about 0, $1 + x^{-\alpha}$. Therefore the limit of the above expression is the same as that of

$$\frac{x^{-\alpha}x^{\alpha}(1+\alpha) - e^{x^{-\alpha}}\alpha}{\alpha} = \frac{1+\alpha - e^{x^{-\alpha}}\alpha}{\alpha} \to \frac{1}{\alpha}$$

as $x \to \infty$. Clearly this is greater than 1 for $0 < \alpha < 1$ so that for sufficiently large x DMR is violated.

2. Weibull distribution: USP, this distribution is $F(x) = 1 - e^{-x^{\alpha}}$. Again algebra yields:

$$\mu'(x) = \frac{1-\alpha}{\alpha x^{\alpha}}.$$

Clearly for any $\alpha < 1$ as $x \to 0$ this expression goes to infinity, so that for sufficiently small x DMR is violated.

3. Gamma distribution: USP, this distribution is $F(x) = \frac{\gamma(\alpha,x)}{\Gamma(\alpha)}$ where $\gamma(\cdot,\cdot)$ is the lower incomplete Gamma function, $\Gamma(\cdot,\cdot)$ is the upper incomplete Gamma function and $\Gamma(\cdot)$ is the (complete) Gamma function:

$$\mu'(x) = \frac{e^{x}(1-\alpha+x)\Gamma(\alpha,x)}{x^{\alpha}} - 1.$$

By definition, $\lim_{x\to 0}\Gamma(\alpha,x)=\Gamma(\alpha)>0$ so

$$\lim_{x\to 0}\mu'(x)=+\infty$$

as $1-\alpha>0$ for $\alpha<1$. Thus clearly for small enough x, the Gamma distribution with shape $\alpha<1$ violates DMR.

We now turn to the categorization of demand functions as having increasing or decreasing pass-through. As price always increases in cost, this can be viewed as either pass-through as a function of price or pass-through as a function of cost.

1. Normal (Gaussian) distribution: USP, this distribution is given by $F(x)=\Phi(x)$, where Φ is the cumulative normal distribution function. Algebraic computations show that for this distribution

$$\rho'(x)=\frac{\big(1-\Phi[x]\big)(1+x^2)\sqrt{2\pi}e^{\frac{x^2}{2}}-x}{\left(e^{\frac{x^2}{2}}\sqrt{2\pi}x\big[1-\Phi(x)\big]-2\right)^2}. \tag{20}$$

This has the same sign as

$$\big(1-\Phi[x]\big)(1+x^2)\sqrt{2\pi}e^{\frac{x^2}{2}}-x. \tag{21}$$

This is positive if and only if

$$\chi(x)\equiv 1-\Phi(x)-\frac{x}{\sqrt{2\pi}(1+x^2)e^{\frac{x^2}{2}}}>0.$$

Note that $\lim_{x\to\infty}\chi(x)=0$. Therefore if $\chi'(x)<0$ for all x, as we show below[25], the $\chi>0$ for any finite x which establishes the result. We now show that $\chi'<0$.

$$\chi'(x)=-\Phi'(x)-\left(\frac{x}{\sqrt{2\pi}(1+x^2)e^{x^2/2}}\right)'=$$

$$-\frac{1}{\sqrt{2\pi}e^{x^2/2}}-\frac{1}{2\pi}\frac{(1+x^2)e^{x^2/2}-(x.2xe^{x^2/2}+x(1+x^2)e^{x^2/2}x)}{(1+x^2)^2e^{x^2}}=$$

$$-\frac{1}{\sqrt{2\pi}e^{x^2/2}}-\frac{1}{2\pi}\frac{(1+x^2)-(2x^2+(1+x^2)x^2)}{(1+x^2)^2e^{x^2/2}}=$$

$$-\frac{1}{\sqrt{2\pi}e^{x^2/2}}\left(1+\frac{(1+x^2)-2x^2-x^2-x^4)}{(1+x^2)^2}\right)=-\frac{1}{\sqrt{2\pi}e^{x^2/2}}\frac{2}{(1+x^2)^2}<0$$

[25]The demonstration that $\chi'<0$ is due to research assistance by Rosen Kralev. Thanks again to him.

2. Logistic distribution: USP, this distribution is $F(x) = \frac{e^x}{1+e^x}$. Again algebra yields

$$\rho'(x) = \frac{e^x}{(1+e^x)^2} > 0.$$

Thus the logistic distribution exhibits increasing pass-through.

3. Type I Extreme Value (Gumbel) distribution : USP, this distribution has two forms. For the minimum version it is $F(x) = 1 - e^{-e^x}$. Algebra shows that for this distribution

$$\rho'(x) = \frac{e^x}{(1+e^x)^2}.$$

Note that this is the same as for the logistic distribution; in fact the pass-through rates for the Gumbel minimum distribution are identical to the logistic distribution. This is not surprising given the close connection between these distributions (McFadden, 1974).

For the maximum version it is $F(x) = e^{-e^{-x}}$. Again algebra yields

$$\rho'(x) = \frac{e^{-x}\left(e^{e^{-x}}[1+e^x] + e^{2x}[e^{e^{-x}} - 1]\right)}{\left(1 + e^{e^{-x}} + e^x - e^{e^{-x}+x}\right)^2}.$$

But clearly $e^{-x} > 0$ so $e^{e^{-x}} > 1$ and therefore the numerator and the entire expression is greater than 0 and the demand function generated by the Gumbel distribution therefore exhibits increasing pass-through.

4. Laplace distribution: USP, this distribution is

$$F(x) = \begin{cases} 1 - \frac{e^{-x}}{2} & x \geq 0, \\ \frac{e^x}{2} & x < 0. \end{cases}$$

For $x > 0$, $\rho = 1$ (so in this range pass-through is not strictly increasing). For $x < 0$

$$\rho'(x) = \frac{2e^x}{(2+e^x)^2} > 0.$$

So the Laplace distribution exhibits globally weakly increasing pass-through, strictly increasing for prices below the mode. The pass-through rate for this distribution is $\frac{e^x}{2+e^x}$ as opposed to $\frac{e^x}{1+e^x}$ for Gumbel and Logistic. However these are very similar, again pointing out the similarities among pass-through functions assumed by common demand forms.

5. Type II Extreme Value (Fréchet) distribution with shape $\alpha > 1$: From the formula above it is easy to show that the derivative of the pass-through rate is

$$\rho'(x) = \frac{x^{-(1+\alpha)}\alpha^2\Big([1+\alpha]\big[x^{2\alpha}(e^{x^{-\alpha}} - 1) + e^{x^{-\alpha}}x^\alpha\big] + \alpha e^{x^{-\alpha}}\Big)}{\left(\alpha[1 + e^{x^{-\alpha}}] - [e^{x^{-\alpha}}] - 1]x^\alpha(1+\alpha)\right)^2} > 0$$

as $x > 0$ in the range of this demand function and $e^x > 1$ for positive x. Thus this distribution, as well, exhibits increasing pass-through.

