A Tractable Approach to Pass-Through Patterns *

Michal Fabinger† E. Glen Weyl‡

December 2014

Abstract

For tractability, researchers often use equilibrium models that can be solved in closed-form. In practice, this means imposing unintended substantive restrictions on incidence properties that are central to many policy questions. To overcome this limitation, we characterize a set of joint supply and demand systems yielding closed-form solutions. This class is broad enough to allow substantial flexibility and thus realism, and it nests virtually all other tractable systems in the literature. We apply these more realistic structures to a range of international trade models typically solved in closed-form, thereby deriving several applied insights about, e.g., the cyclicality of wage bargaining and the organization of supply chains. Beyond parametric examples, the Laplace Transform, a standard tool used in applied mathematics and physics in analogous settings that we exploit, provides a general approach to characterizing and approximating incidence at any degree of desired tractability.

*This paper replaces a now-defunct paper "Pass-Through and Demand Forms". We are grateful to Takanori Adachi, Pol Antràs, Jeremy Bulow, Davin Chor, Swati Dhingra, Taiji Furusawa, James Heckman, Johannes Hörner, Hidehiko Ichimura, Jota Ishikawa, Sergey Kokovin, Kala Krishna, Marc Melitz, Pavel Molchanov, John Morrow, Yasusada Murata, Peter Neary, Mathieu Parenti, Andrés Rodríguez-Clare, Lars Stole, Bruno Strulovici, Yoichi Sugita, James Tybout, Alex White, and Stephen Yeaple and conference and seminar participants at the Asian Development Bank Institute, the 2013 American Economic Association Meetings, the Bates White Antitrust Conference 2012, Harvard University, Hitotsubashi University, HSE CMSSE, IDE-JETRO, Japanese Economic Association Meeting in Fukuoka, Microsoft Research New York and New England, Nagoya University, the National Graduate Institute for Policy Studies of Japan, the National University of Singapore, OEIO Tokyo Conference, Tel Aviv University, Tohoku University, Tsinghua University, the University of California, Berkeley, the University of Chicago and the University of Tokyo for helpful comments. We appreciate the work of Eric Guan and Yali Miao in creating the Apt Demand Toolkit and the more general research assistance of Konstantin Egorov, Franklin Liu, Eva Lyubich, Daichi Ueda, Ryo Takahashi, and Xichao Wang. This research assistance, except in the case of Konstantin Egorov, Eva Lyubich and Xichao Wang, was funded by the Kauffman Foundation, the Becker Friedman Institute for Research in Economics, the Japan Science and Technology Agency and the Japan Society for the Promotion of Science to which we are grateful. We are particularly indebted to Jeremy Bulow for detailed discussion and for inspiring this work and to James Heckman for advice on relevant theorems in duration analysis and nonparametric estimation.

†Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan: fabinger@e.u-tokyo.ac.jp.

‡Microsoft Research New England, One Memorial Drive, Cambridge, MA 02142 USA and Department of Economics, University of Chicago: glenweyl@microsoft.com.

NB: This paper is preliminary and incomplete. Comments, especially on our applications, are very welcome. We are also updating the paper frequently, in response to such comments and to fill gaps.

1 Introduction

For tractability, static and deterministic equilibrium models are often solved in closed-form, especially in pedagogical contexts, but also in many influential research papers (e.g., Farrell and Shapiro, 1990; Melitz, 2003; Melitz and Ottaviano, 2008; Antràs and Chor, 2013). To obtain these simple forms, researchers typically impose restrictions, such as linear or constant elasticity demand and constant or linearly increasing marginal cost, with little grounding in available empirical evidence. Yet while such restrictions are *sufficient* to obtain closed-form solutions, we show in this paper that they are not *necessary*. In fact, a wide range of behaviors, including those matching plausible empirical patterns in various contexts, can be solved in closed-form. Applying these more realistic closed-form solutions to canonical models in international trade, we predict more intuitive patterns of supply chain outsourcing, identify neglected comparative statics of wage bargaining and significantly generalize the class of explicitly aggregable models of monopolistic competition.

Our approach builds off the classic analysis of Bulow and Pfleiderer (1983, henceforth BP), who study what makes demand forms tractable in imperfectly competitive models with constant marginal cost. In the context of monopoly with constant marginal cost c, they observe that demand functions leading to closed-from solutions correspond to inverse demand $P(q) = p_0 + p_t q^{-t}$, where q represents the quantity of the good, and p_0, p_t and t are constants. This leads to equilibrium conditions (monopolist's first-order conditions) of the form

$$p_0 + (1-t)p_t q^{-t} - c = 0.$$

Because this equation is linear in q^{-t} it is straightforward to solve for the optimal quantity q.

However a broader class is equally tractable. Suppose that marginal cost and inverse demand can both be written in this form

$$f_t q^{-t} + f_u q^{-u} = 0 \tag{1}$$

for the same constants t and u, but possibly different constants f_t and f_u. Then the first-order condition also takes the same form and after multiplication by q^u, we obtain a linear equation for q^{u-t}. The BP case above is clearly the special case when $u = 0$. However, unlike this case, a flexible u allows marginal costs to be increasing, decreasing or even non-monotone, and allows demand to be concave in some regions while convex in others. Furthermore, as we show in the next section, this form is the maximally flexible class of equilibrium conditions yielding a linear solution as the BP class does. This flexibility is needed to match plausible patterns of supply and demand, such as U-shaped cost curves and demand derived from bell-shaped willingness-to-pay distributions under unit demand.[1]

Realism in these patterns, in turn, is crucial for policy analysis because the economic incidence of a range of policy interventions turns on these patterns, as we showed in Weyl and Fabinger (2013). The pass-through properties determined by the shape of equilibrium conditions play a central role in signing, among other things, the welfare impacts of third-degree price discrimination and interchange fee regulation in credit card markets. Furthermore, as we highlight in Section 3, the BP form restricts precisely the property of pass-through, whether it rises or falls with quantity, on which the qualitative answers to these policy questions turn. To make matters worse, it does so in a manner that directly contradicts the results that obtain under plausible empirical patterns such as constant marginal cost with a bell-shaped demand or U-shaped cost with constant elasticity demand.

The form of equilibrium in Equation 1 is thus both more realistic than and yet equally tractable as those presently in use. By equally tractable we mean not only that the solutions are equally parsimonious but also that analyzing a model based on one of these forms requires learning *no additional analytic tools* beyond those needed to solve existing tractable models.

Yet we did not simply arrive at this particular form by serendipity, as we explain in Section 4. Instead we utilized generic equilibrium conditions' *inverse Laplace transform*, a standard tool used for analogous purposes in applied mathematics, physics and related fields.[2] A simple analysis based on this transform allows us to characterize not only the class of equilibrium conditions that is as tractable as the BP class, but also the set of all conditions yielding solutions that are one step less tractable, in that they admit quadratic solutions, two steps less tractable (viz. cubic solutions) and so forth.

This provides a natural method for approximating incidence properties of general static,

[1] In fact, Toda (2012) and Toda and Walsh (2014) show that the double Pareto distribution, which is closely related and extremely similar to this demand form, is an accurate model of the distribution of (conditional) income and consumption distributions. The only difference between these distributions is that the transition between the two dominant power laws under our form is smoother than under the double Pareto distribution, a feature which Toda (2014)'s work suggests may fit more accurately.

[2] Technically we use the inverse Mellin transform, which is the inverse Laplace transform in the logarithm, rather than level, of q. However the Mellin transform is a much less common term and thus we maintain the Laplace terminology. Our reliance on, and characterization of tractability in terms of, polynomial solutions derived from this form builds on the approach of Kubler and Schmedders (2010).

deterministic equilibrium models at any desired level of tractability, ranging from closed-form solutions through systems that can be solved at negligible computational cost to ones that are increasingly costly to compute, with a natural corresponding trade-off in terms of accuracy. Thus we identify the full hierarchy of maximally tractable parametric approximations to non-parametric incidence patterns. Our approximations also a) focus attention on the most policy-relevant properties of demand when data is scarce, b) may be computed trivially and c) may be aggregated analytically in many standard models. These derive from approximating the non-parametric incidence pattern by a sum of constant elasticity terms where the gap between the powers in adjacent terms is constant. The degree of solvability is one less than the number of such terms.[3]

To highlight the insights allowed by the generality our framework, in Section 6 we consider a setting in which, largely due to a need for aggregation, closed-form solutions have been particular influential: international trade. First, in Subsection 6.1, we show how introducing realistic demand patterns into Antràs and Chor (2013)'s model of vertical integration along supply chains implies an arguably more plausible pattern of organization, where both basic inputs and final retailing are separated from firms that integrate intermediate points on the value chain. Then, in Subsection 6.2, we show that in the widely-applied Stole and Zwiebel (1996a,b) model of employment bargaining, labor hoarding cushions the economic cycle under realistic demand forms, while it is constant across the cycle under the BP class typically used. Counter-intuitively, given the interpretations in previous literature, this is because demand and thus profit is more *concave* during recessions, leading to more labor hoarding. Finally, in Subsection 6.3, we show that a far broader class of cost and demand models for monopolistic competition (with or without heterogeneous firms) can be solved in closed form or aggregated explicitly than those considered in the literature. These can allow, for example, for realistic U-shaped cost curves that can help explain the size distribution of exporting firms better than existing models, as we hope to show in a future draft. All of these models are solved in closed-form according to precisely the same logic we apply to the simplest monopoly models in the next section.

We conclude in Section 7 with a discussion of a variety of directions for future research. The text generally takes an expository tone, with extensive derivations and computational techniques appearing in appendices following the main text.

[3] For example $1 + q^{0.5} - q = 0$ is quadratically solvable because $0 - 0.5 = 0.5 - 1$ but $q^{-0.5} + 1 - q = 0$ is not because $-0.5 - 0 \neq 0 - 1$. Instead this second equation is cubically solvable because it is a special case of $aq^{-0.5} + b + cq^{0.5} + dq = 0$ where $a = b = -d = 1$ and $c = 0$.

2 Tractable Equilibrium Conditions

In this section we develop the central ideas behind our approach in the simplest setting that suffices to illustrate them. Our results apply much more broadly, as we discuss in detail in Appendix A, to essentially any static, deterministic equilibrium model that can be reduced to a single equation by aggregation. These include partial equilibrium models of perfect competition, essentially all symmetric models of imperfect competition (including with differentiated products), supply chain models with perfectly complementary goods and imperfect competition to produce each, aggregable models of general equilibrium (with perfect or monopolistic competition), etc. Our results can also be extended, somewhat less directly, to models that cannot be reduced to a single equation or closely-linked system of equations (as we show by example in Appendix F) or applied to parts of more complicated models to provide intermediate analytic solutions.

However, this generality distracts from exposition of our basic insight and thus we do not focus on it here. Instead in this section we exposit our general approach in the context of the simplest possible application (the classical monopoly model). We then illustrate the breadth with which our approach applies with richer, but specific, examples in Section 6 below. The observations here are stated and proved more formally in Appendix B.

Consider the classical monopoly problem. If a monopolistic firm chooses output level q, it incurs cost $C(q)$ and can sell its output at price $P(q)$. The first-order condition for maximization of the firm's profits $\Pi(q) \equiv P(q)q - C(q)$ equates marginal revenue $MR(q) = P(q) + P'(q)q$ to marginal cost $MC(q) \equiv C'(q)$:

$$P(q) + P'(q)q - MC(q) = 0. \tag{2}$$

Much analysis of monopoly is concerned with comparing monopolistic outcomes to the social optimum characterized by price equal to marginal cost:

$$P(q) = MC(q), \tag{3}$$

which would be achieved under perfect competition.

Now consider Equations 2 and 3 when price $P(q)$ takes perhaps the most canonical form (Mill, 1848): constant elasticity. In this case inverse demand may be written as $P(q) = p_t q^{-t}$ for some positive constants p_t and t, and marginal revenue

$$P(q) + P'(q)q = p_t(1-t)q^{-t}$$

has exactly the same functional form, differing only by a multiplicative constant.

More generally, if price $P(q)$ is a sum of such constant elasticity terms (viz. if it is a linear combination of powers of q) marginal revenue inherits its functional form.[4] Since inverse demand is average revenue, we can say that one functional form characterizes both average revenue and marginal revenue. A similar statement applies to the cost side as well: if average cost is a linear combination of powers of q, then the marginal cost does not differ in functional form.

Observation 1. *Average and marginal quantities have the same mathematical form (viz. finite functional basis) differing only in the coefficients if they can be written as a finite sum of constant elasticity terms. We refer to this property as* average-marginal form preservation *or* form preservation *for short.*

A formal theorem describing most general average-marginal form-invariant functions may be found in Appendix C.

Now suppose that the functional choices for price and cost lead to the first-order condition (Equation 2) of the form

$$f_t q^{-t} + f_u q^{-u} = 0. \tag{4}$$

Then the equilibrium quantity can be found easily by multiplying the equation by q^u and solving the resulting linear equation for q^{u-t}, which yields $q = (-f_u/f_t)^{1/(u-t)}$. A 'linear' solution of this kind is possible precisely when the equilibrium condition is a sum of two constant elasticity terms. Of course linear solutions are possible for transformations other than powers of q (e.g. $\tan q = 5$), but these will fail to be form-preserving and thus will not typically be equally tractable in both the monopoly and social optimization problem (viz. if $P(q) = \tan(q)$ then $MR(q) = \tan(q) + q\sec^2(q)$).

What does the form in Equation 4 require of inverse demand and cost? It is possible that they could have terms that look nothing like the left-hand side of Equation 4 if and only if these terms are exactly balanced between marginal revenue and marginal cost. Absent such non-generic coincidences, however, *both marginal revenue and marginal cost* must be of the form of the left-hand side of Equation 4. This requires that:

1. By Observation 1, inverse demand must be of the form $P(q) = p_t q^{-t} + p_u q^{-u}$.[5]

2. Average cost $AC(q) \equiv C(q)/q$ must have the same form and thus equal $AC(q) = ac_t q^{-t} + ac_u q^{-u}$ for constants ac_t and ac_u.

[4]In statistics this is known as a "quantile mixture" of Pareto distributions.

[5]The first mention of this form in economics we are aware of is in Mrázová and Neary (2014a) as a demand form. However they do not assume cost takes the same form and thus do not gain any tractability benefits from this form. As a statistical distribution (viz. quantile function) this form is well-known in the modeling of flood flows as the Wakeby distribution (Houghton, 1978).

3. The exponents t and u in the expression for inverse demand must be the same as those for marginal cost and average cost. For example, if demand is linear ($P(q) = p_0 - p_{-1}q$) then average cost must also be linear. If demand has constant elasticity ϵ ($P(q) = p_{1/\epsilon}q^{-1/\epsilon}$), then one of the terms determining average cost must have exponent $-1/\epsilon$, i.e. $AC(q) = ac_t q^{-t} + ac_{1/\epsilon}q^{-1/\epsilon}$. Average cost linearly proportional to an arbitrary power q^{-t} may be thought of as a special case in which $ac_{1/\epsilon} = 0$.

Observation 2. *The only linearly-solvable, form-preserving monopoly problems have both inverse demand and average cost that are a sum of two constant elasticity terms with the same pair of elasticities on the cost and demand side.*

Many studies are concerned with the case where cost has a constant average component, either imposed as a specific tax by the government or arising naturally from the structure of production. Including such a component is needed, for example, if one wishes to solve explicitly for the inverse of marginal cost (the supply function) or of inverse demand (the direct demand function). Such inverses are useful for obtaining closed-form solutions in some models, as we illustrate in Subsection 6.1. As a result, we primarily focus, in what follows, on cases admitting such an explicit inverse solution.

BP (Bulow and Pfleiderer, 1983) studied this case and proposed a class of demands with $P(q) = p_0 + p_t q^{-t}$. It follows directly from our previous observation that this is the maximal class allowing a constant average component of cost and yielding linear solutions. BP focused on the case when cost has only a constant average component, but evidently linear solutions arise if average cost also has a term matching the elasticity of the non-constant term in inverse demand.

Observation 3. *The only linearly solvable, form-preserving monopoly problems that include a constant component in average cost and thus admit explicit inverse solutions have inverse demand of the BP form and average cost of a matching form.*

While linear solutions are the simplest, they are far from being the only ones that are analytically tractable. Other polynomials in q^t, such as quadratic and cubic functions, also admit closed-form solutions. To take a simple example that we will return to in Subsection 4.1, consider

$$f_0 + f_{-t}q^t + f_{-2t}q^{2t} = 0, \tag{5}$$

where f_0, f_{-t}, f_{-2t}, and t are constants. If we denote $x \equiv q^t$, we obtain a quadratic equation

$$f_{-2t}x^2 + f_{-t}x + f_0 = 0,$$

which immediately yields closed-form solutions for q^t and thus q.

Here a quadratic solution is possible precisely because the powers of the three constant elasticity terms ($0, -t$ and $-2t$) are *evenly spaced* : $0 - (-t) = -t - (-2t)$. A quadratic solution would not be possible, for example, to an equation of the form

$$q^{-t} + 1 - q^{2t} = 0,$$

as $2t - 0 \neq 0 - (-t)$. However, in this particular case we can think of the equilibrium condition as a cubic equation in $x \equiv q^{-t}$, namely

$$x + 1 - x^{-2} = 0 \iff x^3 + x^2 - 1 = 0.$$

We see that to determine the degree of polynomial needed to solve the equation, a natural measure of its tractability (Kubler and Schmedders, 2010; Kubler et al., 2014), we must first determine the *minimum number of evenly-spaced power terms necessary to write the equation.*[6] While Observation 2 might suggest that it is simply the number of constant elasticity terms that determines the complexity of the solution to equilibrium conditions, it is in fact the number of evenly-spaced terms needed for the equation. When there are only two terms, the even-spacing requirement is automatically satisfied and thus did not need to be mentioned.

Observation 4. *The order of polynomial required to solve a form-preserving monopoly problem is one less than the minimum number of evenly-spaced power terms required to write its first-order condition.*

Let us now return to the equilibrium condition in Equation 5. It is worth noting that it gives much more flexibility than is possible in the BP class. For example, consider $t = \frac{3}{4}$, as we do in Subsection 4.1. Then inverse demand can have a convex component $(-q^{\frac{3}{4}})$ *and* a concave component $(-q^{\frac{3}{2}} = -q^{2\cdot\frac{3}{4}})$. Similarly, marginal cost may have both increasing and decreasing components, with different shapes, so long as these shapes match with those used to compose inverse demand. Furthermore, this is only one of many possible quadratic forms for Equation 2. If one is willing to employ cubic forms, even greater flexibility is possible; once one reaches the limit of analytically solvable equations (i.e. quartic equations) the flexibility is very large indeed.[7]

[6] Barnett and Jonas (1983) discussed a demand system based on linear combinations of fractional power functions, where the powers were not necessarily equally spaced. In that context, these were functions of prices rather than quantities and for that reason the demand specification was not tractable in our sense. For additional functional specifications that contain fractional powers see Barnett and Lee (1985); Diewert (1971, 1973); Lau (1986).

[7] By the Abel–Ruffini theorem, there is no general algebraic solution to polynomial equations of degree five or higher.

Observation 5. *If one is willing to allow quadratic, and certainly if one allows cubic or quartic, solutions then equilibrium conditions with analytic solutions allow significant flexibility in the structure of inverse demand and marginal cost.*

Note that exploiting these simple closed-form solvable systems does not require the analyst to learn any additional analytical tools. All that is required is to use any one of the richer functional forms discussed in Section 4 below and solve them using familiar, pen-and-paper techniques. This provides substantial additional flexibility at little or no analytical cost; thus we believe there is *no argument*, beyond familiarity, for researchers to continue to limit themselves to traditionally-used functional forms of demand and cost. In fact, as we will see in the next section, if authors do wish, for some reason, to continue to limit the number of parameters, our classes offer far more plausible approximations to real-world demand and cost forms at no analytical cost.

Despite their flexibility, however, even our broader simple closed-form solvable systems will not *perfectly* match arbitrary demand and cost structures, because even in the quartic case they have only a finite number of polynomial terms in q^{-t}. Nonetheless, if we continue to add more terms, we can approximate any smooth cost and demand structure by a process analogous to Taylor series approximation.

If we require a perfect fit, then for a generic equilibrium condition $F(q) = 0$ we need to add a *continuum* of constant elasticity terms. In this case Equation 2 becomes

$$\int_{-\infty}^{\infty} f(t) q^{-t} dt = 0, \tag{6}$$

where f is some function of t, possibly with mass points.[8] The function $f(t)$ may be computed from the original equilibrium function $F(q)$ using an "inverse Laplace transform" (ILT).[9] For brevity, we refer to $f(t)$ as the Laplace equilibrium function; in general, we will use the adjective "Laplace" to refer to similar counterparts of various economic variables.[10]

The simple equilibrium conditions discussed above involving a finite number of constant-elasticity terms may be thought of as Equation 6 with the Laplace equilibrium function $f(t)$ consisting of a finite number of (positive or negative) mass points, located at values of $-t$ that match the original powers of q. From this perspective, what makes these equilibrium systems tractable is that they have "simple" Laplace equilibrium functions with a small number

[8]More precisely, f is a 'generalized function', or in equivalent terminology, a 'distribution'.

[9]The left-hand side of a general equilibrium condition $F(q) = 0$, when written as a function of $\log q$, is the (bilateral) Laplace transform of $f(t)$, which means that $f(t)$ may be recovered from $F(q)$ by inverse Laplace transform. The numerous common techniques for computing such inverse transforms will be discussed in the next draft.

[10]For example, the Laplace inverse demand $p(t)$ associated with inverse demand $P(q) = 3q^{-1/2}$ is a mass point of size 3 at $t = 1/2$, and similarly for Laplace average cost $ac(t)$, etc.

of evenly spaced mass points. Such simple Laplace equilibrium functions will accurately approximate broader Laplace equilibrium functions when most of the (positive or negative) mass of f is clustered near these points, as we illustrate in detail in Subsections 4.2-4.3.

Throughout the piece we use t and u to represent a fixed value in the case of these simple forms and to represent variables of integration or summation when discussing more general Laplace functions: when they serve the former role we place them in subscripts and in the later we write them as functional arguments.

The Laplace versions of demand and supply curves are often simpler to manipulate than are the objects themselves in many economic problems. For example, marginal revenue associated with inverse demand $P(q)$ and Laplace inverse demand $p(t)$ may be expressed as

$$MR(q) \equiv P(q) + qP'(q) = \int_{-\infty}^{\infty} (1-t)p(t)q^{-t}dt \equiv \int_{-\infty}^{\infty} mr(t)q^{-t}dt.$$

That is, the Laplace marginal revenue is simply $mr(t) = (1-t)p(t)$. Similarly, if Laplace average cost is $ac(t)$, then Laplace marginal cost equals $(1-t)\,ac(t)$. If we combine these two expressions, we conclude that the Laplace equilibrium function is just $(1-t)(p(t) - ac(t))$. More generally, analyzing Laplace equilibrium functions is useful in revealing a variety of properties of these systems.

Observation 6. *Inverse demand, marginal cost and equilibrium conditions can be easily related to one another through their inverse Laplace transforms, which exists very generally and to which we refer to with the adjective* Laplace. *Among other useful properties, the tractability of a monopoly's first-order conditions corresponds to the Laplace equilibrium function consisting of a small number of evenly-spaced mass points.*

Solving for the optimal price, quantity and cost is an intermediate step to determining welfare quantities of interest. Calculating these from cost and demand involves simple integration for which their Laplace representation is useful:

$$CS(q) = \int_0^q P(x)dx - P(q)q = \int_{-\infty}^{\infty}\int_0^q p(t)\tilde{q}^{-t}d\tilde{q}dt - \int_{-\infty}^{\infty} p(t)q^{1-t}dt.$$

Note that the inner integral converges only for $t < 1$ and thus we assume that the Laplace inverse demand function $p(t)$ has support in $(-\infty, 1)$, as well as well-behaved asymptotics. Under this assumption we obtain

$$CS(q) = \int_{-\infty}^{1} \left[\frac{p(t)}{1-t} - p(t)\right] q^{1-t}dt = \int_{-\infty}^{1} \frac{t}{1-t}p(t)q^{1-t}dt = -\int_{-\infty}^{0} \frac{t+1}{t}p(t+1)q^{-t}dt,$$

where the last equality was obtained by relabeling $t \to t+1$. We see that the Laplace consumer surplus equals $cs(t) = -\frac{t+1}{t} p(t+1)$.

Similarly if we denote Laplace profits as $\pi(t)$, analogous calculations give that $\pi(t) = p(t+1) - ac(t+1)$. The Laplace marginal cost $mc(t)$ should have support in $(-\infty, 1)$ if cost $C(q)$ is to be finite. These results imply representation of utility, consumer surplus, profit, and cost of a form that resembles the form of the equilibrium function: simple Laplace inverse demand and average cost imply simple Laplace profit and surplus.[11]

Observation 7. *Laplace welfare quantities inherit the simplicity Laplace inverse demand and average cost. If the functional forms of inverse demand and average cost are valid for any level of quantity, then their Laplace counterparts are supported in* $(-\infty, 1)$.

3 Existing and Desirable Forms

In this section we discuss why allowing for the greater flexibility our tractable forms permit is necessary to obtain correct answers to important policy questions and explain the sense in which this flexibility is missing from existing tractable forms. While we provide motivation here by briefly discussing a range of issues for which pass-through patterns are relevant in a several field, we develop more detailed applications to international trade models in Section 6.

3.1 What difference does functional specification make?

We begin by discussing how the form of the equilibrium conditions impacts policy questions. In Weyl and Fabinger (2013) we argue that a wide range of welfare properties are determined by the most basic incidence quantity: the *pass-through rate* ρ at which a specific tax on a producer translates into an increase in price for consumers, $\rho \equiv \frac{dp^\star}{d\tau}$. For example, focusing on the monopoly case we used to motivate the previous section:

1. The local incidence I of a specific tax (per-unit tax) τ under a general model of symmetric imperfect competition is equal to the pass-through rate ρ. Here we define local incidence as $I \equiv \frac{dCS}{d\tau} / \frac{dPS}{d\tau}$, where CS and PS are consumer and producer surplus, respectively.

[11] If a fixed cost of entry or participation is introduced, this adds one more term (placed at 0) to the resultant equation and may thus will raise the order of solvability. However, with this concern in mind the Laplace inverse demand and average cost may be chosen appropriately to ensure a minimum of impact on tractability in cases where a fixed cost is relevant.

2. The global incidence $\overline{I}$ is equal to a weighted average $\overline{\rho}$ of the pass-through rate ρ. The global incidence is defined as the ratio of welfare gains associated with the existence of the market, namely $\overline{I} \equiv CS/PS$. The precise form the average pass-through rate is $\overline{\rho} \equiv \int_{\tau=0}^{\infty} w_Q(x)\rho(x)dx$ with non-negative weight function $w_Q(x) \equiv Q(x)/(\int_{\tau=0}^{\infty} Q(x')dx')$. In other words, $\overline{\rho}$ is the *quantity-weighted average pass-through rate* for specific tax rates ranging from zero to infinity.

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.[12]

4. In a supply chain consisting of a consecutive monopolies, the levels and slopes of pass-through rates determine the relationships between markups at various stages of the supply chain, as well as the effects of changes in industrial structure of these production stages (Adachi and Ebina, 2014a,b).[13]

5. The slopes of pass-through rates play an important role in determining whether third-degree price discrimination is socially beneficial or harmful (Fabinger and Weyl, 2014).

6. Pass-through plays similarly important role in a range of other applications, such as behavioral welfare analysis, merger analysis, strategic effects in oligopoly and the welfare impacts of regulation of platforms.

To see how pass-through relates to the form of the equilibrium conditions, note that a specific tax τ enters linearly into (2). If we use the notation $\Pi(q)$ for pre-tax profits, the firm maximizes $\Pi(q) - \tau q$, and its first-order condition (2) may be written as $\Pi'(q^\star) = \tau$, where $q^\star$ is the optimally chosen quantity, corresponding to optimal price $p^\star \equiv P(q^\star)$. The related comparative statics equation $\Pi''(q^\star)dq^\star = d\tau$ then provides an expression for the pass-through rate $\rho \equiv \frac{dp^\star}{d\tau} = \frac{dq^\star}{d\tau}P'(q^\star)$:

$$\rho = \frac{dp^\star}{dt} = \frac{P'\left(q^\star(t)\right)}{\Pi''\left(q^\star(t)\right)} = \frac{\int_{-\infty}^{1} tp(t)q^{-1-t}dt}{\int_{-\infty}^{1}\left[(1-t)p(t) - c(t)\right]tq^{-1-t}dt} \tag{7}$$

We see that the inverse pass-through rate $\frac{1}{\rho}$ may be written as a weighted average of $1-t-\frac{c(t)}{p(t)}$:

[12] Relatedly, Kremer and Snyder (2014) show that the bias of R&D expenditures that firms choose to engage in, relative to the preferences of a social planner, is driven by pass-through.

[13] A related, recent results are that pass-through plays a key role in determining which modes of electronic book pricing are most pro-competitive (Gaudin and White, 2014a) and whether a federal authority, faced with revenue-maximizing states but placing greater welfare on resident welfare than do the states, would regulate states to use ad-valorem or specific taxes to raise revenue (Gaudin and White, 2014b).

$$\frac{1}{\rho}=\int_{-\infty}^{1}\left(1-t-\frac{c(t)}{p(t)}\right)w\left(t\right)dt,$$

with weight function

$$w\left(t\right)\equiv\frac{tp(t)q^{-1-t}}{\int_{-\infty}^{1}\tilde{t}p(\tilde{t})q^{-1-\tilde{t}}d\tilde{t}},$$

which does not necessarily have to be always positive, since t and $p\left(t\right)$ may have opposite signs. Using this representation of pass-through ρ we can intuitively understand how its value is influenced by the behavior of Laplace inverse demand $p\left(t\right)$ and Laplace cost $c\left(t\right)$.

If $p(t)t$ puts a lot of (positive) weight on negative t values (or negative weight on positive values), pass-through will tend to be less than one. This corresponds to the fact that log-concavity tends to lead to pass-through below 1. The reverse occurs if the weight is placed on positive values of t. Similarly weight on t values below -1 (concave terms) will lead the pass-through to be below $\frac{1}{2}$ while convex terms will lead it to be above $\frac{1}{2}$.

Contributions to the marginal cost play identical roles, though with opposite signs. If $tp(t)>0$ then $\frac{c(t)}{p(t)}$ is positive if $c(t)>0$ for $t>0$ or if $c(t)<0$ and $t<0$. Thus declining marginal cost terms will tend to raise pass-through by decreasing $1-t-\frac{c(t)}{p(t)}$. Conversely increasing components of marginal cost will raise pass-through. This is consistent with the classical intuitions that pass-through is highest when supply is most elastic (and even higher when it is "more than perfectly" elastic). Thus the form of equilibrium conditions, both through the structure of cost and demand, is crucial to shaping pass-through rates and thus a wide range of standard policy questions in imperfectly competitive markets.

3.2 Existing tractable functional forms

A simple example of this is the ubiquitous BP (Bulow and Pfleiderer, 1983) class. Consider the case of inverse demand $P(q)=p_0+p_tq^{-t}$ and constant average cost $AC\left(q\right)=ac_0$. Then Expression 7 for the pass-through rate simplifies to

$$\rho=\frac{\int_{-\infty}^{1}tp(t)q^{-1-t}dt}{\int_{-\infty}^{1}(1-t)\left[p(t)-ac(t)\right]tq^{-1-t}dt}=\frac{p_{-t}tq^{-1-t}}{(1-t)tp_{-t}q^{-1-t}}=\frac{1}{1-t},$$

because in this case Laplace inverse demand $p\left(t\right)$ consists just of a point mass p_t at the choice t and a point mass p_0 at 0, and similarly $ac(t)$ consists of a point mass ac_0 at 0.