6. Type III Extreme Value (Reverse Weibull) distribution: USP, this distribution is $F(x) = e^{-(-x)^\alpha}$ and has support $x < 0$. Algebra shows

$$\rho'(x) = (-x)^{\alpha-1}\alpha^2 \frac{1-\alpha+e^{(-x)^\alpha}\Big([1-\alpha]\big[(-x)^\alpha - 1\big] + [-x]^{2\alpha}\alpha\Big)}{\Big(\alpha - 1 + [-x]^\alpha\alpha + e^{[-x]^\alpha}\Big[1+\big([-x]^\alpha - 1\big)\alpha\Big]\Big)^2},$$

which has the same sign as

$$1-\alpha+e^{(-x)^\alpha}\Big([1-\alpha]\big[(-x)^\alpha - 1\big] + [-x]^{2\alpha}\alpha\Big). \tag{22}$$

Note that the limit of this expression as $x \to 0$ is

$$1-\alpha-(1-\alpha) = 0$$

and its derivative is

$$\frac{e^{(-x)^\alpha}(-x)^{2\alpha}\alpha\big(1+\alpha+[-x]^\alpha\alpha\big)}{x},$$

which is clearly strictly negative for $x < 0$. Thus expression (22) is strictly decreasing and approaches 0 as x approaches 0. It is therefore positive for all negative x, showing that again in this case $\rho' > 0$.

7. Weibull distribution with shape $\alpha > 1$: As with the Fréchet distribution algebra from the earlier formula shows

$$\rho'(x) = \frac{x^{\alpha-1}(\alpha-1)\alpha^2}{(\alpha-1+x^\alpha\alpha)^2},$$

which is clearly positive for $\alpha > 1$ as the range of this distribution is positive x. Thus the Weibull distribution with $\alpha > 1$ exhibits increasing pass-through.

8. Gamma distribution with shape $\alpha > 1$: Again using the formula calculated above for μ', a bit of algebra and a derivative yield:

$$\rho'(x) = \frac{\alpha-1-x+\frac{e^x}{x^\alpha}\big(x^2-2x[\alpha-1]+[\alpha-1]\alpha\big)\Gamma(\alpha,x)}{x\Big(\frac{e^x}{x^\alpha}[1+x-\alpha]\Gamma[\alpha,x]-2\Big)^2}.$$

Because the Gamma distribution is only defined for positive x, this has the same sign as

$$\alpha-1-x+\frac{e^x}{x^\alpha}\big(x^2+[\alpha-2x][\alpha-1]\big)\Gamma(\alpha,x). \tag{23}$$

Note that as long as $\alpha > 1$

$$x^2+(\alpha-2x)(\alpha-1) = x^2-2(\alpha-1)x+\alpha(\alpha-1) > x^2-2(\alpha-1)x+(\alpha-1)^2 = (x+1-\alpha)^2 > 0.$$

Therefore so long as $x \leq \alpha - 1$ this is clearly positive. On the other hand when $x > \alpha - 1$ the proof depends on the following result of Natalini and Palumbo (2000):

Theorem (Natalini and Palumbo, 2000). *Let a be a positive parameter, and let q(x) be a function, differentiable on* $(0, \infty)$ *, such that* $lim_{x\to\infty} x^\alpha e^{-x} q(x, \alpha) = 0$*. Let*

$$T(x, \alpha) = 1 + (\alpha - x)q(x, \alpha) + x\frac{\partial q}{\partial x}(x, \alpha).$$

If $T(x, \alpha) > 0$ *for all* $x > 0$ *then* $\Gamma(\alpha, x) > x^\alpha e^{-x} q(x, \alpha)$*.*

Letting

$$q(x, \alpha) \equiv \frac{x - (\alpha - 1)}{x^2 + (\alpha - 2x)(\alpha - 1)},$$

$$T(x, \alpha) = \frac{2(\alpha - 1)x}{\left(\alpha^2 + x[2 + x] - \alpha[1 + 2x]\right)^2} > 0$$

for $\alpha > 1, x > 0$. So $\Gamma(\alpha, x) > x^\alpha e^{-x} q(x, \alpha)$. Thus expression (23) is strictly greater than

$$\alpha - 1 - x + x - (\alpha - 1) = 0$$

as, again, $x^2 + (\alpha - 2x)(\alpha - 1) > 0$. Thus again $\rho' > 0$.

This establishes the second part of the proposition. Turning to my final two claims, algebra shows that the pass-through rate for the Fréchet distribution is

$$\rho(x) = \frac{\alpha}{\alpha + e^{x^{-\alpha}}\left(\alpha - x^\alpha[1 + \alpha]\right) + x^\alpha(1 + \alpha)} = \frac{\alpha}{\alpha(1 + e^{x^{-\alpha}}) - (e^{x^{-\alpha}} - 1)x^\alpha(1 + \alpha)}.$$

Note for any $\alpha > 1$ this is clearly continuous in $x > 0$. Now consider the first version of the expression. Clearly as $x \to 0$, $x^\alpha \to 0$ and $e^{x^{-\alpha}} \to \infty$ so the denominator goes to ∞ and the expression goes to 0. So for sufficiently small $x > 0$, $\rho(x) < 1$ and demand is cost-absorbing. On the other consider the second version of the expression. Its denominator is

$$\alpha(1 + e^{x^{-\alpha}}) - (e^{x^{-\alpha}} - 1)x^\alpha(1 + \alpha).$$

By the same argument as above with the Fréchet distribution the limit of the above expression as $x \to \infty$ is the same as that of

$$\alpha(1 + e^{x^{-\alpha}}) - x^{-\alpha}x^\alpha(1 + \alpha) = \alpha(1 + e^{x^{-\alpha}}) - 1 - \alpha \to \alpha - 1$$

as $x \to \infty$. Thus

$$\lim_{x\to\infty}\rho(x)=\frac{\alpha}{\alpha-1}>1$$

and thus for sufficiently large x and any $\alpha>1$, this distribution exhibits cost-amplification.
Finally, consider my claim about AIDS. First note that for this demand function

$$\mu'(p)=1+\frac{b\big(a-2b+b\log[p]\big)}{\big(a-b+b\log[p]\big)^2}<1$$

as $b<0$ and $p\le e^{-\frac{a}{b}}<e^{2-\frac{a}{b}}$.

$$\rho(p)=-\left(\frac{a}{b}+\log[p]+\frac{b}{a-2b+b\log[p]}\right).$$

This is less than 1 if and only if

$$a^2+2ab\big(\log[p]-2\big)+b^2\Big(1+\log[p]\big[\log(p)-2\big]\Big)<b^2\big(2-\log[p]\big)-ab$$

or

$$\big(a+b\log[p]\big)^2-b^2\big(\log[p]+1\big)<0.$$

Clearly as $p\to 0$ the second term is positive; therefore there is always a price at which $\rho(p)>1$. On the other hand as $p\to e^{-\frac{a}{b}}$ this expression goes to

$$0-b^2\left(1-\frac{a}{b}\right)=b(a-b)<0.$$

Thus there is always a price at which $\rho(p)<1$.

$$\rho'(p)=\frac{b^2-\big(a-2b+b\log[p]\big)^2}{p\big(a-2b+b\log[p]\big)^2},$$

which has the same sign as

$$b^2-\big(a-2b+b\log[p]\big)^2<b^2-(2b)^2=-3b^2<0.$$

Thus AIDS exhibits decreasing pass-through.