We see that pass-through is constant and depends only on the fixed number t. If average cost has the BP form with the same value of t as does inverse demand (i.e. $AC\left(q\right)=ac_0+ac_tq^{-t}$), then pass-through will remain independent of q, but in addition to the constant

t, it will also depend on the relative size of the constants mc_t and p_t. This constant pass-through class of demand functions includes every imperfectly competitive model component we are aware of that has been solved in closed-form, including

1. Constant elasticity demand, $P(q) = p_t q^{-t}$ with t and p_t positive, combined with constant average cost (e.g. Dixit and Stiglitz, 1977; Krugman, 1980; Melitz, 2003) or power-law cost, $AC(q) = ac_u q^{-u}$ with u negative and ac_u positive (e.g. Antràs and Chor, 2013).[14]

2. Linear demand, $P(q) = p_0 + p_{-1}q$ with $p_0 > 0$ and $p_{-1} < 0$, combined with constant marginal cost (e.g. Singh and Vives, 1984; Melitz and Ottaviano, 2008) or linearly increasing marginal cost, $MC(q) = mc_0 + mc_{-1}q$ with mc_{-1} positive, (e.g. Farrell and Shapiro, 1990).

3. Exponential demand, $P(q) = -a \log q = -\lim_{t \to 0+} \frac{a}{t}\left(q^{-t} - 1\right)$ with a positive, combined with constant marginal cost (e.g. Baker and Bresnahan, 1985, 1988).

The fact that so many influential results have been founded on this particular structure would be justifiable if there were empirical or intuitive reasons to believe that equilibrium conditions must have one of these form. Are there?

3.3 Functional forms that are plausible, but not tractable

Unfortunately, there is little empirical evidence to guide an answer to this question as essentially all empirical studies we are aware of have considered the level rather than the slope of pass-through. These have generally found that pass-through rates range widely depending on circumstances, though few systematic patterns have emerged.[15]

[14] Although Antràs and Chor formulate their supply chain problem in terms of constant marginal costs, it may be rephrased in terms of power-law costs, as we discuss in Subsection 6.1.

[15] 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). However, recent work in this literature has found higher pass-through rates, near unity (Fabra and Reguant, 2014).

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 rate of individual firm cost shocks to that firm's price to be small, at about 25 to 60%. (They measured pass-through elasticity, so we used the relative markup to convert their result to the corresponding range for the pass-through rate ρ stated here.) 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

More systematic patterns emerge from the large literatures on demand and cost estimation. Since its emergence in the 1980's (Bresnahan, 1989), the "new" empirical literature in industrial organization has typically, building off the pioneering work of Berry et al. (1995), employed demand curves based on statistical distributions with a single, bell-shaped peak such as Normal, lognormal, logistic or Type I Extreme Value. These shapes have strong theoretical under-pinnings as outcomes of various statistical processes such as order statistics or sums of independent additive or multiplicative components of demand.

More importantly, they have been supported by two types of empirical evidence. The first is indirect, but is the oldest quantitative tradition in demand estimation in economics. Say (1819) argued that willingness-to-pay is likely to be proportional to income and thus that the distribution of willingness-to-pay has the same shape as the income distribution. Based on extrapolations of early measurements of top incomes following power laws (Garnier, 1796; Say, 1828), Dupuit (1844) and Mill (1848) argued that demand would have a constant elasticity, an observation that appears to be the origin of the modern focus on constant elasticity demand form (Ekelund and Hébert, 1999). However evidence on broader income distributions that became available in the 20th century as the tax base expanded (Piketty, 2014) shows that, beyond the top incomes that were visible in 19th century data, the income distribution is roughly lognormal through the mid-range and thus bell-shaped. Distributions that accurately match income distributions throughout their full range (Reed and Jorgensen, 2004; Toda, 2012, 2014) have a similar bell shape, but incorporate the Pareto tails perceived in the 19th century data.

Of course, consumers' idiosyncratic values also contribute to demand shapes. More direct measurements of the distribution of willingness-to-pay in contexts, like auctions, where it can be recovered non-parametrically using structural models (Haile and Tamer, 2003; Cassola et al., 2013) suggest that lognormal distributions fit well. Even in contexts where valuation distributions are complicated (have multiple peaks), bell-shaped distributions fit much better than do the Generalized Pareto Distributions that generate the BP demands (Burda et al., 2008).

Thus there is a strong empirical and theoretical case for ensuring demand forms can match the basic shape of such distributions. One consistent pattern they exhibit, as we show in Appendix D, is that these demand curves are more convex at higher prices (lower quantities). This implies that, when paired with constant average cost, they give rise to pass-through rates that monotonically decrease in quantity. For the normal, logistic and Type I Extreme Value distributions, curvature ranges from 1 at very small quantities, implying

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.

constant average cost pass-through of 1, to $-\infty$ at very high quantities (implying constant average cost pass-through of 0).

For lognormal and other approximations to the income distribution the pattern is similar, but the exact range over which curvature spans as quantities change depends on exact parameter values and at sufficiently small quantities in the case of the lognormal distribution curvature modestly rises (in quantity) because the distribution has thin tails. More realistic, Pareto-tailed income distribution approximations have curvature that steadily declines in quantile, reaching a high plateau at very low quantile/quantity (high income). In any case, because all these forms have curvature that range broadly as a function of quantity, they are all inconsistent with the BP form for demand.

On the cost side, few studies have systematically considered the shape of cost curves. However, at least since Marshall (1890) economists have commonly assumed cost curves have a U-shape, with economies of scale at small size and diseconomies of scale at sufficiently large scale. Some limited evidence confirms this widely-held hypothesis (MacDonald and Ollinger, 2000). Despite this, we know of no common tractable equilibrium model incorporating this feature. Nearly all standard tractable imperfect competition models have firms with constant marginal cost; a few (Farrell and Shapiro, 1990; Antràs and Chor, 2013) feature firms with linearly increasing marginal cost and some incorporate fixed costs.[16] U-shaped costs curves lead to pass-through rates that monotonically fall with quantity, just as bell-shaped demand curves do under constant marginal cost when demand determines pass-through. The reason is that at low quantities economies of scale reinforce pass-through while at high quantities diseconomies reduce it just as under perfect competition.

Thus, for reasons both of supply and demand, equilibrium conditions that, at very least, allow for pass-through rates that fall significantly as quantities increase are necessary to correspond with what we expect from both theory and empirical work. Recent empirical work (Einav et al., Forthcoming) has found other patterns in the context of resale goods, where demand is so convex at low quantities that pass-through is infinite and convexity is more moderate at higher quantities. We know that equilibrium conditions of both of these forms will have significantly different implications for a range of policy questions than do currently available tractable forms, motivating the extensions we explore in the next section.

[16]See Subsection 6.1 for the transformation of the Antràs and Chor (2013) model into our language that makes this and the above discussions accurate.

4 Closed-Forms and Approximation

In this section we elaborate on the outline we gave Section 2 by discussing more specifically the nature and advantages of our approach to approximating equilibrium conditions.

4.1 Second-degree forms

From Section 2 we know that the equilibrium function has the same form as average cost and inverse demand if it must takes the form of a polynomial in $x \equiv q^{-t}$.When does this polynomial have a quadratic solution?

Consider an equation of the form

$$ax^{2-\kappa} + bx^{1-\kappa} + cx^{-\kappa} = 0.$$

It has a quadratic solution because if we multiply both sides of the equation by x^{κ}, we obtain $ax^2 + bx + c = 0$. It is clear that only polynomials of this and equivalent forms admit quadratic solutions. The form requires that the spacing between the exponents of the highest power term and middle power term is the same as that between the exponents of the middle power term and lowest power term. For example

$$aq^{11/2} + bq^{5/2} + cq^{-1/2} = 0 \iff a\left(q^3\right)^2 q^{-1/2} + bq^3 q^{-1/2} + cq^{-1/2} = 0$$

would be an admissible form if we want to obtain closed-form solutions. For many purposes, such as studying the impact of an ad-valorem tax or proportional "iceberg" trade costs, this class of possibilities offers a broad range of flexibility.[17] However a limitation of this broad class is that it may not allow for tractability if constant components of marginal cost, corresponding to constant terms in the equilibrium conditions, are introduced, say by specific taxes. Such constant components are needed, for example, to analytically calculate non-local pass-through rates, i.e. price responses to large specific tax changes or equivalent cost changes. They are also, more importantly, needed for an explicit direct demand or supply function, as is important for obtaining closed-form solutions in some models, as we illustrated in Subsection 6.1.

If it is important to allow for specific taxes, we must restrict ourselves to the subclass where $2t - \kappa = 0$, $t - \kappa = 0$ or $\kappa = 0$. The first and third of these cases are equivalent,

[17] In fact, an even more tractable form with linear solutions, discussed in the introduction, is possible if one does not require the possibility of constant components to marginal cost such as specific taxes. If equilibrium conditions take the form $f_t q^{-t} + f_u q^{-u} = 0$ for any $t, u < 1$ then by the same logic as in the quadratic case the system admits a linear solution, obviously allowing log-concave and log-convex components of demand and increasing and decreasing components of marginal cost.

because t may take either sign. The demand forms giving rise to them were discussed in a previous version of this paper under the name "Adjustable Pass-Through" (Apt) demand (Fabinger and Weyl, 2012), and have been applied by work building on that paper (Gaudin and White, 2014b). We now briefly consider the range of economic behaviors accommodated in each of these cases.

First, consider the case when both terms have negative exponents; that is the equilibrium conditions take the form

$$f_0 + f_t q^{-t} + f_{2t} q^{-2t} = 0 \tag{8}$$

with $t \in (0, 1/2)$. To ensure positive surplus, f_{2t} should be positive. Therefore at least one of f_0 or f_t must be negative for an equilibrium to exist. Pass-through is always greater than unity in this form. If $f_t > 0$ then pass-through monotonically decreases in quantity (increases in the tax), while if $f_t < 0$ it monotonically increases in quantity (decreases in the tax).

Equation 8 corresponds to log-convex demand and to marginal cost functions with components that have constant, declining (economies of scale) components and concavely increasing (diseconomies of scale) components that manifest at high quantities. On the demand side these can mimic log-convex increasing pass-through demand forms such as lognormal distribution with very high variance or log-convex decreasing pass-through demand forms, possibly with infinite pass-through at sufficiently low quantities as found by Einav et al. (Forthcoming). On the cost side they can match cost curves that have economies of scale at low quantities that either persist throughout he full range of quantities or turn into diseconomies at large quantities, creating a U-shape.

Second, consider the case when one term has a negative and the other an equal and opposite exponent, as discussed several times above. Then equilibrium conditions take the form

$$f_0 + f_t q^{-t} + f_{-t} q^{t} = 0 \tag{9}$$

for $t \in (0, 1)$. Again, $f_t > 0$ and either f_0 or f_{-t} must be negative. When $f_{-t} > 0$ this case is quite similar to Equation 8 when $f_t < 0$. When $f_{-t} < 0$ pass-through monotonically decreases in quantity and passes from being above to below unity.

This form allows for both log-convex and log-concave contributions to demand arising, respectively, from positive coefficients on q^{-t} and negative coefficients on q^t in demand. However, the latter must be convex because $t \in (0, 1)$, so it does not actually allow for bell-shaped willingness-to-pay distributions. On the cost side it allows for a U-shaped form when marginal cost places positive weight on both q^t and q^{-t}. However, again, the increasing portion of marginal cost must be concave, so convexly increasing marginal costs are

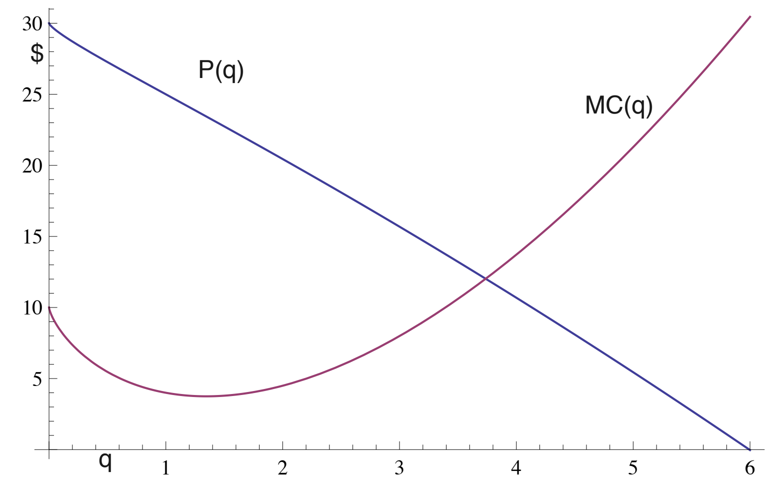

Figure 1: An example of a demand and cost curve contributing to equilibrium conditions that can be solved quadratically. The demand curve correspond to a bell-shaped willingness-to-pay distribution and the cost curve has the standard U-shape. $P(q) = 30 - 4q^{\frac{3}{4}} - q^{\frac{3}{2}}$ and $MC(q) = 10 - 10q^{\frac{3}{4}} + 4q^{\frac{3}{2}}$.

impossible.

Third, let us discuss a case where this is actually possible, namely:

$$f_0 + f_{-t}q^t + f_{-2t}q^{2t} = 0 \tag{10}$$

for $t > 0$. One of f_{-t} and f_{-2t} must be negative, and at least one term of the three positive, for a solution to exist. When $f_{-2t} < 0$ pass-through monotonically falls in quantity. If $f_{-2t} > 0$ then pass-through rises monotonically in quantity from below unity to above unity.

When the coefficients of non-constant terms are both negative, the qualitative shape of demand depends on the magnitude of t. While demand always becomes more concave the larger q becomes, it may be convex at both low and high quantities (if $t < \frac{1}{2}$), concave in both ranges (if $t > 1$) or first convex then concave, and thus having a bell-shaped willingness-to-pay distribution, if $t \in \left(\frac{1}{2}, 1\right)$. In all these cases demand is globally log-concave, however. When the coefficient of q^{2t} is positive, demand transitions from being log-concave to being arbitrarily log-convex at sufficiently high quantities, a useful form for representing resale settings as in Einav et al..

Similarly a rich combination of behaviors of cost functions can be supported. In fact these are too rich to describe thoroughly here. But consider just one example. If $t \in \left(\frac{1}{2}, 1\right)$, the coefficient in the marginal cost expression on q^t is negative and the term on q^{2t} is positive,

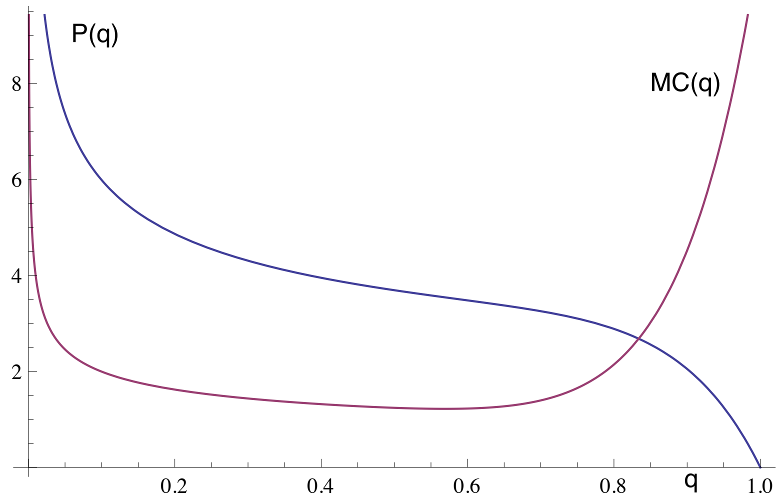

Figure 2: An example of a demand and cost curve contributing to equilibrium conditions that can be solved linearly, but with no constant component of marginal cost. The demand curve correspond to a pronounced bell-shaped willingness-to-pay distribution and the cost curve has the standard U-shape. $P(q) = 3\left(q^{-.3} - q^{10}\right)$ and $MC(q) = q^{-.3} + 10q^{10}$.

then marginal cost have a nice, smooth U-shape as shown in Figure 1. In the figure we also show the inverse demand corresponding to a bell-shaped willingness-to-pay distribution that can be represented by the same t value.

While attractive and capable of fitting nearly-linear bell shaped distributions as found empirically by Dinerstein et al. (2014), this form does not allow the convexity of demand at low quantities, nor its concavity at high quantities to be very pronounced. As a result, the bell shape in the distribution of willingness-to-pay is quite subtle and the distribution is fairly close to uniform. Allowing greater contrasting curvature at the two ends of the inverse demand curve is possible with a quadratic, even linear, solution if one is willing to sacrifice the possibility of solving explicitly for the impacts of specific taxes. An example with a linear solution is shown in Figure 2. However, if one wants to include specific taxes and achieve these contrasting curvatures at the two ends, one must accept a higher-order polynomial solution as discussed in the next subsection.

4.2 Higher-degree forms

We have seen that replacing the linear solutions of the previous section by simple quadratic solutions, we can study a substantially wider range of economic behaviors. However, more still is possible if we allow cubic or quartic equations, which always yield a closed-form solution in terms of radicals. Although traditional paper-and-pencil manipulation of cubic or quartic solutions may require substantially more effort than that of quadratic solutions,

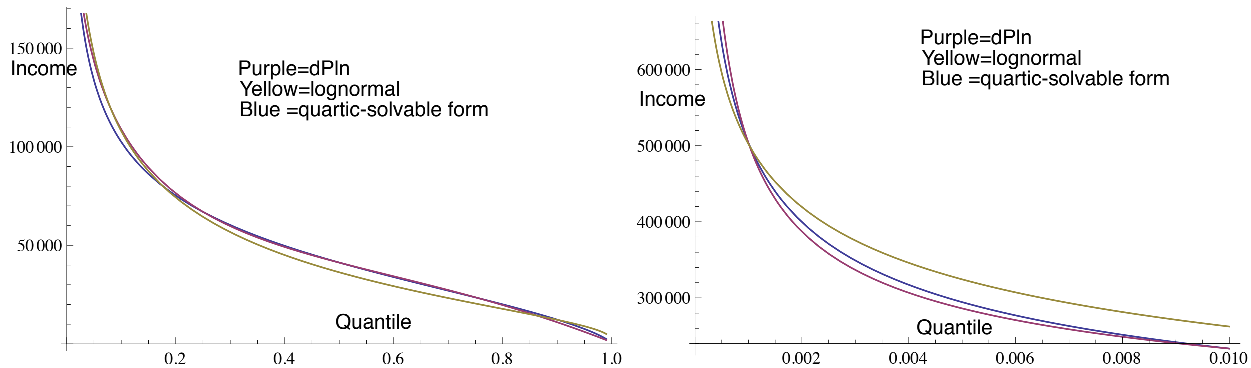

Figure 3: Fitting a quartic-solvable demand form to standard models of the income distribution. In purple is a double Pareto lognormal distribution with $(\alpha, \beta, \mu, \sigma) = (3, 1.43, 10.9, .45)$, which fits the US income distribution well as discussed in Subappendix D.1. In yellow is the common lognormal approximation, with parameters again fitted to the US income distribution, $(\mu, \sigma) = (10.5, .85)$. The quartic-solvable approximation, in blue is $P(q) = 10^4 \left(.56q^{-.59} + 18 - 49q^{.59} + 61q^{1.18} - 30q^{1.77}\right)$, where all numbers are rounded to two significant digits and the curve was fitted to the .001, .01, .25, .5, .75, .99 and .999 quantiles of the income distribution using Mathematica's non-linear fit function. The left panel shows the main part of the income distribution (quantiles on the x-axis, income on the y-axis) and the right panel shows its upper tail.

symbolic mathematical software can do the job with ease.[18] As a result, moving beyond the quadratic form decreases tractability only incrementally.

The permutations of possible behaviors become so large with such higher-order forms that it is beyond the scope of this paper to analyze them in any detail. Instead we simply consider a single example: we try to match some basic features of the US income distribution using a demand form yielding quartic solutions and admitting specific taxation.

Figure 3 shows two standard distributions used to model the US income distribution with standard calibrated values, the lognormal (yellow) and double Pareto-lognormal (dPln) (Reed, 2003; Reed and Jorgensen, 2004) (purple) distributions. The exact parameters are calibrated to the US income distribution as discussed in detail in Subappendix D.1. The left panel shows the main range of the distribution and the right panel focuses on the upper tail (up to the first quantile). The two distributions hug one another tightly except at the tails, where the dPln distribution (realistically) has Pareto rather than lognormal tails. The curve in blue is our quartically-solvable approximation; its equation is $P(q) = 10^4 \left(.56q^{-.59} + 18 - 49q^{.59} + 61q^{1.18} - 30q^{1.77}\right)$.

This form is a closer approximation to the dPln income approximation, to which it was fit given that this fits the US income distribution significantly better over a broad range

[18]For example, one can use Mathematica, or more generally the newly released Wolfram Language.

(Toda, 2014), than is the lognormal approximation. The only region over which it performs less well is upper middle-income, from about \$60k to \$160k, though even there the fit is quite close. In the true tail of the distribution, shown in the right panel, its fit is far closer than the lognormal distribution, given the latter's thin tails. Thus our closed-form solvable approximation appears to overall be a *better* fit to the US income distribution than is the canonical lognormal approximation. This offers a dramatic increase in tractability at no or negative cost to accuracy. In subsection 6.2 we show that fit similar to the lognormal except for q very close to 1 may be obtained using a quadratically solvable form.

Before moving on to discuss the logic behind these forms, we once more wish to emphasize that using any of these forms does not require the analyst to know anything about Laplace transforms or any other logic used to derive these forms. The properties discussed in this subsection are more than sufficient to allow the analyst to choose a functional form appropriate to her setting that nests the existing tractable forms discussed in Subsection 3.2 above. This form can then be solved using precisely the same pen-and-paper methods with which existing forms are solved; this is precisely what we mean by tractability. The broader theory of deriving these approximations based on the Laplace transform is needed only in coming up with new forms and understanding approximations methodologically, not in applying those proposed here.

4.3 Approximating arbitrary systems

Beyond the quartic form, higher-degree equilibrium conditions in terms of polynomials in q^{-t} generically do not have a solution by the method of radicals. However, pre-set functions for the solution for roots of even very high-degree polynomial equations are standardly embedded into mathematical software.[19] Thus, if one is willing to use such software, approximating equilibrium systems in terms of minimal degree polynomials allows for straightforward computation to replace the numerical searches for roots necessary to solve for equilibrium given the form of many standard computational equilibrium systems.

A very wide range of equilibrium systems may be approximated this way, even for a non-constant conduct parameter, such as symmetrically differentiated Bertrand competition derived from discrete choice models (Weyl and Fabinger, 2013) or supply chains with diverse cost functions at each stage; see Appendix A for details. However because the basic approach is independent of the particular system being approximated, here we focus our attention on approximating the demand function based on the most canonical discrete-choice statistical distribution, the Logistic distribution (viz. the difference between two Type I Extreme Value

[19] For more general purposes, Mathematica 10 includes fast numeric solutions of polynomial equations based on algebraic homotopy computation.

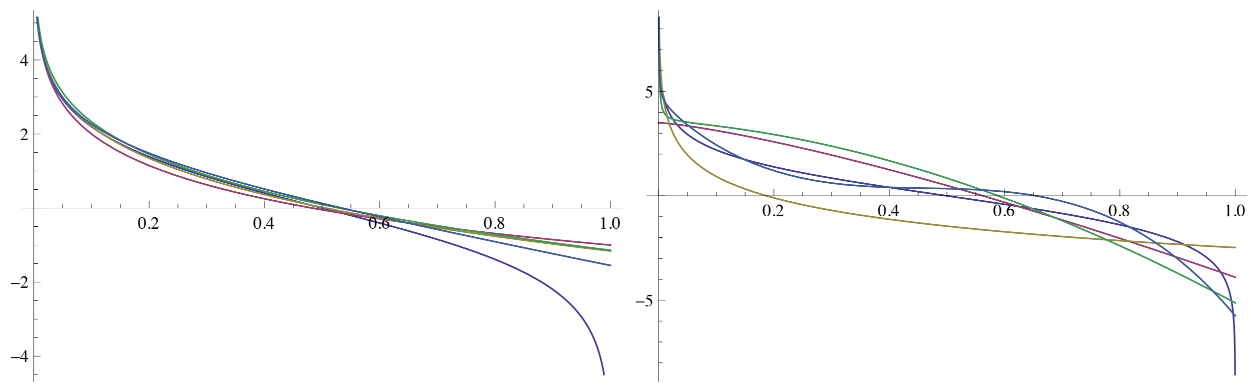

Figure 4: Approximating the logistic demand's quantile function with quartic-or-lower-solvable inverse demands functions.

distributions).

To begin, we take the Laplace inverse demand function corresponding to the standard Logistic distribution. This can be done by any standard Mathematical software package and expressions for many standard distributions other than the Logistic are given in Appendix D. Let $P_L(q)$ be the inverse demand function corresponding to the standard Logistic distribution (in the language of statistics, $P_L(q)$ is the quantile function). We obtain:

$$P_L(q) = -\log(q) - \sum_{i=1}^{\infty} \frac{q^i}{i} = \lim_{t \to 0} \frac{q^{-t} - 1}{t} - \sum_{i=1}^{\infty} \frac{q^i}{i}.$$

From the perspective of providing a tractable approximation, this form has both attractions and challenges. On the positive side, the Laplace inverse demand function takes and extremely simple form and consists of evenly-spaced mass points with the exception of the two points. On the negative side, for the first term to be accurate, weight must be placed on values of t very close to 0, as well as weight on a constant term, while to include terms as in the summation weight must be placed on terms with large values of t. This means that an accurate approximation requires a high-degree solution.

Figure 4 shows what low-degree approximations can and cannot accomplish. All forms are fit using Mathematica's nonlinear fit function. The left panel are fit to only points from below the median quartile (above the median draw) with a focus on points from the upper tail. The right panel is fit to a variety of points symmetric about the median. All demands fail to match the extreme concavity at high quantities as a result of attempting to match the convexity at low quantities while maintaining a low-degree solution. Approximations run from BP (Bulow and Pfleiderer, 1983) approximations (which is first-degree solvable and is crudest in its fit) up to the closest approximations which is quartic-solvable. Panels on the left refine, in the process, the fit at low quantities while panels on the right refine the fit

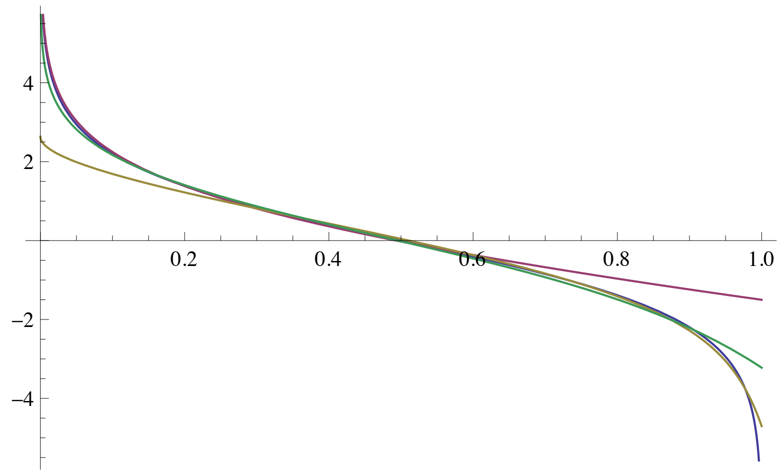

Figure 5: 100-order-solvable approximations to logistic demand.

evenly throughout the function.

This exercise shows some of the trade-offs in approximation through closed-form solvable demands that allow for specific taxation; similar issues arise in forms that are similarly solvable but do not admit specific taxation, except that the poor fit occurs for intermediate quantities rather than at the tails. Approximations fit to the upper tail upper tail capture the high curvature and therefore pass-through rate (arbitrarily close to unity) of the logistic distribution at low quantities. They do a much worse job, however, at middle and high quantities than do the curves on the right. Very tight fit is possible to the upper tail (low quantities) if the lower tail is neglected and a high-order solution is admitted. Only a loose fit to the lower tail (high quantities) is possible even with a quartic solution.

Much more is possible, however, if we are willing to admit systems that cannot be solved by radicals, but whose solutions all are stored in standard mathematical software. This is true for polynomials up to order 100 in Mathematica 10; beyond this order solving equations explicitly becomes quite costly, though more modern software versions may be able to handle higher-order polynomials at low cost. Kubler et al. (2014) provide a detailed discussion of computations of high-order polynomial solutions for economic problems.

Figure 5 shows three possible 100-order-solvable approximations to the logistic demand. One matches almost perfectly the upper tail (but is very inaccurate beyond the median), one almost perfectly the lower tail (but very poorly approximates the upper tail) and one fits both tails reasonably well. In our next draft we plan to make this approximation process more rigorous and supply software allowing researchers to do it at no cost to themselves.

While this case is clearly far less tractable than the low-order approximations discussed in the previous subsections, it may still be used free of almost any computational costs. With improved algorithms for this canonical problem that seem likely to be forthcoming, even better approximations may be used. This contrasts sharply with what would be required if the logistic function were used directly: its marginal revenue has no standard solution when equated to a constant or polynomial marginal cost and thus every solution requires a costly computational search for a root. We discuss further tractability advantages of these polynomial forms in the next section. While our focus here was on approximating one particular demand function, the same basic technique may be applied to any demand function, cost function or even more complicated equilibrium system, such as ones with non-constant conduct parameter generated by substitution patterns that change as a function of price. In all of these cases, our approach offers a much more fine-grained trade-off between flexibility and tractability than exists at present.

5 Formal theorems and properties of demand functions

We have recently developed a new point of view on demand functions, which makes pass-through properties that might seem hard to understand manifest in terms of Laplace transform and complete monotonicity properties.[20] We also found the most general functions invariant under average-marginal transformations. This section, when properly written, will discuss these new developments, as well as formal mathematical statements. In the current draft, formal specification of utility functions we consider may be found in Appendix B. The newly developed point of view on demand functions based on complete monotonicity may be found in Subappendix B.2. A theorem specifying most general functions invariant under average-marginal transformations is formulated in Appendix C.

6 International Trade Applications

The previous sections dealt, for expositional purposes, with a general characterization of tractability in the context of a highly stylized monopoly model. However the usefulness of these techniques arise from their ability to generate concrete insights in the richer and more realistic models of contemporary interest. In this section we therefore consider examples of such applied insights the approach can generate. For definiteness and coherence, we focus on applications to models most commonly employed in international trade as this is a field where models that can be explicitly aggregated into a single equation are particularly prevalent. However the approach is equally relevant in other active areas of economics where similar models are common such as industrial organization theory, perfect and imperfect

[20]We are grateful to James Heckman for pointing out to us the formal similarity of our equations and those used in duration analysis, which motivated the development of the complete monotonicity viewpoint.

competition in selection markets, asymmetric information bargaining, symmetric auctions, financial market microstructure, etc.