$\square$

B Properties of Apt demand

Definition 1. *A function Q defined for $p\in\mathbb{R}_+$ is an* Adjustable pass-through *(Apt) demand function if*

$$Q(p)=\sigma\left[\frac{1+\overline{\mu'}\,(1+\alpha)\left(\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}-1\right)}{1+\overline{\mu'}\,(1+\alpha)\left(\sqrt{\frac{\alpha}{1+\alpha}}-1\right)}\right]^{-\frac{2}{\overline{\mu'}}} \tag{24}$$

when not otherwise specified, $m, \sigma > 0$, $\overline{\mu'} < 1$ and one of the three following sets of conditions is satisfied:

1. $\overline{\mu'} < 0$, $\alpha < -1$ *and equation (24) holds for* $p < m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{\overline{\mu'}^2(1+\alpha)}\mathbf{1}_{1>\frac{1-\overline{\mu'}}{\alpha\overline{\mu'}}} - \alpha\right)$ *while* $Q(p) = 0$ *for* $p \geq m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{\overline{\mu'}^2(1+\alpha)}\mathbf{1}_{1>\frac{1-\overline{\mu'}}{\alpha\overline{\mu'}}} - \alpha\right)$.

2. $\overline{\mu'} > 0, \alpha \geq 0$ *and equation (24) holds for* $p > \max\left\{m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)} - \alpha\right), 0\right\}$, $Q(p) = \frac{1}{a+bp}$ *for* $-\frac{a}{b} \leq p \leq m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)} - \alpha\right)$ *and* $Q(p) = \infty$ *for* $p < -\frac{a}{b}$ *where*

$$b \equiv \frac{4\left(1-\overline{\mu'}\right)^2(1+\alpha)}{m\sigma\left[1-\overline{\mu'}(1+\alpha)\right]^2\left(2-\overline{\mu'}\right)}\left[\frac{\left[1-\overline{\mu'}(1+\alpha)\right]\left(2-\overline{\mu'}\right)}{2\left(1-\overline{\mu'}\right)\left(1+\overline{\mu'}\left[\sqrt{\alpha(1+\alpha)}-(1+\alpha)\right]\right)}\right]^{\frac{2}{\overline{\mu'}}}$$

and

$$a \equiv \frac{1}{\sigma}\left[\frac{\left[1-\overline{\mu'}(1+\alpha)\right]\left(2-\overline{\mu'}\right)}{2\left(1-\overline{\mu'}\right)\left(1+\overline{\mu'}\left[\sqrt{\alpha(1+\alpha)}-(1+\alpha)\right]\right)}\right]^{\frac{2}{\overline{\mu'}}} - b \cdot m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)} - \alpha\right).$$

3. $\overline{\mu'} < 0 \leq \alpha$ *and $Q(p)$ follows the special cases from 1) and 2) that differ from equation (24) when p is in the regions there-specified.*

or any of the following three conditions is satisfied:

1. $Q(p) = \sigma e^{2(1+\alpha)\left(\sqrt{\frac{\alpha}{1+\alpha}}-\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}\right)}$ *for* $p < -m\alpha$ *with* $\alpha < -1$ *and* $Q(p) = 0$ *for* $p \geq -m\alpha$.

2. $Q(p) = \sigma e^{2(1+\alpha)\left(\sqrt{\frac{\alpha}{1+\alpha}}-\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}\right)}$ *for* $p > m\left(\frac{1}{4(1+\alpha)} - \alpha\right)$ *with* $\alpha \geq 0$,

$$Q(p) = \frac{2\sigma e^{2\sqrt{\alpha(1+\alpha)}-1}}{1+4(1+\alpha)\left(\frac{p}{m}+\alpha\right)}$$

for

$$m\left(\frac{1}{4(1+\alpha)} - \alpha - \frac{2(1+\alpha)e^{2\left(1-2\sqrt{\alpha(1+\alpha)}\right)}}{\sigma^2 m^2}\right) < p \leq m\left(\frac{1}{4(1+\alpha)} - \alpha\right)$$

and $Q(p) = \infty$ *for* $m\left(\frac{1}{4(1+\alpha)} - \alpha - \frac{2(1+\alpha)e^{2\left(1-2\sqrt{\alpha(1+\alpha)}\right)}}{\sigma^2 m^2}\right) \geq p$.

3. $Q(p) = \sigma e^{-\frac{p}{m}}$.

These last cases are interpreted to have $\overline{\mu'} = 0$ and in the last case $\alpha = 0$ as well.

Proof of Proposition 2. 1. *Positivity*: Clearly $0, \infty \geq 0$ so we need not consider the constant portions of any form of Apt demand. In the other regions, first note that $\sqrt{\frac{\alpha}{1+\alpha}}$ and $\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}$ are always real numbers because either $\alpha > 0$ so both factors are positive or $\alpha < -1$ and $\frac{p}{m} < -\alpha$. For the rest, we go case-by-case:

- The sign of the whole expression in (24) is determined by that inside the parentheses when $\overline{\mu'} \neq 0$. When $\overline{\mu'} \cdot \alpha \geq 0$, given that either $\alpha \geq 0$ or $\alpha < -1$,

$$1 + \overline{\mu'}\left(1+\alpha\right)\left[\sqrt{\frac{\alpha}{1+\alpha}} - 1\right] > 0.$$

  The numerator is also positive by the same argument.

- When $\overline{\mu'} < 0 \leq \alpha$, the denominator is positive as $\sqrt{\frac{\alpha}{1+\alpha}} < 1$. To see that the numerator is positive note that when $p \leq m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{\overline{\mu'}^2(1+\alpha)} - \alpha\right)$,

$$1 + \overline{\mu'}\left(1+\alpha\right)\left[\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}} - 1\right] \geq 1 + \overline{\mu'}\left(1+\alpha\right)\left[\sqrt{\frac{\left[1-\overline{\mu'}\left(1+\alpha\right)\right]^2}{\overline{\mu'}^2\left(1+\alpha\right)^2}} - 1\right] =$$

$$1 + \overline{\mu'}\left[\frac{\overline{\mu'}\left(1+\alpha\right) - 1}{\overline{\mu'}} - 1 - \alpha\right] = 0.$$

- When $\overline{\mu'} = 0$ or $Q(p) = \sigma e^{-\frac{p}{m}}$ demand is positive by the exponential form.
- Outside of the main range, when demand takes the form $\frac{1}{a+bp}$, positivity follows from the fact that $p > -\frac{a}{b}$ and b is positive in both cases by inspection.

2. *Monotonicity*: Over the main range described by equation (24), we can compute the derivative of demand (which exists by the next item) as

$$-\sigma\frac{2}{\overline{\mu'}}\frac{\overline{\mu'}(1+\alpha)}{2\sqrt{\left(\frac{p}{m}+\alpha\right)(1+\alpha)}\left(1+\overline{\mu'}\left[\sqrt{\alpha(1+\alpha)}-1-\alpha\right]\right)}\left(\frac{1+\overline{\mu'}\left[\sqrt{\left(\frac{p}{m}+\alpha\right)(1+\alpha)}-(1+\alpha)\right]}{1+\overline{\mu'}\left[\sqrt{\alpha(1+\alpha)}-(1+\alpha)\right]}\right)^{-\left(\frac{2}{\overline{\mu'}}+1\right)}.$$

By the same logic as in the proof of positivity, the last factor in this expression and $\sqrt{\left(\frac{p}{m}+\alpha\right)\left(1+\alpha\right)}$ are both positive and thus the sign of the whole expression is the same as that of

$$-\frac{1+\alpha}{1+\overline{\mu'}\left[\sqrt{\alpha\left(1+\alpha\right)}-1-\alpha\right]}.$$

Reviewing the arguments from the proof of positivity, note that the sign of $1+\alpha$ is always the same as that of $1 + \overline{\mu'}\left[\sqrt{\alpha\left(1+\alpha\right)} - 1 - \alpha\right]$, establishing monotonicity.

In the non-core region, the result follows by the positivity of b and the positivity of demand. For $\overline{\mu'} = 0$ the derivative of demand is

$$-\frac{\sigma\left|1+\alpha\right|Q(p)}{m\sqrt{\left(\frac{p}{m}+\alpha\right)(1+\alpha)}} < 0.$$

3. *Differentiability*: First note that away from functional break points, continuity and infinite differentiability are immediate, because the functions are analytic, so long as $\frac{p}{m}$ never moves to the other side of $-\alpha$ and the argument that is raised to a power never changes signs. But we know neither of these occur from the proof of positivity. So we need only consider break points of the piecewise definitions. Furthermore the proposition only claims properties about demand at points where it is weakly positive and finite, so we need not consider the break points where demand is 0 above these points or ∞ below these, as demand is constructed right, rather than left, continuous at these points in the first case and left continuous at these points in the second case. This leaves only the break point between the two (finite) expressions when $\alpha > 0$. The demand functions on both sides of the break are differentiable so we must just show that their levels and derivatives match at the break point.

   We begin with levels. Evaluating expression (24) at the break point and noting that the break point arises only when $\alpha \leq \frac{1-\overline{\mu'}}{\overline{\mu'}}$ yields

$$\sigma\left(\frac{1+\overline{\mu'}\left[\sqrt{\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2}}-1-\alpha\right]}{1+\overline{\mu'}\left[\sqrt{\alpha\left(1+\alpha\right)}-1-\alpha\right]}\right)^{-\frac{2}{\overline{\mu'}}} = \sigma\left(\frac{1+\overline{\mu'}\left[\frac{1-\overline{\mu'}(1+\alpha)}{2\left(1-\overline{\mu'}\right)}-1-\alpha\right]}{1+\overline{\mu'}\left[\sqrt{\alpha\left(1+\alpha\right)}-1-\alpha\right]}\right)^{-\frac{2}{\overline{\mu'}}} =$$

$$\sigma\left(\frac{\left(2-\overline{\mu'}\right)\left[1-\overline{\mu'}\left(1+\alpha\right)\right]}{1+\overline{\mu'}\left[\sqrt{\alpha\left(1+\alpha\right)}-1-\alpha\right]}\right)^{-\frac{2}{\overline{\mu'}}}.$$

   Note that this is just $\frac{1}{a+b\cdot m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)}-\alpha\right)}$ and thus the levels of demand match at the break point. The value of μ under formula (24) at the break point is given by evaluating expression (3) at the break point:

$$m\left(\overline{\mu'}\frac{\left[1-\overline{\mu'}\left(1+\alpha\right)\right]^2}{4\left(1-\overline{\mu'}\right)^2\left(1+\alpha\right)}+\left[1-\overline{\mu'}\left(1+\alpha\right)\right]\sqrt{\frac{\left[1-\overline{\mu'}\left(1+\alpha\right)\right]^2}{4\left(1-\overline{\mu'}\right)^2\left(1+\alpha\right)^2}}\right) =$$

$$\frac{m\left[1-\overline{\mu'}\left(1+\alpha\right)\right]^2}{4\left(1-\overline{\mu'}\right)^2\left(1+\alpha\right)}\left[\overline{\mu'}+2\left(1-\overline{\mu'}\right)\right] = \frac{m\left(2-\overline{\mu'}\right)\left[1-\overline{\mu'}\left(1+\alpha\right)\right]^2}{4\left(1-\overline{\mu'}\right)^2\left(1+\alpha\right)}.$$

The value of μ for $Q(p) = \frac{1}{a+bp}$ is

$$-\frac{\frac{1}{a+bp}}{-\frac{b}{(a+bp)^2}} = \frac{a}{b} + p = \frac{m\left(2-\overline{\mu'}\right)\left[1-\overline{\mu'}\left(1+\alpha\right)\right]^2}{4\left(1-\overline{\mu'}\right)^2\left(1+\alpha\right)}.$$

Finally, we compare the second derivatives. Clearly $\mu' = 1$ below the break point. Above it

$$\mu' = \overline{\mu'} + \frac{1-\overline{\mu'}\left(1+\alpha\right)}{2\left(1+\alpha\right)\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}}.$$

Evaluating this at the break point yields

$$\overline{\mu'} + \frac{1-\overline{\mu'}\left(1+\alpha\right)}{\left(1+\alpha\right)\sqrt{\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)^2}}} = \overline{\mu'} + 1 - \overline{\mu'} = 1.$$

Thus the level, first and second derivatives match, implying that the function is twice continuously differentiable at the break point. An essentially identical argument applies in the corresponding $\overline{\mu'} = 0$ case at the break point, so we omit it for the sake of brevity.

4. *Declining marginal revenue*: Over the range of values where equation (24) applies, the above analysis indicates that

$$\mu' = \overline{\mu'} + \frac{1-\overline{\mu'}\left(1+\alpha\right)}{2\left(1+\alpha\right)\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}}. \tag{25}$$

When $\left[1-\overline{\mu'}\left(1+\alpha\right)\right]\left(1+\alpha\right) < 0$ this expression is no greater than $\overline{\mu'} < 1$ by our parametric restrictions. When $\left[1-\overline{\mu'}\left(1+\alpha\right)\right]\left(1+\alpha\right) > 0$ there are two possibilities. If $\alpha > 0$, the formula applies only if $p > m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)} - \alpha\right)$, in which case expression (25) is less than

$$\overline{\mu'} + \frac{1-\overline{\mu'}\left(1+\alpha\right)}{2\left(1+\alpha\right)\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}} < \overline{\mu'} + \frac{1-\overline{\mu'}\left(1+\alpha\right)}{\left(1+\alpha\right)\sqrt{\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)^2}}} = 1.$$

by the logic of the proof of differentiability. The same logic applies in the $\overline{\mu'} = 0$ case. The other possibility is that $\alpha < \frac{1-\overline{\mu'}}{\overline{\mu'}} < 0$. In this case $p < m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{\overline{\mu'}^2(1+\alpha)}\mathbf{1}_{1>\frac{1-\overline{\mu'}}{\alpha\overline{\mu'}}} - \alpha\right) \implies$

$$\overline{\mu'} + \frac{1-\overline{\mu'}\left(1+\alpha\right)}{2\left(1+\alpha\right)\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}} < \overline{\mu'} + \frac{1-\overline{\mu'}\left(1+\alpha\right)}{2\left(1+\alpha\right)\sqrt{\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{\overline{\mu'}^2(1+\alpha)^2}}} = \overline{\mu'} + \frac{1-\overline{\mu'}\left(1+\alpha\right)}{-2\left(1+\alpha\right)\frac{1-\overline{\mu'}(1+\alpha)}{\overline{\mu'}(1+\alpha)}} = \frac{\overline{\mu'}}{2} < \frac{1}{2}.$$

From the analysis above, in the alternate range $\mu' = 1$, completing the proof.