6.1 Non-monotonic value chain organization

Antràs and Chor (2013, henceforth AC) model the decisions of firms that manufacture complex, multistage products about vertical structure (in-sourcing v. out-sourcing) to address hold-up problems as in, e.g., Grossman and Hart (1986). Firms contributing to critical stages of the production process, where the marginal revenue associated with their contributions is very high, should be in-sourced to avoid hold-up, while those at more marginal stages of the production process should be out-sourced to avoid hold-up by the main firm that discourages quality production. For tractability, they assume a Dixit and Stiglitz (1977) structure, implying that marginal revenue is monotone and thus that either the early or the late stages of production are out-sourced, but not both. However, a more natural assumption may be that while marginal revenue rises at early production stages, as the product is first taking shape, it falls at later stages once it is nearly finished and thus its quality is reaching saturation, causing standard downward sloping demand to kick in. This would lead to out-sourcing of both early and late stages, an arguably more plausible conclusion. In this application we show how a model exhibiting these features can be formulated and solved as simply as that studied by AC.

We generalize AC's model by relaxing constant-elasticity assumptions for final demand functions and production functions.[21] We use notation compatible with the rest of this paper both for consistency and because it simplifies the exposition, but in footnotes provide explicit relations to the notation of AC.

A firm desires to provide a final good to its customers, but production requires a continuum of customized inputs each provided by a different supplier indexed by $j \in [0, 1]$. The firm must choose whether to contract with independent suppliers or integrate its suppliers into the firm.

The production of the final good is sequential, with one input added at a time. If production proceeds smoothly the *effective* (quality-adjusted) quantity q of the final good the integral of the quality contributed by intermediate input j, which we denote $q_s(j)$:[22]

[21] Our main focus is on more general demand functions. In the illustrative example we discuss we keep the production function unchanged, although closed-form solutions may be obtained straightforwardly by the same method with more general assumptions on technology.

[22] We use the symbol $\tilde{q}$ to refer to a quantity measure denoted q in AC, which is *distinct* from what we call quantity q. In order to recover AC's original model as a special case, we identify their output $\tilde{q}$ with $q^{1/\alpha}$, where $\alpha \in (0, 1)$ is a constant defined there. For the present discussion we do not need q to be linearly proportional to the number of units produced. It is just some measure of the output, which may or may not

$q = \int_0^1 q_s(j)\, dj$. Note that this effective quantity represents both the quantity of the good and its quality, and is a non-linear transformation of the notion of quantity used by AC. The lower is j, the further *upstream* a supplier is; that is, the more basic inputs to the good she supplies. However, if production is "disrupted" by the failure of some supplier, $\overline{j} \in [0, 1)$, to cooperate, then only the quality accumulated to that point in the chain is available, with all further quality-enhancement impossible, and thus effective quantity is $q = q(j) \equiv \int_0^{\overline{j}} q_s(j)\, dj$. The firm faces an inverse demand function $P(q)$, which need not be decreasing as is a standard inverse demand is, because, for example, consumers may have little willingness-to-pay for a very early stage product. Thus, over various ranges, increasing effective quantity may either raise or lower price.[23] If there is no disruption in production, $q = q(1)$.

Following the property rights theory of the firm (Grossman and Hart, 1986; Hart and Moore, 1990; Antràs, 2003), the key feature of AC's model is that input production requires relationship-specific investments that are subject to hold-up resulting from incomplete contracts. The marginal surplus from additional quality brought by supplier j, $MR\,(q(j))\, q_s(j)$ is therefore split between the firm and supplier j, where $MR = P + P'q$. In particular, the supplier receives a fraction $1 - \beta(j)$ that is interpreted as the firm's *effective* bargaining power, which may be adjusted by the organizational form: integrating the supplier lowers her bargaining power and leaving her independent raises it.

Effort by any individual supplier is subject to diminishing returns in producing $q_s(j)$. The cost function $C(q_s(j))$ associated with producing $q_s(j)$ is convex, and it is assumed to be the same for all suppliers.[24] Thus the first-order condition of supplier j equates the share of marginal revenue she bargains for with her marginal cost:

$$MC\,(q_s(j)) \equiv C'\,(q_s(j)) = [1 - \beta\,(j)]\, MR\,(q\,(j))\,. \tag{11}$$

The cost to the firm of obtaining a contribution $q_s(j)$ from supplier j is therefore the surplus it must leave in order to induce $q_s(j)$ to be produced, $q_s MC\,(q_s(j))$.[25]

be quite abstract. A similar statement applies to the customized intermediate input. Our measure $q_s\,(j)$ of a particular input is related to AC's measure $x(j)$ by $q_s\,(j) = \theta^\alpha (x\,(j))^\alpha$, where θ is a positive productivity parameter defined in their original paper.

[23]If physical quantity is a non-linear transformation of the effective quantity then the downward sloping nature of demand may be restored.

[24]The AC model corresponds to $C(q_s) = (q_s)^{1/\alpha} c/\theta$, where c and θ are positive constants defined in their paper. In our notation, the suppliers' cost is convex but their contributions towards the final output are linear. In the original paper the suppliers' cost is linear, but their contributions towards the final output have diminishing effects. These are two alternative interpretations of the same economic situation from the point of view of two different systems of notation. As mentioned before, in our interpretation, the product of a supplier is q_s, whereas in the original paper the supplier's product is x, related to q_s by $q_s\,(j) = \theta^\alpha (x\,(j))^\alpha$.

[25]Production of $q_s(j)$ requires $1 - \beta\,(j) = \frac{MC(q_s(j))}{MR(q(j))}$, so that the revenue received by supplier j, and thus

When the firm decides on the organizational structure of the supply chain, it chooses $\beta(j)$ optimally for each supplier to maximize its profits. Let us first focus on the *relaxed* problem where $\beta(j)$ may be adjusted freely and continuously. This provides most of the intuition for what happens when the firm is constrained to choose between two levels, corresponding to integration (in-sourcing) or separation (out-sourcing), for β. This relaxed problem may even be more realistic in some cases as various joint ventures and long-term relationships can lead to intermediate β values (Holmström and Roberts, 1998).[26] Note that by convexity $MC' > 0$, while each q_s makes a linearly separable contribution to q. Thus for any fixed q the firm wants to achieve, it does so most cheaply by setting all $q_s = q$ by Jensen's Inequality.[27] This observation allows us to avoid the complex Euler-Lagrange equation AC use to analyze their model. Thus Equation 11 becomes, at any optimum $q^\star$,

$$\beta^*(j) = 1 - \frac{MC(q^\star)}{MR(jq^\star)}. \tag{12}$$

From this we immediately see that $\beta^\star$ is co-monotone with MR: in regions where marginal revenue is increasing, $\beta^\star$ will be rising and conversely when marginal revenue is decreasing. The marginal revenue associated with constant elasticity demand is in a constant ratio to inverse demand and thus whenever inverse demand is declining in quantity, marginal revenue must be declining if it is positive. However if inverse demand is increasing in (effective) quantity, then so is marginal revenue (so long as it is positive). This implies AC's principal result that when revenue elasticity is less than unity the firm will tend out-source upstream and when revenue elasticity is less than unity the firm will tend to out-source downstream.[28]

This conclusion is somewhat counter-intuitive: we often think of firms out-sourcing both the very early stages of production and its final (retailing) stages in most industries. The reason for this conclusion may be the equally counter-intuitive assumption about the shape of demand. Constant elasticity preferences imply that either $P(q)$ is globally increasing or globally decreasing. However it seems natural to think that $P(q)$ would initially rise, as consumers are willing to pay very little for a product that is nowhere near completion, and would eventually fall as the product is completed according to the standard logic of downward-sloping demand. We now show how this logic can be embedded in assumptions

the cost to the firm, is $[1-\beta(j)]\, MR(q(j))\, q_s(j) = MC(q_s(j))\, q_s(j)$.

[26] AC consider the relaxed problem following analogous considerations in Antràs and Helpman (2004, 2008).

[27] Formal details of this argument will be supplied in Appendix E in the next draft.

[28] As BP argued, the tight connection between the slope of *marginal revenue* (which depends crucially on demand curvature) and the slope of inverse *demand* arises only under the constant elasticity demand form. Intuitively there is no reason the first and second derivatives of demand need have any particular relationship to one another. For example, using a natural experiment Einav et al. (Forthcoming) identify an (auction) inverse demand curve that is decreasing but has increasing marginal revenue and an increasing, but sufficiently concave inverse demand P yields declining marginal revenue, as we will shortly see.

that yield as simple a solution to the model as does constant elasticity demand, while yielding these arguably more realistic results.

Equation 12 implies that the surplus left to each supplier is $q_s MC(q)$ and thus total cost is $qMC(q)$. Thus the problem reduces to choosing q to maximize revenue $qP(q)$ less cost $qMC(q)$, giving first-order condition

$$MR(q) = MC\left(q\right) + q\, MC'\left(q\right). \tag{13}$$

This differs from the familiar neoclassical first-order condition $MR\left(q\right) = MC\left(q\right)$ only by the presence of the (positive) term $q\, MC'\left(q\right)$. Note that $MC + qMC'$ bears the same relationship to MC that MC bears to AC; this equation therefore similarly inherits the tractability properties of the standard monopoly problem.

Let us now consider the class of all linearly solvable examples for this model, namely $P(q) = p_{-t}q^t + p_{-u}q^u$ and $MC(q) = mc_{-t}q^t + mc_{-u}q^u$. This includes AC's specification as the special case when $p_t = 0$ and $mc_u = 0$ so that each has constant elasticity.[29] However many other cases can also be considered and we focus on one that corresponds to our intuition that price should initially increase and then decline in q. In particular, we assume that $t, u, mc_{-u}, p_{-t} > 0 = mc_{-t} > p_{-u}$ and that $u > t$ so that the first term of the inverse demand dominates at small quantities while the second dominates at large quantities.[30]

The expression resulting for $\beta^*\left(j\right)$ is:

$$\beta^*\left(j\right) = 1 - \frac{1}{\left(1+u\right)\left[\left(1 - \frac{p_{-u}}{mc_{-u}}\right)j^t + \frac{p_{-u}}{mc_{-u}}j^u\right]}. \tag{14}$$

Note that because $mc_{-u} > 0 > p_{-u}$, the numerator and first denominator term in the ratio are positive and the second denominator term is negative. This implies that at small q, where j^t dominates, $\beta^\star$ increases in j, while at large j it decreasing in j. In the AC complements case when $p_{-u} = 0$, or even if p_{-u} is sufficiently small, this large j behavior is never manifested and all out-sourcing (low $\beta^\star$) occurs at early stages. Also note that only the ratio of coefficients $\frac{p_{-u}}{mc_{-u}}$ matters for the sourcing pattern; p_{-t} is irrelevant, as the the joint level of p_{-u} and mc_{-u}.

However, for many parameters an inverted U-shape emerges. For example, Figure 6 shows the case when $t = .35, u = .7, p_{-t} = 1.8, \frac{p_{-u}}{mc_{-u}} = -4$. The curve corresponds to the shape

[29] In particular, in their notation, AC have $t = \frac{1}{\alpha}$, $u = 1 + \frac{\rho}{\alpha}$, $mc_t = c/\alpha\theta$ and $p_u = A^{1-\rho}$, where θ and ρ is are positive constants defined in AC, not to be confused with the pass-through rate denoted by ρ or the conduct parameter denoted by θ in other parts of this paper.

[30] We focus on this case because, interestingly, if $mc_t \neq 0$ then if β has any lower bound there is no positive production solution yielding positive profits because it becomes extremely costly to incentivize the first suppliers.

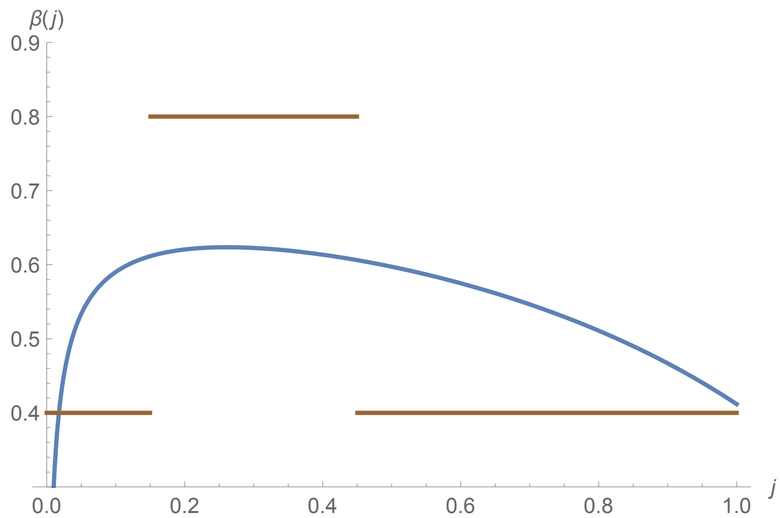

Figure 6: Optimal relaxed and restricted $\beta^{\star}$ in the AC model when $t = .35, u = .7, \frac{p_{-u}}{mc_{-u}} = -4$.

of the relaxed solution. Depending on precisely which values of β we take in-sourcing and out-sourcing to correspond to, this can lead to in-sourcing in the middle of the production and out-sourcing at either end.

To investigate this we turn now to the restricted case. As we show in Subappendix E.1 a nearly closed-form solution to this problem, one with only a single, univariate explicit equation to be solved numerically, is possible when the Laplace marginal cost and price include a mass point at 0 (admit a closed-form inverse solution). This equation determines the overall level of production, not the pattern of sourcing which can be solved for fully analytically. This is not the case with the form we chose above, which is the simplest one that gives the non-monotone effect while being closed-form solvable in the relaxed problem. However, even in our case, many steps can be performed analytically, though it does require a two-dimensional numerical search. We use this procedure to construct optimal constrained solutions for the case when out-sourcing gives $\beta_O = .8$ and in-sourcing gives $\beta_I = .4$. This is illustrated by the lines in Figure 6, which show the constrained optimum. This gives the same qualitative answer as the relaxed problem, unsurprisingly.

The comparison between the case here and that we consider in the appendix illustrates the trade-off between the greater improved flexibility-tractability trade-off possible without a mass point at 0 and the additional tractability benefits for some problems afforded by including a mass point at 0.

6.2 Comparative statics of employment bargaining

Stole and Zwiebel (1996a,b, henceforth SZ) study a model of wage bargaining where firms employ workers mutually at-will and where hiring new workers happens with delay. This leads to "labor hoarding" (viz. over-employment relative to the neo-classical benchmark) in order make each worker more expendable and thus weaken their bargaining position. This distortion helps offset firm product and labor market power. However, the model has been primarily applied (Helpman et al., 2010; Helpman and Itskhoki, 2010; Helpman et al., 2014) in settings with constant elasticity demand and power law technology, as in the last subsection. In this context, the quantity of labor hoarding is unaffected by the prevailing state of the economic cycle, as measured by the size of the workers' outside option relative to demand. Furthermore, because the model is governed by a complex differential equation, most intuitions about the model arise from these very special cases and may be misleading as a result.

In fact, we show that stepping just slightly outside this particular class yields qualitatively different results. In particular if one allows for a more plausible bell-shaped demand, labor hoarding is *counter-cyclical*, implying that the possibility of bargaining dampens economic fluctuations. We derive this result by mechanically applying our tractable forms; in fact, we still do not understand the intuition behind this result. Such insights unguided by intuition are possible because the fairly elaborate equilibrium of the Stole-Zwiebel model is simple when framed in terms of the Laplace first-order conditions of a standard monopoly model. Thus, despite how different this model appears to be from the neoclassical monopoly model we focused on in Section 2, it can be tractably solved precisely under the conditions (on demand and cost) as the monopoly model.

In SZ, at the beginning of a period, a firm hires workers, each of whom supplies one unit of labor if employed.[31] When this process has been completed but before production takes place, the workers are free to bargain over their wages for this period. At that time the firm cannot hire any additional workers, so if any bargaining is not successful and any worker leaves the firm, fewer workers will be available for production in this period. Moreover, after the worker's departure, the remaining employees are free to renegotiate their wages, and in principle the process may continue until the firm loses all its employees. Assuming its revenues are concave in labor employed, this gives the firm an incentive to "over-employ" or *hoard* workers as hiring more workers makes holding a marginal worker less valuable to the firm and thus reduces workers' bargaining power.