$\square$

Proof of Proposition 3. First we prove the log-curvature results, relying on the identification of the sign of μ' with positive v. negative log-curvature. From above we have that

$$\mu' = \overline{\mu'} + \frac{1 - \overline{\mu'}(1+\alpha)}{2(1+\alpha)\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}}.$$

When $\overline{\mu'}, \alpha < 0$ and $\alpha \geq \frac{1-\overline{\mu'}}{\overline{\mu'}}$, the result holds by the global negativity of the second term, as in the case when $\alpha > 0$ and $\alpha < \frac{1-\overline{\mu'}}{\overline{\mu'}}$ it holds by global positivity of the second term. When $\overline{\mu'}, \alpha < 0$ and $\alpha < \frac{1-\overline{\mu'}}{\overline{\mu'}}$, we have from the last part of the proof of Proposition 2 that $\mu' \leq \frac{\overline{\mu'}}{2} < 0$. When $\overline{\mu'} > 0$ and $\alpha > \frac{1-\overline{\mu'}}{\overline{\mu'}}$

$$\overline{\mu'} + \frac{1 - \overline{\mu'}(1+\alpha)}{2(1+\alpha)\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}} > \overline{\mu'} + \frac{1 - \overline{\mu'}(1+\alpha)}{2(1+\alpha)\sqrt{\frac{\alpha}{1+\alpha}}} = \frac{1 + \overline{\mu'}(1+\alpha)\left(2\sqrt{\frac{\alpha}{1+\alpha}} - 1 - \alpha\right)}{2(1+\alpha)\sqrt{\frac{\alpha}{1+\alpha}}}$$

which has, strictly, the same sign as

$$1 + \overline{\mu'}\left(2\sqrt{\alpha(1+\alpha)} - 1 - \alpha\right). \tag{26}$$

Note that this is strictly increasing in α, as we can see by taking its derivative

$$\overline{\mu'}\left(\frac{1+\alpha+\alpha}{\sqrt{\alpha(1+\alpha)}} - 1\right) = \frac{\overline{\mu'}}{\sqrt{\alpha(1+\alpha)}}\left(\sqrt{(1+2\alpha)^2} - \sqrt{\alpha(1+\alpha)}\right) > 0.$$

So because $\alpha > 0$, expression (26) is at least

$$1 + \overline{\mu'}(-1) = 1 - \overline{\mu'} > 0.$$

The result follows directly from the above formula for the case of $\overline{\mu'} = 0$. For $Q(p) = \sigma e^{-\frac{p}{m}}$, demand is globally log-linear (and therefore has $\mu' = \mu'' = 0$ everywhere) because the exponential function is log-linear. Taking the derivatives of μ' from equation (25) yields

$$-\frac{1 - \overline{\mu'}(1+\alpha)}{4m(1+\alpha)^2\left(\frac{\frac{p}{m}+\alpha}{1+\alpha}\right)^{\frac{3}{2}}},$$

which clearly has the same sign as

$$\overline{\mu'}(1+\alpha) - 1.$$

$\square$

Proposition 5. *When $\overline{\mu'} \neq 0$, Apt demand has consumer surplus*

$$CS\left(p\right) = V(p) \equiv \frac{m}{\left(1+\alpha\right)\left(2-\overline{\mu'}\right)\left(1-\overline{\mu'}\right)} \cdot$$

$$\left[\left(\left[1-\overline{\mu'}(1+\alpha)\right]\left[2(1+\alpha)\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}+1-\overline{\mu'}(1+\alpha)\right]+(1+\alpha)\overline{\mu'}\left(2-\overline{\mu'}\right)\left(\frac{p}{m}+\alpha\right)\right)Q(p)-\sigma\left(\frac{1-\overline{\mu'}(1+\alpha)}{1+\overline{\mu'}(1+\alpha)\left(\sqrt{\frac{\alpha}{1+\alpha}}-1\right)}\right)^{-\frac{2}{\overline{\mu'}}}\left[1-\overline{\mu'}(1+\alpha)\right]^2\mathbf{1}_{1<\frac{1-\overline{\mu'}}{\alpha\overline{\mu'}}}\cdot\mathbf{1}_{\overline{\mu'}<0}\right].$$

except when $\alpha < 0$ and $p \geq m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{\overline{\mu'}^2(1+\alpha)}\mathbf{1}_{1>\frac{1-\overline{\mu'}}{\alpha\overline{\mu'}}}-\alpha\right)$ in which case $CS(p)=0$, $\alpha \geq 0$ and $-\frac{a}{b} < p < m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)}-\alpha\right)$, in which case it has surplus

$$CS\left(p\right) \equiv V\left(m\left(\frac{\left[1-\overline{\mu'}\left(1+\alpha\right)\right]^2}{4\overline{\mu'}^2\left(1+\alpha\right)}-\alpha\right)\right)+\frac{1}{b}\log\left(\frac{a+bm\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)}-\alpha\right)}{a+bp}\right)$$

or $p \leq -\frac{a}{b}$ in which case surplus is ∞. When $\overline{\mu'} = 0$ it has surplus

$$CS(p) = W\left(p\right) \equiv \frac{m\left(\left[1+2\left(1+\alpha\right)\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}\right]Q(p)-\sigma e^{2(1+\alpha)\sqrt{\frac{\alpha}{1+\alpha}}}\mathbf{1}_{\alpha<0}\right)}{2\left(1+\alpha\right)}$$

except when $\alpha < -1$ and $p \geq -m\alpha$ in which case $CS(p)=0$, $\alpha \geq 0$ and

$$m\left(\frac{1}{4(1+\alpha)}-\alpha-\frac{2(1+\alpha)e^{2\left(1-2\sqrt{\alpha(1+\alpha)}\right)}}{\sigma^2m^2}\right) < p \leq m\left(\frac{1}{4(1+\alpha)}-\alpha\right)$$

in which case it has surplus

$$CS(p) = W\left(m\left[\frac{1}{4(1+\alpha)}-\alpha\right]\right)-\log\left(e^{2\left(1-2\sqrt{\alpha(1+\alpha)}\right)}+\frac{\sigma^2m^2}{2(1+\alpha)}\left(\frac{p}{m}+\alpha-\frac{1}{4(1+\alpha)}\right)\right)$$

or $p \leq m\left(\frac{1}{4(1+\alpha)}-\alpha-\frac{2(1+\alpha)e^{2\left(1-2\sqrt{\alpha(1+\alpha)}\right)}}{\sigma^2m^2}\right)$ in which case surplus is ∞. Finally if $Q(p) = \sigma e^{-\frac{p}{m}}$ then surplus is $\sigma m e^{-\frac{p}{m}}$.

Proof. Surplus is given by the standard formula $CS(p) = \int_p^\infty Q(p)dp$. For $\overline{\mu'} < 0$, demand is 0 for $p \geq m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{\overline{\mu'}^2(1+\alpha)}\mathbf{1}_{1>\frac{1-\overline{\mu'}}{\alpha\overline{\mu'}}}-\alpha\right)$ so $CS(p)=0$ over this range as well. Standard integration techniques give that the indefinite integral of the formula in equation (24) is

$$U(p) \equiv -\frac{m}{(1+\alpha)\left(2-\overline{\mu'}\right)\left(1-\overline{\mu'}\right)}\left(\left[1-\overline{\mu'}(1+\alpha)\right]\left[2(1+\alpha)\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}+1-\overline{\mu'}(1+\alpha)\right]+(1+\alpha)\overline{\mu'}\left(2-\overline{\mu'}\right)\left(\frac{p}{m}+\alpha\right)\right)Q(p),$$

where $Q(p)$ is given by expression (24). When $\overline{\mu'} < 0, \alpha < 0$, demand takes the form from equation (24) over the full range until $p = m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{\overline{\mu'}^2(1+\alpha)}\mathbf{1}_{1>\frac{1-\overline{\mu'}}{\alpha\overline{\mu'}}} - \alpha\right)$. Thus CS in these cases is $W\left(m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{\overline{\mu'}^2(1+\alpha)}\mathbf{1}_{1>\frac{1-\overline{\mu'}}{\alpha\overline{\mu'}}} - \alpha\right)\right) - U(p)$. Noting that when $1 \geq \frac{1-\overline{\mu'}}{\alpha\overline{\mu'}}$ expression (24) is equal to 0 at $\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{\overline{\mu'}^2(1+\alpha)} - \alpha\right)$ and that when $1 < \frac{1-\overline{\mu'}}{\alpha\overline{\mu'}}$ all terms involving $\frac{p}{m} + \alpha$ become 0 at $-m\alpha$ so it simplifies to the expression desired.

When $\alpha > 0$, the same argument applies above the break point where $p = m\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)} - \alpha\right)$, as $W(p) \to 0$ as $p \to \infty$ given that $\mu' < 1$ and $Q(p)$ goes at a rate of $p^{-\frac{1}{\mu'}}$ to 0, which exceeds the at-most-linear rate of the factor multiplying $Q(p)$ in the expression for W. Below this break point, $Q(p) = \frac{1}{a+bp}$ which has indefinite integral $\frac{\log(a+bp)}{b}$. So surplus is

$$-V\left(m\left(\frac{\left[1-\overline{\mu'}\left(1+\alpha\right)\right]^2}{4\left(1-\overline{\mu'}\right)^2\left(1+\alpha\right)} - \alpha\right)\right) + \frac{\log\left(\frac{a+bm\left(\frac{\left[1-\overline{\mu'}(1+\alpha)\right]^2}{4\left(1-\overline{\mu'}\right)^2(1+\alpha)}-\alpha\right)}{a+bp}\right)}{b},$$

as reported in the proposition statement.

When $\overline{\mu'} = 0$, integrating $\sigma e^{2(1+\alpha)\left(\sqrt{\frac{\alpha}{1+\alpha}}-\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}\right)}$ yields

$$-\frac{m\left[1+2\left(1+\alpha\right)\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}\right]e^{2(1+\alpha)\left(\sqrt{\frac{\alpha}{1+\alpha}}-\sqrt{\frac{\frac{p}{m}+\alpha}{1+\alpha}}\right)}}{2\left(1+\alpha\right)}.$$

The same logic as in the proof of the $\overline{\mu'} \neq 0$ establishes from here that the formulas desired hold. We omit repetition of these derivations for brevity.

In the case when $Q(p) = \sigma e^{-\frac{p}{m}}$ taking the indefinite integral yields $-\sigma m e^{-\frac{p}{m}}$ and thus the formula desired.

□

C Flexibility and Estimation of Apt Demand

To reduce notation, we refer to $\left(q^\star, p^\star, \mu\left(p^\star\right), \mu'\left(p^\star\right), \mu''\left(p^\star\right)\right)$ as $\left(q, p, \mu, \mu', \mu''\right)$ in this appendix. Additionally, to simplify the exposition we often adopt the notation of Section 4 and 5 . In this notation the Apt form is given by

$$p(q) = A_0 + A_1 q^{\theta-1} + A_2 q^{2(\theta-1)},$$

where $A_i, i = 0, 1, 2$ are defined in terms of the more primitive parameters in equations (9). Then

$$\mu(p) = -\frac{q(p)}{q'(p)} = -p'(q)q.$$

If we define $x \equiv q^{\theta-1}$ and $a = \frac{2A_2}{A_1}$ then

$$\frac{dp}{dx} = A_1 (1 + ax)$$

so that,

$$\mu = (1 - \theta) A_1 (1 + ax) x, \tag{27}$$

$$\mu' = \frac{\frac{d\mu}{dx}}{\frac{dp}{dx}} = (1 - \theta) \frac{1 + 2ax}{1 + ax} = (1 - \theta) \left(2 - \frac{1}{1 + ax}\right) \tag{28}$$

and

$$\mu'' = \frac{\frac{d\mu'}{dx}}{\frac{dp}{dx}} = (1 - \theta) \frac{a}{A_1 (1 + ax)^3}. \tag{29}$$

Thus we obtain that

$$\frac{\mu''\mu}{(\mu')^2} = \frac{ax}{(1 + 2ax)^2}, \tag{30}$$

an equation that will be crucial both in the following proof and in the identification of Apt demand.

Proof of Proposition 4. First, let us maximize expression (30) over $\gamma \equiv ax$. The first derivative of this expression with respect to γ is

$$\frac{(1 + 2\gamma)^2 - 4 (1 + 2\gamma) \gamma}{(1 + 2\gamma)^4} = \frac{1 + 4\gamma + 4\gamma^2 - 4\gamma - 8\gamma^2}{(1 + 2\gamma)^4} = \frac{1 - 4\gamma^2}{(1 + 2\gamma)^4}.$$

Equating this to 0 yields $4\gamma^2 = 1 \implies \gamma = \frac{1}{2}$. This is the unique critical point but also clearly a global maximizer as the expression is negative for lower values of γ and approaches 0 for values of γ approaching infinity while for $\gamma = \frac{1}{2}$, expression (30) becomes

$$\frac{1}{2 (2)^2} = \frac{1}{8}.$$

Thus, as discussed in the text, Apt demand disallows $\frac{\mu''\mu}{(\mu')^2}$ from exceeding $\frac{1}{8}$. It remains to be shown that any values obeying these bounds and the additional bound in the proposition statement, as well as the condition that $\overline{\mu}' < 1, q, \mu, p > 0$, may be obtained.

To do this, note that to match arbitrarily low values of $\frac{\mu''\mu}{(\mu')^2}$ or values as high as the global maximum of $\frac{1}{8}$, γ needs have domain $\left(-\frac{1}{2}, \frac{1}{2}\right]$ as, by the above analysis, the expression (30) is monotone increasing and continuous over this range and clearly approaches $-\infty$ as

$\gamma \to -\frac{1}{2}$. Next note that, employing the definitions from equations (9),

$$ax = \frac{2A_1}{A_2 q^{1-\theta}} = \frac{\left(\frac{\sigma}{q}\right)^{\frac{\overline{\mu'}}{2}} \left(\frac{1}{\overline{\mu'}(1+\alpha)} + \sqrt{\frac{\alpha}{1+\alpha}} - 1\right)}{1 - \frac{1}{\overline{\mu'}(1+\alpha)}} \equiv (s)^{\frac{\overline{\mu'}}{2}} \beta.$$

Now using equation (28) we obtain that

$$\mu' = \frac{\overline{\mu'}}{2}\left(2 - \frac{1}{1+\gamma}\right) \implies \overline{\mu'} = 2\mu' \frac{1+\gamma}{1+2\gamma} \tag{31}$$

and thus from the preceding logic

$$\frac{\gamma}{s^{\mu' \frac{1+\gamma}{1+2\gamma}}} = \beta.$$

Thus it suffices to show that for relevant values of μ' and $\gamma \in \left(-\frac{1}{2}, \frac{1}{2}\right]$, β can be chosen to match the left hand side of this equation.