If the bargaining weight of the worker relative to that of the firm's owner is λ, then

[31] The model is formally dynamic but is usually studied in its steady state as described here.

the relationship surplus splitting condition is $S_w = \lambda S_f$. The worker's surplus is simply the equilibrium wage corresponding to the current employment level minus the outside option: $S_w = W(l) - W_0$, where W is the wage as a function of l, the labor supplied. For expositional simplicity, we assume the firm transforms labor into output one-for-one, though analytic solutions also exist for any power law production technology when $\lambda = 1$ and in other cases. Thus we assume $q = l$ and henceforth use q as our primary variable analysis for consistency with previous sections.

The firm faces inverse demand $P(q)$ and thus its profits are $\Pi(q) = [P(q) - W(q)]\, q$. The firm's surplus from hiring an additional worker is then $\Pi'(q)$. This gives differential equation

$$W(q) - W_0 = \lambda MR(q) + \lambda \left(W(q)\, q\right)' \Rightarrow \lambda (W(q)\, q^{1+\frac{1}{\lambda}})' = q^{\frac{1}{\lambda}} \left(\lambda MR(q) + W_0\right),$$

where $MR \equiv P + P'q$ and the implication can be verified by simple algebra and is a standard transformation for an ordinary differential equation of this class. Integrating both of the sides of the equation, imposing the boundary condition that the wage bill shrinks to 0 at $q = 0$ and solving out yields wages

$$W(q) = q^{-\left(1+\frac{1}{\lambda}\right)} \int_0^q x^{\frac{1}{\lambda}} MR(x)dx + \frac{W_0}{1+\lambda}$$

and thus profits

$$\Pi(q) = P(q)q - q^{-\frac{1}{\lambda}} \int_0^q x^{\frac{1}{\lambda}} MR(x)dx - \frac{W_0}{1+\lambda}.$$

The firm's optimal q solves its first-order condition, $\Pi'(q) = 0$, which, after some algebraic manipulations, is

$$\frac{(1+\lambda)\int_0^q x^{\frac{1}{\lambda}} MR(x)dx}{\lambda q^{1+\frac{1}{\lambda}}} = W_0. \tag{15}$$

Let us define (relative) labor hoarding as $h \equiv \frac{q^\star - q^{\star\star}}{q^{\star\star}}$, where $q^\star$ is SZ employment and $q^{\star\star}$ is the employment level that a neoclassical firm with identical technology would choose: $MR(q^{\star\star}) = W_0$. Combining these definitions with (15) gives a useful condition for h in terms of the equilibrium employment level $q^\star$:

$$MR\left(\frac{q^\star}{1+h}\right) = \frac{(1+\lambda)\int_0^{q^\star} x^{\frac{1}{\lambda}} MR(x)dx}{\lambda (q^\star)^{1+\frac{1}{\lambda}}}. \tag{16}$$

Note that this equation, and Equation 15, involves only a) marginal revenue and b) integrals of it multiplied by a power of q and then divided by one power higher of q. It can easily be shown that the support of the Laplace marginal revenue is preserved by this transformation using essentially the same argument we used in Section 2 to show this support was shifted by exactly one unit in when consumer surplus is calculated. This implies that Equations 15 and 16 have precisely the same tractability characterization as does the basic monopoly model we studied in Section 2.[32]

We are interested in the response of hoarding to changes in the economic cycle. In particular, in good economic times the real value of W_0 falls so that firms optimally produce more. Given the complexity of Equation 16 from any perspective other than our tractable forms, we investigate it using these forms, following Helpman et al. (2010) who study the model under constant elasticity demand. First consider the BP (Bulow and Pfleiderer, 1983) class, $P(q) = p_0 + p_t q^{-t}$, which nests the constant elasticity case when $p_0 = 0$. Solving Equation 16 for h yields [33]

$$h = \left(\frac{1+\lambda}{1+\lambda - t\lambda}\right)^{\frac{1}{t}} - 1. \tag{17}$$

Thus hoarding is constant in $q^\star$ and consequently in W_0. Thus under constant elasticity demand, or more broadly under the BP class of demand, the economic cycle (the nominal outside option) has *no effect* on relative hoarding. It can easily be shown that h monotonically increases in t, so that the less concave demand (and thus profits) are, the more hoarding occurs. To illustrate this Figure 7 illustrates the value of Expression 17 for $\lambda = 1$ as a function of t running from -10 (very concave demand) to $t = 1$ (the maximally convex demand consistent with the firm's second-order conditions). Hoarding clearly rises with demand convexity.

We found this result counter-intuitive, as we believed, building off the intuition supplied by SZ about the relationship beween the "front-loading" that drives hoarding and concavity, that labor hoarding was driven by concavity in the firm's profit function. Instead it appears that the reverse is the case: at least This shows one advantage of considering an explicit functional forms: they help correct false intuitions. In particular, because t clearly parameterizes concavity the comparative static has a natural interpretation.

Yet this new intuition suggests that the hoarding may not be constant over the economic cycle if, during that cycle, the curvature of firm profits change. For example, if during booms broad parts of the population are served and during recessions only wealthier individuals are served, then labor hoarding should be counter-cyclical as the distribution of income among the wealthy is more convex than among the middle-class and poor.

[32] Note that Equation 15 also involves a constant and thus only our tractable forms with a constant term will maintain their tractability in this model. This is why we focus on this class below.

[33] (16)$\Rightarrow p_t\,(q^\star)^{-t}(1+h)^t = \frac{1+\lambda}{\lambda} p_t\,(q^\star)^{-\left(1+\frac{1}{\lambda}\right)} \int_0^{q^\star} x^{\frac{1}{\lambda}} x^{-t} dx \Rightarrow (q^\star)^{-t}(1+h)^t = \frac{1+\lambda}{\lambda}\frac{1}{\frac{1}{\lambda}+1-t}(q^\star)^{-t} \Rightarrow$(17).

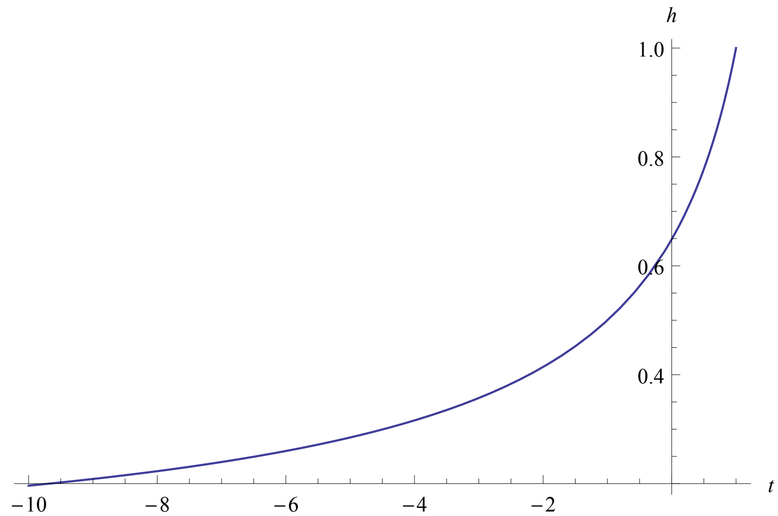

Figure 7: Labor hoarding as a function of demand curvature in the BP class; demand and thus profit is increasingly convex as t rises.

To analyze this issue we considered one of our quadratically solvable forms that does a reasonable job of matching salient features of the income distribution:

$$P(q) = p_t q^{-t} + p_0 + p_{-t} q^t.$$

Using the standard quadratic formula, one can obtain explicit expressions for $q^\star$ and $q^{\star\star}$ by solving (15) and $MR(q^{\star\star}) = W_0$, respectively:

$$h = \left(\frac{\sqrt{1 - \frac{\beta(1-t)^2 p_t p_{-t}}{(W_0 - p_0)^2}} \pm 1}{\sqrt{1 - \frac{(1-t)^2 p_t p_{-t}}{(W_0 - p_0)^2}} \pm 1} \right)^{\frac{1}{t}} \gamma - 1, \tag{18}$$

where the newly defined constants are just combinations of t and λ:

$$\beta \equiv \frac{(1+\lambda)^2}{(1+\lambda-t\lambda)(1+\lambda+t\lambda)} \qquad \gamma \equiv \left(\frac{1+\lambda+t\lambda}{1+\lambda}\right)^{\frac{1}{t}}$$

and the $\pm$ is positive if $W_0 < p_0$ and negative otherwise.

We can use Equation 18 to study the cyclic behavior of labor hoarding by considering the effect of a change in W_0 on h. We interpret a reduction in W_0, or equivalently a multiplicative scaling up of P, to be a boom (as it leads to higher production) and a rise in W_0 to be a recession. This expression can be shown algebraically to be always increasing. This implies that in a recession hoarding rises, cushioning the force of the recession. This contrasts with the standard intuition that unions exacerbate recessions by creating nominal

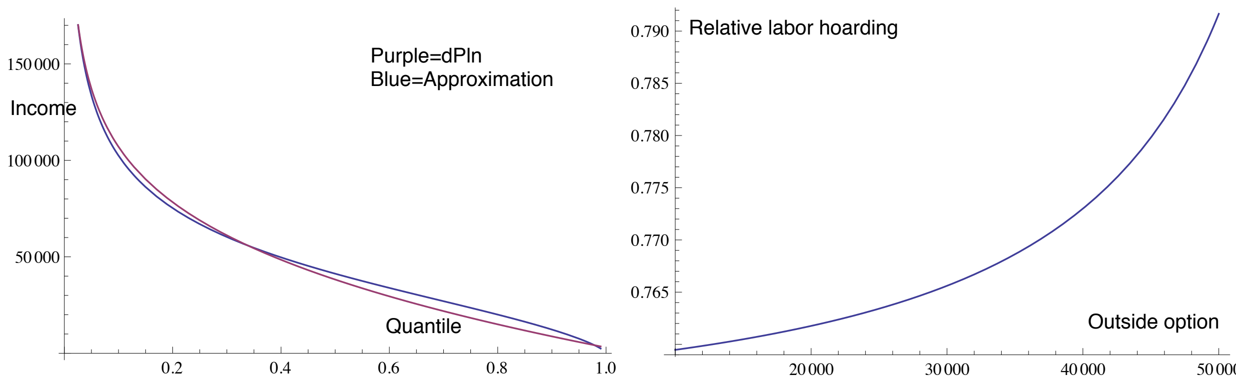

Figure 8: Left: US income distribution approximated by a dPln distribution with parameters $(\alpha, \beta, \mu, \sigma) = (3, 1.43, 10.9, .45)$ (income on the y-axis, quatile on the x-axis) and an approximation to it $P(q) = 10^4 \cdot (2.2q^{-.42} + 9.1 - 11q^{.42})$. Right: relative labor hoarding in the SZ model with $\lambda = 1$ and demand given by the approximation for $W_0 \in 10^4 \cdot [1, 5]$

wage rigidity, suggesting the effects of individual workers' bargaining may have qualitatively different comparative statics than collective bargaining does.

The size of this effect can be seen in a calibrated example, where P is calibrated to the US income distribution, λ is set, as is conventional, to 1 and we vary W_0.[34]

Figure 8 shows the results. The left panel shows the income distribution, as approximated by a dPln with $(\alpha, \beta, \mu, \sigma) = (3, 1.43, 10.9, .45)$ as in Subsection 4.2 and a (quite close) approximation (by non-linear fitting in Mathematica) to it using our tractable form with $t = .42, p_t = 2.2 \cdot 10^4, p_0 = 9.1 \cdot 10^4, p_{-t} = -1.1 \cdot 10^5$. The right-hand panel shows h as a function of W_0 as the latter range from $1 \cdot 10^5$ to $5 \cdot 10^5$. Hoarding is quite large, on the order of 70-80%; however its comparative statics are more subtle. It rises by about 2.5 percentage points when the outside option rises from \$30k to \$50k, perhaps a reasonable range of variation over the economic cycle. Thus the BP approximation of constancy appears not to be very far off. However, these effects may be non-negligible from the perspective of shifting employment, and cushioning unemployment during recessions, in the aggregate economy, and thus are worth including to get a realistic portrait of cyclic unemployment.

6.3 Monopolistic competition model

In our final application, we extend a framework in which the combination of constant marginal cost and Bulow-Pfleiderer demand is used perhaps most frequently: monopolistic competition building on the model of Dixit and Stiglitz (1977). The applications of this framework are ubiquitous and we consider three of the most common: the models of

[34]Note, however, that $\lambda = 1$ has no special tractability advantage in our class when technology is linear. However, it does for more general power-law technology.

Krugman (1980), Melitz (2003) and Melitz and Ottaviano (2008). These international trade papers all feature constant marginal cost of production and transportation and either constant elasticity or linear demand. A number of recent papers have noted the substantive restrictions imposed by the assumptions of BP demand and have conducted analyses of topics to which these models are typically applied that relax the demand-side restrictions, e.g. Zhelobodko et al. (2012); Arkolakis et al. (2012); Bertoletti and Etro (2013); Bertoletti and Epifani (2014); Bertoletti and Etro (2014); Dhingra and Morrow (2014); Kichko et al. (2014); Mrázová and Neary (2014a,c,b); Parenti et al. (2014).[35] To do so, these papers sacrifice analytic tractability and thus the ability to consider many of the quantitative aggregate equilibrium questions with which the literature has traditionally been concerned. At the same time, they typically, with the notable exception of Zhelobodko et al., maintain the assumption of constant marginal cost of production. Whenever trade costs are considered, these are either constant per-unit costs, or much more frequently *iceberg* type costs (i.e. losses of the transported good itself), which under the assumption of constant marginal costs of production also imply constant marginal costs of transport.[36]

In this section we show how these limitations can be overcome using our technology. In particular we show how all three of these canonical trade models with monopolistic competition fall squarely into our framework and thus can be studied flexibly on both the demand and cost side while maintaining to a large extent the same tractability and aggregability of the original papers. Thus *there is often no trade-off*, or at most a gradual one, between the tractability of the original frameworks and the richer insights yielded by the growing literature relaxing the assumption of a BP demand form. At the same time, restrictions on production and transport cost may be relaxed just as easily as those on demand.

In a future draft of this paper we plan to discuss instances where the usual functional form assumptions are inconsistent with empirical observations, and where relaxing these assumptions is particularly desirable because it leads to novel economic insights. In the current draft, we only describe the mathematical solutions to the models.

[35] Behrens and Murata (2007, 2012) use a tractable approach based on exponential inverse demand functions, for which equilibrium variables may be expressed in terms of the Lambert W function. Note also that the original Dixit and Stiglitz (1977) paper does discuss non-constant elasticity demand, albeit in a way that is not explicitly tractable.

[36] For recent empirical evidence documenting the size of non-iceberg trade costs, see Hummels and Skiba (2004) and Irarrazabal, Moxnes and Opromolla (2014).

6.3.1 Tractable generalizations of the Dixit-Stiglitz framework

In the baseline monopolistic competition model consumers derive their utility from a continuum of varieties $\omega \in \Omega$ of a single heterogeneous good:

$$U_\Omega = \int_\Omega u_\omega \left(q_\omega\right) \, d\omega. \tag{19}$$

In the original Dixit-Stiglitz model with constant elasticity of substitution σ, $u_\omega(q_\omega)$ is a power of the consumed quantities q_ω: $u_\omega(q_\omega) \propto q_\omega^{1-1/\sigma}$. In our generalization $u\left(q_\omega\right)$ is assumed to be a function of a combination different powers of q_ω . More explicitly, consumer optimization requires that marginal utility of extra spending is equalized across varieties: $u'_\omega\left(q_\omega\right) = \lambda P_\omega$, where P_ω is the price of variety ω and λ is a Lagrange multiplier related to consumers' wealth. To ensure tractability, we let the residual inverse demand $P_\omega\left(q_\omega\right) = {}^{u'_\omega(q_\omega)}/_{\lambda}$ and the corresponding revenue $R_\omega\left(q_\omega\right)$ be linear combinations of equally-spaced powers of q_ω:

$$P_\omega\left(q_\omega\right) = \sum_{t\in T} p_{\omega,t} q_\omega^{-t}, \quad R_\omega\left(q_\omega\right) = \sum_{t\in T} p_{\omega,t} q_\omega^{1-t}$$

for some finite and evenly-spaced set T, with the number of elements of T determining the precise degree of tractability. For convenience of notation, we choose a numéraire in a way that keeps $P_\omega\left(q_\omega\right)$ for a given q_ω independent of macroeconomic circumstances.

Each variety of the differentiated good is produced by a single firm. We assume that the marginal cost and average cost of production can be written as

$$MC_\omega(q) = \sum_{t\in T} mc_{\omega,t} q_\omega^{-t}, \quad AC_\omega(q) = \sum_{t\in T\cup\{1\}} ac_{\omega,t} q_\omega^{-t},$$

where $mc_{\omega,t} = (1-t)\, ac_{\omega,t}$. A constant component of average cost (and marginal cost) would correspond to $ac_{\omega,0}$ and a fixed cost would correspond to $ac_{\omega,1}$. However given the generality possible here we do not necessarily have to assume that these components are present in all models under consideration.

6.3.2 Flexible Krugman model

The Krugman (1980) model of trade, featuring monopolistic competition and free entry of identical single-product firms, may be solved explicitly for the tractable demand and cost functions mentioned above, not just constant-elasticity demand and constant marginal cost specified in the original paper. Here we consider these solutions in the case of two symmetric countries, which leads to a symmetric equilibrium.

There is a continuum of identical consumers with preferences as in Equation 19 who earn labor income. The amount of labor a firm needs to hire in order to produce quantity q may be split into a fixed part f and a variable part $L(q)$ that vanishes at zero quantity. Both $L(q)$ and the revenue function $R(q)$ are assumed to allow for a linear term. The firm only uses labor for production, so its total cost is $w(L(q)+f)$, where w is the competitive wage rate. Having produced quantity q, the firm splits it into q_d to be sold domestically, and τq_x to be shipped abroad. Due to iceberg-type trade costs ($\tau \geq 1$), a fixed fraction of the shipped good is lost during transport, and only quantity q_x is received in the other country. (Non-iceberg trade costs are considered in the appendix.) Let us denote the equilibrium level of marginal cost, measure of firms, international trade flows, and welfare by $MC^\star$, $N^\star$, $X^\star$, and $W^\star$, respectively, and similarly for other variables. The total labor endowment of one of the two symmetric economies is L_E.

Observation 8. *There exists an explicit map $MC^\star \to (f, q_d^\star, q_x^\star, w^\star)$ and an explicit map $(MC^\star, L_E) \to (N^\star, X^\star, W^\star)$. These relationships represent a closed-form solution to the model in terms of $MC^\star$ and exogenous parameters.*

We provide details in Subappendix E.3. Analogous solutions for several symmetric countries with fully symmetric trade costs or even asymmetric countries may be obtained straightforwardly. Additionally, it is possible to relax the assumption of exogenous labor supply.

6.3.3 Flexible Melitz model

The Melitz (2003) model is again based on monopolistic competition and assumes constant elasticity of substitution between heterogeneous-good varieties. Relative to the Krugman (1980) model, it introduced a novel channel for welfare gains from trade, namely increased average firm productivity resulting from trade liberalization or analogous decreases in trade costs. Here we generalize the model to allow for more flexible demand functions, non-constant marginal costs of production, and trade costs that may have components that are neither iceberg-type nor constant per unit.

Single country. For clarity of exposition, we first describe the flexible and tractable version of the Melitz model in the case of a single country and later discuss its generalization. Just like the Krugman model, it involves two types of agents: monopolistic single-product firms and identical consumers, who supply their labor in a competitive labor market and consume the firms' products.[37]

[37] For simplicity, consumers do not discount future, although it would be easy to incorporate an explicit discount factor. Formally, the model includes an infinite number of periods, but it may be thought of as a static model because the equilibrium is independent of time.

Labor is the only factor of production: all costs have the interpretation of labor costs and are proportional to a competitive wage rate w. Each heterogeneous-good variety is produced by a unique single-product firm, which uses its monopolistic market power to set marginal revenue equal to marginal cost. Demand and costs are specified tractably as discussed above; this time we do not need to assume that variable cost and revenue functions allow for a linear term.

If the firm is not able to make positive profits, it is free to exit the industry. In situations of main interest, this endogenous channel of exit is active: there exist firms that are indifferent between production and exit. There is also an exogenous channel of exit: in every period with probability δ_e the firm is forced to permanently shut down.

Entry into the industry is unrestricted, but comes at a fixed one-time cost wf_e. Only after paying this fixed cost, the entering firm observes a characteristic a, drawn from a distribution with cumulative distribution function $G(a)$, that influences the firm's cost function. In the original Melitz model the constant marginal cost of production is equal to wa. Here we leave the specification more general, while maintaining the convention that increasing a increases the firm's cost at any positive quantity q. In expectation, the stream of the firm's profits must exactly compensate the (risk-neutral) owner for the entry cost, which implies the *unrestricted entry condition* $wf_e = \mathbb{E}\Pi(q;a)/\delta_e$, with the profit $\Pi(q;a)$ evaluated at the optimal quantity.[38]

The amount of labor needed to produce quantity q is $L(q;a) + f$, where $L(q;a)$ corresponds to variable cost ($L(0;a) = 0$) and f to a fixed cost. $L(q;a)$ is assumed to be tractable with respect to q, but also with respect to a.[39] In terms of the labor requirement function $L(q;a)$, the firm profit maximization condition and the z*ero cutoff profit condition* are $R'(q) = wL'(q;a)$ and $R(q_c) = wL(q_c;a_c) + wf$, where q_c and a_c correspond to a *cutoff firm*, i.e. a firm that is in equilibrium indifferent between exiting and staying in the industry. We denote by L_E, $M^\star$,$M_e^\star$, $W^\star$ the labor endowment, and the equilibrium measure of firms, measure of entering firms, and level of welfare, respectively.

Observation 9. *There exists an explicit map* $(q_c, G) \to (f_e, w^\star, a_c)$ *and an explicit map* $(q_c, L_E, G) \to (M^\star, M_e^\star, W^\star)$*. These relationships represent a closed-form solution to the model in terms of* q_c *and exogenous parameters.*

[38]In the case of a single country, the profit is simply $\Pi(q;a) = q[P(q;a) - AC(q;a)]$. Also note that the unrestricted entry condition is often referred to as the *free entry condition*, but here we avoid this term since there is a positive entry cost.

[39]For example, the function $L(q;a)$ could be linear in a, as would be the case in the original Melitz model. A simple example of a tractable choice of functional forms is $L(q) = \tilde{L}(q) + a\hat{L}(q)$, $\hat{L}(q) \equiv q^t$, $\tilde{L}(q) \equiv \tilde{\ell}_t q^{-t} + \tilde{\ell}_u q^{-u}$, and $R(q) = r_t q^{-t} + r_u q^{-u}$.

Again, we provide details in Subappendix E.3. Since these maps allow for an arbitrary cumulative distribution function G, they involve integrals. For truncated Pareto G and $L(q;a)$ linear in a, there exist explicit expressions for these integrals, often involving special functions. As in the case of the Krugman model, it is possible to relax the assumption of exogenous labor supply.

Note also that comparative statics that assumes constant f_e may be also performed explicitly in terms of q_c by constructing variations of the model's parameters that leave the value of f_e intact.

Two symmetric countries, with non-iceberg and iceberg trade costs. Just like in the case of the flexible Krugman model, it is convenient to write the model in terms of equilibrium marginal cost, which this time is firm-specific and also depends on the firm's chosen export status. For tractability we will need the revenue function $R(q)$ and the production labor requirement function $L(q;a)$ to allow for a linear term. The same is true for labor corresponding to the non-iceberg trade costs, here denoted by $L_T(q_x)$. As in the original Melitz (2003) paper, we consider equilibria characterized by two cutoffs, here denoted a_1 and a_2, such that least productive firms with $a > a_1$ exit, more productive firms with $a \in (a_2, a_1]$ serve only their domestic market, and most productive firms with $a \leq a_2$ serve both countries. In general, we denote the equilibrium marginal cost of a non-exporting firm as $MC_n^\star$ and that of an exporting firm as $MC_x^\star$. Variables corresponding to the two cutoffs are distinguished by subscripts 1 and 2, so for example $MC_{1n}^\star$ is the optimal marginal cost of a firm with $a = a_1$, and $MC_{2x}^\star$ and $MC_{2n}^\star$ are optimal marginal costs of a firm with $a = a_2$ that decides to export or not to export, respectively. We denote by $M_x^\star$ and $X^\star$ the equilibrium measure of exporting firms and international trade flows.

Observation 10. *There exists an explicit map* $(MC_{1d}^\star, MC_{2x}^\star, G) \to (f_x, f_e, w^\star, a_1, a_2)$ *and an explicit map* $(MC_{1d}^\star, MC_{2x}^\star, L_E, G) \to (M^\star, M_x^\star, M_e^\star, X^\star, W^\star)$. *These relationships represent a closed-form solution to the model in terms of* $MC_{1d}^\star$, $MC_{2x}^\star$, *and exogenous parameters.*

Again we discuss details in Subappendix E.3. As in the case of a single country, for truncated Pareto G and $L(q;a)$ linear in a, there exist explicit expressions for the relevant integrals, often involving special functions. Likewise, the assumption of exogenous labor supply may be relaxed.

We see that the Melitz model, which originally involved a very special specification of demand, production costs and transportation costs, is not an isolated tractable model, but rather a first instance of a wide class of models that describe heterogeneous firms in a tractable way that leads to closed-form solutions. As mentioned before, future versions of this draft will discuss some of the cases where this greater flexibility is most needed.

6.3.4 Flexible Melitz/Melitz-Ottaviano model with non-separable utility

If we relax the assumption of separable utility, we obtain models that are rich enough to contain also the Melitz and Ottaviano (2008) model as a special case (provided we are willing to add a homogeneous good). Depending on the exact case under consideration, we may obtain explicit solutions as in the separable utility case. However, even in the other cases, one can explicitly perform aggregation under similar assumptions as previously. This means that it is still possible to reduce the problem involving an infinite number of heterogeneous firms to a problem with a finite number of equations for aggregate variables, which are straightforward to solve numerically. We plan to make these observations concrete in a future draft of this paper.

7 Conclusion

This paper makes three contributions. First, it identifies classes of (static, deterministic) equilibrium systems that are analytically tractable, nest nearly every known tractable equilibrium systems as special cases and allow much greater flexibility than existing systems. Second, it shows how these equilibrium systems can be used to overcome implausible substantive assumptions imposed by existing tractable systems. Finally, it uses this framework to study, and even in some cases overturn, the robustness and realism of prominent conclusions in the theory of international trade.

Our work suggests several directions for future research. First, the process of adding additional terms to match features of an equilibrium system closely resembles sieve approximation in non-parametric statistics. Determining "optimal" procedures for using increasingly less tractable equilibrium systems to approximate empirical equilibrium systems as statistical precision increases seems a natural method to maintain maximal tractability and flexibility on incidence features for a given statistical resolution.

Second, inverse Laplace transforms may be a useful representation of various properties of demand, cost and equilibrium systems. In on-going work (Fabinger and Weyl, 2014) we are using characterizations of demand curves in terms of properties of inverse Laplace transforms to determine sufficient conditions for price discrimination to be welfare enhancing. The properties we use and others stated in terms of inverse Laplace transforms may be useful hypotheses for other theoretical results.

Third, while we focused on international trade in this paper, many results in the theory of industrial organization are based on explicitly-solved equilibrium systems. Revisiting these results in light of more flexible but equally soluble equilibrium systems could be enlightening. Conversely models of optimal taxation take similar forms to monopoly models but have been notoriously difficult to solve in closed for, except for with unrealistic uniform or Pareto talent distributions. Lockwood et al. (2014) plan to include in their next draft an explicitly solved, realistic example based on our tractable forms.

Finally, we considered only models that can eventually be reduced to a single equation (or a few closely-related equations). However our approach here does not directly apply to models, such as general equilibrium models with several asymmetric sectors or imperfectly competitive models with multiple choice dimensions of quality or asymmetric firms, characterized by several, non-aggregable equations. Nonetheless, as we discuss briefly in Appendix F, the techniques from applied mathematics we use apply in these richer settings, albeit with an increasingly unattractive trade-off between flexibility and tractability as the number of equations increase. This approach should be more fully developed.

More broadly, the analysis of imperfect competition has become somewhat divided between approaches that employ simple, explicitly soluble systems, like those we discussed in Subsection 3.2, and other work that focuses on more complex, realistic systems that require significant computational cost to analyze. The approximation approach we developed here can bridge between these two extremes by allowing systems that match key policy-relevant features of empirical structures while remaining nearly as tractable as the systems that are usually employed for their convenience rather than their realism. Researchers may then choose more freely which position along the tractibility-plausibility spectrum is most appropriate to their purposes, as well as choosing the features to most closely approximate depending on the policy question of interest.

References

Adachi, Takanori and Takeshi Ebina, "Cost Pass-Through and Inverse Demand Curvature in Vertical Relationships with Upstream and Downstream Competition," *Economics Letters*, 2014, *124* (3), 465–468.

_ **and** _ , "Double Marginalization and Cost Pass-Through: Weyl–Fabinger and Cowan meet Spengler and Bresnahan–Reiss," *Economics Letters*, 2014, *122* (2), 170–175.

Antràs, Pol, "Firms, Contracts, and Trade Structure," *Quarterly Journal of Economics*, 2003, *118* (4), 1375–1418.

_ **and Davin Chor**, "Organizing the Global Value Chain," *Econometrica*, 2013, *81* (6), 2127–2204.

_ **and Elhanan Helpman**, "Global Sourcing," *Journal of Political Economy*, 2004, *112* (3), 552–580.

_ **and** _ , "Contractual Frictions and Global Sourcing," in Elhanan Helpman, Dalia Marin, and Thierry Verdier, eds., *The Organization of Firms in a Global Economy*, Cambridge, MA: Harvard University Press, 2008.

Arendt, Wolfgang, Charles JK Batty, Matthias Hieber, and Frank Neubrander, *Vector-valued Laplace transforms and Cauchy problems*, Vol. 96, Springer, 2011.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare, "The Elusive Pro-Competitive Effects of Trade," 2012. https://dl.dropboxusercontent.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.

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez, "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 2011, *49* (1), 3–71.

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, "The Gains from Merger or Collusion in Product-Differentiated Industries," *Journal of Industrial Economics*, 1985, *33* (4), 427–444.

— **and** — , "Estimating the Residual Demand Curve Facing a Single Firm," *International Journal of Industrial Organization*, 1988, *6* (3), 283–300.

Barnett, William A. and Andrew B. Jonas, "The Müntz-Szatz demand system: An application of a globally well behaved series expansion," *Economics Letters*, 1983, *11* (4), 337–342.

— **and Yul W. Lee**, "The global properties of the minflex Laurent, generalized Leontief, and translog flexible functional forms," *Econometrica*, 1985, pp. 1421–1437.

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

Behrens, Kristian and Yasusada Murata, "General Equilibrium Models of Monopolistic Competition: A New Approach," *Journal of Economic Theory*, 2007, *136* (1), 776–787.

— **and** — , "Trade, Competition, and Efficiency," *Journal of International Economics*, 2012, *87* (1), 1–17.

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

Bertoletti, Paolo and Federico Etro, "Monopolistic Competition when Income Matters," 2013. http://economia.unipv.it/pagp/pagine_personali/pberto/papers/MCIM.pdf.

— **and** — , "Pricing to Market in the Krugman Model," 2014. ftp://ftp.repec.org/opt/ReDIF/RePEc/pav/demwpp/DEMWP0061.pdf.