We do this in four cases depending on whether we want to match positive or negative values of μ'' and positive or negative values of μ'. We return to the knife-edge cases of $\mu'' = 0$ at the end; $\mu' = 0$ is not by the proposition by supposition. Note that $s > 1$ everywhere that prices and quantities are positive so we can restrict attention to $s > 1$.

1. $\mu'', \mu' > 0$: In this case it suffices to show that we can match γ values in $(0, \infty)$. When γ takes these values, $\mu' \frac{1+2\gamma}{1+\gamma} > 0$ so, given that $s > 1$, $0 \leq \frac{\gamma}{s^{\mu' \frac{1+2\gamma}{1+\gamma}}} \leq \gamma$. It is thus sufficient to show that β can match any value in $(0, \infty]$. This is clearly possible as when $\alpha \to \infty$, $\beta \to 0$ while when $\alpha \to \frac{1-\overline{\mu'}}{\overline{\mu'}}$ from above $\beta \to \infty$ and it is continuous in between, establishing that any value can be fit in this case.

   However, there is an additional restriction on the values of γ we can match in this case: we need $\overline{\mu'} < 1$ which implies, by equation (31) that

   $$1 \geq 2\mu' \frac{1+\gamma}{1+2\gamma} \implies 1 + 2\gamma \geq 2\mu' + 2\mu'\gamma \implies \gamma \geq \frac{2\mu' - 1}{2(1-\mu')}.$$

   Because the smallest necessary value of γ to achieve the full range of values is $\frac{1}{2}$ this binds only when

   $$\frac{2\mu' - 1}{2(1-\mu')} \geq \frac{1}{2} \implies \frac{2\mu' - 1}{1 - \mu'} \geq 1 \implies 2\mu' - 1 \geq 1 - \mu' \implies \mu' \geq \frac{2}{3}.$$

   When this holds the greatest value of of $\frac{\mu''\mu}{(\mu')^2}$ that is achievable is when γ satisfied the bound with equality yielding

   $$\frac{\mu''\mu}{(\mu')^2} = \frac{\gamma}{(1+2\gamma)^2} = \frac{\frac{2\mu'-1}{2(1-\mu')}}{\left(1 + 2\frac{2\mu'-1}{2(1-\mu')}\right)^2} = \frac{(2\mu'-1)(1-\mu')}{(1-\mu'+2\mu'-1)^2} = \frac{(2\mu'-1)(1-\mu')}{(\mu')^2},$$

establishing the lower bound in the proposition and the flexibility result, given that this is the only additional restriction on these parameters. And, as in the last case, $\overline{\mu'} < 0$ and thus there are no additional restrictions on it.

2. $\mu'' > 0 > \mu'$: Again it suffices to match γ values in $\left(0, \frac{1}{2}\right]$ and again $\frac{\gamma}{s^{\mu' \frac{1+2\gamma}{1+\gamma}}} > 0$, though because $\mu' < 0$ we now need to match β to any strictly positive value. However the same argument as in the previous case holds because again the full range between 0 and ∞ is mapped out as $\alpha \to -\infty$ and $\frac{1-\overline{\mu'}}{\overline{\mu'}}$ from below when $\overline{\mu'} < 0$, which clearly holds in this case. Because $\overline{\mu'} < 0$ and there is are no bounds on $\overline{\mu'}$ so long as it is negative there are no further restrictions and flexibility is established in this case.

3. $0 > \mu'', \mu'$: Now it suffices to match γ values in $\left(-\frac{1}{2}, 0\right)$. In this case $\overline{\mu'} < 0$ and we must therefore match any strictly negative value for β. For $\beta < -1$ we chose $\alpha < -1$. As α approaches $\frac{1-\overline{\mu'}}{\overline{\mu'}}$ from above, $\beta \to -\infty$ for any $\overline{\mu'}$; as $\alpha \to -1$, $\beta \to -1$. Because β is continuous in this range, we can match any value of $\beta < -1$. For $\beta \in [-1, 0)$, we match using $\alpha \geq 0$. When $\alpha = 0$, $\beta = -1$. When $\alpha \to \infty$ and $\overline{\mu'} < 0$, $\beta \to 0$ from below. Again β is continuos in α over this range so we can match any negative value of β.

4. $\mu' > 0 > \mu''$: We will focus here on the case when $\gamma < -1$. Then note that automatically $\overline{\mu'} < 1$ as $2\frac{1+\gamma}{1+2\gamma} < 1$ as $2 + 2\gamma > 1 + 2\gamma$ and both are < 0. But if $\gamma < -1$ then $\frac{\gamma}{(1+2\gamma)^2} > -1$ as, by our logic at the beginning of the proof, $\frac{\gamma}{(1+2\gamma)^2}$ decreasing in γ over this range.

   This case is incomplete as of yet and the additional restrictions impose by the necessity of $m > 0$ have not been explored fully. This characterization will appear in a future version of this paper.

Given that Apt demand can match the desired values, mapping them to the correct parameters is simple. First, we choose any valid solution for γ from the equation $\frac{\mu''\mu}{(\mu')^2} = \frac{\gamma}{(1+2\gamma)^2}$. There will always be two solutions to this equation, though the restrictions discussed above may make only one of these valid. When this is the case, choose the valid solution; otherwise either is acceptable and thus there may be two Apt parameter sets fitting a given collection (p, q, μ, μ', μ'').

Given a value of γ thus derived, equation (31) uniquely defines $\overline{\mu'}$, given μ'. Given $\overline{\mu'}$, $x = q^{-\frac{\overline{\mu'}}{2}}$. We then recover A_1, from equation (27) as

$$A_1 = \frac{2\mu}{\overline{\mu'}\,(1+\gamma)\,x}.$$

This yields A_2 from $A_2 = \frac{A_1\gamma}{2x}$. We then recover A_0 from

$$A_0 = p - A_1 x - A_2 x^2.$$

Similarly parameters in the $\left(\sigma, m, \overline{\mu'}, \alpha\right)$ space follow. $\overline{\mu'}$ is already known from above. $a = (\sigma)^{\frac{\overline{\mu'}}{2}}\beta$. We have not fully completed the identification back to the fundamental Apt

demand parameterization and will do so in a future draft. In the mean time, we include a complete treatment of the $\mu'' = 0$ (Bulow-Pfleiderer) case as this both is missing from the above and provides the reader a flavor of what will eventually appear here.

If $\mu'' = 0$ by the argument establishing Proposition 3 we must have that $\alpha = \frac{1-\overline{\mu'}}{\overline{\mu'}}$ and thus we are in the Bulow-Pfleiderer class. In this case $\mu' = \overline{\mu'}$ at every point and so simply match these parameters. To recover m we use

$$m\left(1 - \overline{\mu'} + \overline{\mu'}\frac{p}{m}\right) = \mu \implies m = \frac{\mu - \overline{\mu'}p}{1 - \overline{\mu'}}.$$

Because $\overline{\mu'} < 1$ this will be positive and thus m valid if and only if $\mu > \overline{\mu'}p$, which is always true so long as p is small enough, as discussed in the proposition statement. So long as this condition (Marshall's Second Law of Demand as discussed in Subsection 5.3) holds, m is valid and we can match the the demand as m may take on any positive value. In this case clearly demand will be strictly positive and σ may thus be adjusted to any positive number necessary to match q. Thus the only restriction is that $\mu > \mu' p = \overline{\mu'}p$ if $\mu'' = 0$. $\square$

While the identification above (in Proposition 4) establishes the range of values of (q, p, μ, μ', μ'') Apt demand can fit and a procedure for fitting the values to these, it does not provide a procedure for empirically measuring these values. Clearly in practice the appropriate procedure will depend on the type of data available and setting in which the model is applied. However, to illustrate at a high level how such a procedure would work in practice we consider the case of a symmetric oligopoly model with a constant conduct parameter θ and local exogenous variation in either the price p or the cost c (that induces variation in p) that is sufficient to estimate either one, two or three local derivatives of endogenous variables with respect to these exogenously varying quantity. The first case corresponds, for example, to the recent work of Einav et al. (2010) and Einav et al. (2012), while the second case corresponds to the work of Baker and Bresnahan (1988), Atkin and Donaldson (2012) and Miller et al. (2012) and the local derivative matching strategy here is explained in greater detail in Weyl (2009). We discuss both the case when $\theta > 0$ is known (to be monopoly, Cournot with a given number of firms, etc.) and when it is not, but for brevity, we assume the parameters derived below are consistent with our restrictions on (q, p, μ, μ', μ'') and do not discuss how these restrictions could be tested in data, though doing so would be analogous to the discussion that follows.

Suppose that only price variation is available. p and q at equilibrium are directly observed. Q' at equilibrium is revealed by first-order variation in p and is sufficiently to identify μ as $\mu = -\frac{q}{Q'}$; this would be sufficient to identify a sub-case of the Bulow-Pfleiderer demand, such as linear, constant elasticity or exponential, but not the full class and certainly not Apt demand. Second-order variation reveals Q'' and is sufficient to estimate μ' as $\mu' = \frac{Q''q}{(Q')^2} - 1$; this would be sufficient to identify the full Bulow-Pfleidierer. Third-order variation in p would also identify $Q^{(3)}$ and thus

$$\mu'' = \frac{\left(Q^{(3)}q + Q'Q''\right)(Q')^2 - 2Q'\left(Q''\right)^2 q}{(Q')^4}.$$

This would suffice to identify Apt demand. The advantage of the identification strategy based p is that it does not rely on any supply-side assumptions (such as constancy of the conduct parameter θ). The downside is that third-order variation is required to identify Apt demand and this may be hard to come by.

On the other hand suppose that cost variation is available. We observe both the impact that a shock to cost has on price *and* the impact that the resultant change in price has on quantity; exogeneity implies that any effect on quantity comes only through the change in price. Even without variation if we observe the level of cost c we observe the markup and thus may back out μ from $p - c = \theta\mu$ if we know θ. In this case with known θ we could estimate one of the Bulow-Pfleiderer special cases even without exogenous variation, but not the full class or Apt demand.

With first-order variation we observe Q' and thus μ but also ρ (from the effect of c on p). If θ is known it can now be tested as μ is over-identified; if θ is unknown it is now identified while leaving μ independently identified. Either way we obtain θ and therefore also μ' from

$$\rho = \frac{1}{1 - \theta\mu'} \implies \mu' = \frac{1 - \rho}{\theta\rho}.$$

This suffices to estimate the Bulow-Pfleiderer class whether θ is known or not, but not Apt demand in either case.

With second-order variation we additionally observe Q'' and thus μ' but also ρ' which identifies μ'' from $\mu'' = \frac{\rho'}{\theta\rho^3}$ as shown in Subsection 1.3 of the text. Our direct identification of μ' and θ allows another test, which can be interpreted as a test of the constancy of θ in price, of whether, in fact, $\mu' = \frac{1-\rho}{\theta\rho}$. But more importantly for our purposes it identifies Apt demand. Thus the advantage of this identification approach, as opposed to that based on exogenous price variation (alone) is that lower-order variation is sufficient and the identifying assumptions can be tested. The disadvantage is that this lower-order variation only suffices if one accepts at least some untestable assumptions, here namely that the second derivative of θ is 0. Perhaps failure to reject on a sufficiently powerful test of the 0 first derivative of θ would persuade that θ's second derivative is also constant, though a sufficiently powerful test might itself require variation not far short of third-order.

D The Effects of Trade on Marginal Costs

In autarky all firms with unit labor requirements higher than a cutoff value a_c decide to exit the industry, and only more efficient firms choose to produce. Under free trade there is another important cutoff value $a_x < a_c$. Only firms with unit labor requirements smaller than a_x will choose to export. Less efficient firms will either focus on the domestic market or exit the industry.

The behavior of wages w under trade liberalization may be obtained as follows. The unrestricted entry condition $E\pi = \delta w f_e$ now instead of (13) takes the form

$$\int_{a \le a_c} \left(\left(p\left(a\right) - wa \right) q\left(a\right) - wf \right) g\left(a\right) da + \int_{a \le a_x} \left(\left(p\left(a\right) - wa \right) q\left(a\right) - wf_x \right) g\left(a\right) da = \delta w f_e. \tag{32}$$

The second integral is positive and represents the export contribution to expected firm profits. The right-hand side is increasing in w.[26] Since the first integral on the left-hand side is a nondecreasing function of w, we conclude that under trade liberalization w increases.[27]

The fact that trade liberalization increases wages has one immediate consequence for the marginal cost wa_c of a firm at the productivity cutoff. Since the firm faces the same demand but needs to cover a larger fixed cost wf of operation, its constant marginal cost wa_c must smaller than in autarky.

In Subsection 5.3 we discussed the properties of the demand elasticity as a function of quantity in a certain case of small A_0. Here we briefly consider the case of A_0 that may significantly differ from zero. We still maintain the assumption that $\theta \in \left(\frac{1}{2}, 1\right)$ (viz. that $\overline{\mu'} > 0$), which is likely to be most relevant for applications. The equation for demand elasticity (18) may be rewritten as

$$2\left(1-\theta\right)\epsilon\left(q\right) = 1 + \frac{1 + \frac{2A_0}{A_1} q^{1-\theta}}{1 + \frac{2A_2}{A_1}\frac{1}{q^{1-\theta}}}.$$

This formula makes manifest the behavior of the elasticity for small or large q. In addition, by taking the derivative of this expression it is possible to show that if A_0 and A_2 are both positive, then as long as $4A_0A_2 < A_1^2$ the elasticity is necessarily a monotone function of q in the entire range relevant for applications. It will be increasing in q for positive A_1 and decreasing for negative A_1. In other cases the elasticity may not be monotone.

[26] Here we consider only the nontrivial case with positive international trade flows.

[27] It is straightforward to show that the left-hand side is increasing. The integral may be rewritten as $\int \max\left\{\left(\left(p\left(a\right) - wa\right) q\left(a\right) - wf\right), 0\right\} g\left(a\right) da$. Thanks to our normalization of b, the first factor of the integrand may be thought of as the maximized profit of a monopolist facing costs proportional to w and demand independent of w. Since this maximized profit can only decrease in response to higher w and since $g\left(a\right)$ is independent of w, the first integral on the left-hand side is indeed a non-decreasing function of w.