— **and Paolo Epifani**, "Monopolisitic Competition: CES Redux?," *Journal of International Economics*, 2014, *93* (2), 227–238.

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.

Bresnahan, Timothy F., "Empirical Studies of Industries with Market Power," in Richard Schmalensee and Robert Willig, eds., *Handbook of Industrial Organization*, Vol. 2, North-Holland B.V.: Amsterdam, Holland, 1989, pp. 1011–1057.

Brockett, Patrick L. and Linda L. Golden, "A class of utility functions containing all the common utility functions," *Management Science*, 1987, *33* (8), 955–964.

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.

Burda, Martin, Hatthew Harding, and Jerry Hausman, "A Bayesian Mixed Logit-Probit Model for Multinomial Choice," *Journal of Econometrics*, 2008, *147* (2), 232–246.

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.

Caplin, Andrew and Barry Nalebuff, "Aggregation and Imperfect Competition: On the Existence of Equilibrium," *Econometrica*, 1991, *59* (1), 25–59.

_ **and** _ , "Aggregation and Social Choice: A Mean Voter Theorem," *Econometrica*, 1991, *59* (1), 1–23.

Cassola, Nuno, Ali Hortaçsu, and Jakub Kastl, "The 2007 Subprime Market Crisis Through the Lens of European Central Bank Auctions for Short-Term Funds," *Econometrica*, 2013, *81* (4), 1309–1345.

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.

Cournot, Antoine A., *Recherches sur les Principes Mathematiques de la Theorie des Richess*, Paris, 1838.

Dhingra, Swati and John Morrow, "Monopolistic Competition and Optimum Product Diversity Under Firm Heterogeneity," 2014. http://www.sdhingra.com/selection3rdGain.pdf.

Diamond, Peter and Emmanuel Saez, "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 2011, *25* (4), 165–190.

Diewert, W. Erwin, "An application of the Shephard duality theorem: a generalized Leontief production function," *The Journal of Political Economy*, 1971, pp. 481–507.

_ , "Functional forms for profit and transformation functions," *Journal of Economic Theory*, 1973, *6* (3), 284–316.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan, "Consumer Price Search and Platform Design in Internet Commerce," 2014. http://web.stanford.edu/∼leinav/Search.pdf.

Dixit, Avinash K. and Joseph E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 1977, *67* (3), 297–308.

Dupuit, Arsène Jules Étienne Juvénal, *De la Mesure de L'utilité des Travaux Publics*, Paris, 1844.

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

_ **and** _ , "Selection in Insurance Markets: Theory and Empirics in Pictures," *Journal of Economic Perspectives*, 2011, *25* (1), 115–138.

_ **, Theresa Kuchler, Jonathan Levin, and Neel Sundaresan**, "Assessing Sale Strategies in Online Markets Using Matched Listings," *American Economic Journal: Microeconomics*, Forthcoming.

Fabinger, Michal and E. Glen Weyl, "Pass-Through and Demand Forms," 2012. http://ssrn.com/abstract=2194855.

_ **and** _ , "Price Discrimination is Typically Welfare-Enhancing," 2014. This paper is under preparation. Contact Glen Weyl at glenweyl@microsoft.com for notes.

Fabra, Natalia and Mar Reguant, "Pass-Through of Emissions Costs in Electricity Markets," *American Economic Review*, 2014, *104* (9), 2872–2899.

Farrell, Joseph and Carl Shapiro, "Horizontal Mergers: An Equilibrium Analysis," *American Economic Review*, 1990, *80* (1), 107–126.

Feller, Willliam, *An introduction to probability theory and its applications*, Vol. 2, John Wiley & Sons, 2008.

Gabaix, Xavier, David Laibson, Deyuan Li, Hongyi Li, Sidney Resnick, and Caspar G. de Vries, "The Impact of Competition on Prices with Numerous Firms," 2013. http://pages.stern.nyu.edu/∼xgabaix/papers/CompetitionEVT.pdf.

Garnier, Germaine, *Abrégé Élémentaire des Principes de L'Économie Politique*, Paris: H. Agasse, 1796.

Gaudin, Germain and Alexander White, "On the Antitrust Economics of the Electronic Book Industry," 2014. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2352495.

_ **and** _ , "Unit vs. Ad Valorem Taxes under Revenue Maximization," 2014. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2440708.

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.

Grossman, Sanford J. and Oliver D. Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 1986, *9* (4), 691–719.

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

Haile, Philip A. and Elie Tamer, "Inference with an Incomplete Model of English Auctions," *Inference with an Incomplete Model of English Auctions*, 2003, *111* (1), 1–51.

Hajargasht, Gholamreza and William E. Griffiths, "Pareto-lognormal Distributions: Inequality, Poverty, and Estimation from Grouped Income Data," *Economic Modeling*, 2013, *33*, 593–604.

Hart, Oliver and John Moore, "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 1990, *98* (6), 1119–1158.

Heckman, James J., Richard Robb, and James R. Walker, "Testing the mixture of exponentials hypothesis and estimating the mixing distribution by the method of moments," *Journal of the American Statistical Association*, 1990, *85* (410), 582–589.

Helpman, Elhanan and Oleg Itskhoki, "Labour Market Rigidities, Trade and Unemployment," *Review of Economic Studies*, 2010, *77* (3), 1100–1137.

_ **,** _ **, and Stephen Redding**, "Inequality and Unemployment in a Global Economy," *Econometrica*, 2010, *78* (4), 1239–1283.

_ **,** _ **, Marc-Andreas Muendler, and Stephen Redding**, "Trade and Inequality: From Theory to Evidence," 2014. http://www.princeton.edu/∼itskhoki/papers/TradeInequalityEvidence.pdf.

Holmström, Bengt and John Roberts, "The Boundaries of the Firm Revisited," *Journal of Economic Perspectives*, 1998, *12* (4), 73–94.

Houghton, John C., "Birth of a Parent: The Wakeby Distribution for Modeling Flood Flows," *Water Resources Research*, 1978, *14* (6), 1105–1109.

Hummels, David and Alexandre Skiba, "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture," *Journal of Political Economy*, 2004, *112* (6), 1384–1402.

Irarrazabal, Alfonso, Andreas Moxnes, and Luca David Opromolla, "The Tip of the Iceberg: A Quantitative Framework for Estimating Trade Costs," 2014. http://www.nber.org/papers/w19236.

Jr., Robert B. Ekelund and Robert F. Hébert, *Secret Origins of Moder Microeconomics*, Chicago: University of Chicago Press, 1999.

Kichko, Sergey, Sergey Kokovin, and Evgeny Zhelobodko, "Trade Patterns and Export Pricing Under Non-CES Preferences," 2014. http://goo.gl/s2v91J.

Kremer, Michael and Christopher M. Snyder, "Preventives Versus Treatments," 2014. http://scholar.harvard.edu/files/kremer/files/vac66_2013.01.11.pdf.

Krugman, Paul, "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 1980, *70* (5), 950–959.

Kubler, Felix and Karl Schmedders, "Competitive Equilibria in Semi-Algebraic Economies," *Journal of Economic Theory*, 2010, *145* (1), 301–330.

_ , Philipp Renner, and Karl Schmedders, "Computing All Solution to Polynomial Equatoins in Economics," in Karl Schmedders and Kenneth L. Judd, eds., *Handbook of Computational Economics*, Vol. 3, Amsterdam, Holland: Elsevier, 2014.

Lau, Lawrence J., "Functional forms in econometric model building," *Handbook of econometrics*, 1986, *3*, 1515–1566.

Lockwood, Benjamin B., Charles G. Nathanson, and E. Glen Weyl, "Taxation and the Allocation of Talent," 2014. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1324424.

MacDonald, James M. and Michael E. Ollinger, "Scale Economies and Consolidation in Hog Slaughter," *American Journal of Agricultural Economics*, 2000, *82* (2), 334–346.

Mahoney, Neale and E. Glen Weyl, "Imperfect Competition in Selection Markets," 2014. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2372661.

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.

Mill, John Stuart, *Principles of Political Economy with some of their Applications to Social Philosophy*, London: Longmans, Green and Co., 1848.

Miller, Kenneth S and Stefan G Samko, "Completely monotonic functions," *Integral Transforms and Special Functions*, 2001, *12* (4), 389–402.

Mrázová, Monika and J. Peter Neary, "Not So Demanding: Preference Structure, Firm Behavior, and Welfare," 2014. http://goo.gl/8CaXPv.

_ **and** _ , "Selection Effects With Heterogeneous Firms," 2014.

_ **and** _ , "Together at Last: Trade Costs, Demand Structure, and Welfare," *American Economic Review Papers and Proceedings*, 2014, *104* (5), 298–303.

Myerson, Roger B., "Optimal Auction Design," *Mathematics of Operations Research*, 1981, *6* (1), 58–73.

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

Parenti, Mathieu, Philip Ushchev, and Jacques-François Thisse, "Toward a Theory of Monopolistic Competition," 2014. http://goo.gl/gHVQLH.

Piketty, Thomas, *Capital in the Twenty-First Century*, Cambridge, MA: Harvard University Press, 2014.

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," *Journal of Economic Theory*, 2014, *152*, 266–290. 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*, 469–486.

Reed, William J and Murray Jorgensen, "The Double Pareto-Lognormal Distribution – A New Parametric Model for Size Distributions," *Communicatoins in Statistics – Theory and Methods*, 2004, *33* (8), 1733–1753.

Saez, Emmanuel, "Striking it Richer: The Evolution of Top Incomes in the United States," 2013. http://eml.berkeley.edu/∼saez/saez-UStopincomes-2012.pdf.

Salinger, Michael A., "Vertical Mergers and Market Foreclosure," *Quarterly Journal of Economics*, 1988, *5* (1), 345–356.

Say, Jean-Baptiste, *Traité D'Économie Politique ou Simple Exposition de la Maniée dont se Forment, se Distribuent et se Consomment les Richesses*, Paris: Guillaumin., 1819.

_ , *Cours Complet D'Économie Politique Practique*, Paris: Rapilly, 1828.

Schilling, René L, Renming Song, and Zoran Vondracek, *Bernstein functions: theory and applications*, Vol. 37, Walter de Gruyter, 2012.

Singh, Nirvikr and Xavier Vives, "Price and Quantity Competition in Differentiated Duopoly," *RAND Journal of Economics*, 1984, *15* (4), 546–554.

Stole, Lars A. and Jeffrey Zwiebel, "Intra-firm Bargaining under Non-Binding Contracts," *Review of Economic Studies*, 1996, *63* (3), 375–410.

_ **and** _ , "Organizational Design and Technology Choice under Intrafirm Bargaining," *American Economic Review*, 1996, *86* (1), 195–222.

Toda, Alexis Akira, "The Double Power Law in Income Distribution: Explanations and Evidence," *Journal of Economic Behavior & Organization*, 2012, *84* (1), 364–381.

_ , "Double Pareto-Lognormal Distribution in Economic Data," 2014. http://goo.gl/bo08lo.

— **and Kieran Walsh**, "The Double Power Law in Consumption and Implications for Testing Euler Equations," 2014. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2319454.

Weyl, E. Glen and Jean Tirole, "Market Power Screens Willingness-to-Pay," *Quarterly Journal of Economics*, 2012, *127* (4), 1971–2003.

— **and Michal Fabinger**, "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition," *Journal of Political Economy*, 2013, *121* (3), 528–583.

Widder, David Vernon, *The Laplace Transform*, Dover, 2010.

Zemanian, Armen H, *Distribution theory and transform analysis: an introduction to generalized functions, with applications*, Courier Dover Publications, 1965.

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti, and Jacques-François Thisse, "Monopolisitic Competition: Beyond the Constant Elasticity of Substitution," *Econometrica*, 2012, *80* (6), 2765–2784.

Appendix

A Additional Applications

In this appendix we consider how our approach may be formally applied to three broad contexts not discussed in the text. For brevity we do not derive any concrete results about these contexts, but instead simply show how our approach allows tractability through the same logic as described in the text.

A.1 Selection markets

Consider the model of perfect competition in selection markets proposed by Einav et al. (2010) and Einav and Finkelstein (2011).[40] In that model the welfare-relevant aggregates they analyze are characterized by an inverse demand P and an average cost function AC from which marginal cost may be derived as above. Perfectly competitive equilibrium is defined as an intersection between P and AC while a necessary condition for the social optimality is $P = MC$. Therefore by Observation 1 if Laplace inverse demand and average cost are chosen to be tractable at any desired level, marginal cost and thus the social optimization problem will inherit the same tractability. This is likely why both papers chose the highly jointly-tractable linear forms for both demand and cost, despite the fact that the demand curve estimated by Einav et al. appears concave. Our results can obviously therefore be used to extend their analysis to more realistic demand and cost shapes while preserving the same tractability.

A.2 Sequential-action supply chains

Consider the model of imperfectly competitive supply chains where each stage of production strategically anticipates the reactions of the subsequent stage proposed by Salinger (1988). There are m stages of production interacting via linear pricing. Producers at each stage act simultaneously and the stages act in sequence. We solve by backwards induction.

Producers at stage m take an input from producers at stage $m-1$ and sell it to final consumers, facing inverse demand P_m. The n_m firms at stage m are symmetric Cournot competitors with average cost AC_m. The linear price clearing the market between stage $m-1$ and m is $\hat{P}_{m-1}$. Using the standard first-order condition for Cournot competition and

[40] The analysis may be easily extended to any of the imperfectly competitive models discussed above (Mahoney and Weyl, 2014), but the method is sufficiently analogous that we ignore it here.

dropping arguments, the first-order equilibrium conditions are

$$P_m + \frac{1}{n_m} P'_m q = \hat{P}_{m-1} + AC_m + \frac{1}{n_m} AC'_m q \iff$$

$$\hat{P}_{m-1} = P_m + \frac{1}{n_m} P'_m q - AC_m - \frac{1}{n_m} AC'_m q.$$

Thus the effective inverse demand facing the firms at stage $m-1$ is

$$P_{m-1} \equiv P_m + \frac{1}{n_m} P'_m q - AC_m - \frac{1}{n_m} AC'_m q,$$

as all output produced at stage $m-1$ is used as an input at stage m. Effectively the inverse demand at stage $m-1$ is the (competition-adjusted) marginal profit (competition-adjusted marginal revenue less marginal cost) at stage m.

This analysis may be back-propagated up the supply chain to obtain a first-order condition at the first stage determining the quantity in the industry. However, at each stage one higher derivative of P_m, at least and also of some of the cost curves, enters the first-order conditions. Thus the implicit equation for the first-order conditions characterizing the supply chain is usually quite elaborate and is both difficult to analyze in general and highly intractable, even computationally, for many functional forms.

However we now derive a simple explicit transformation of the Laplace inverse demand and average cost characterizing the supply chain and discuss how this can be used to overcome these difficulties. Note that

$$P_m + \frac{1}{n_m} P'_m q = \left(1 - \frac{1}{n_m}\right) P_m + \frac{1}{n_m} MR_m,$$

where $MR_m = P_m + P'_m q$. Let p_m be the Laplace inverse demand. From Section 2 we have that the Laplace marginal revenue is $(1-t)p_m$ and thus that the ILT of $\left(1 - \frac{1}{n_m}\right) P_m + \frac{1}{n_m} MR_m$ is just $\left(1 - \frac{t}{n_m}\right) p_m$. By the same logic, if we denote the Laplace average cost by ac_m the ILT of $AC_m + \frac{1}{n_m} AC'_m q$ is $\left(1 - \frac{t}{n_m}\right) ac_m$.

Iterating this process, one obtains that the Laplace first-order condition at the initial stage, which we denote f_1, is

$$p_m \prod_{i=1}^{m} \left(1 - \frac{t}{n_i}\right) - \sum_{i=1}^{m} \left[ac_i \prod_{j=1}^{i} \left(1 - \frac{t}{n_j}\right) \right].$$

This obviously differs only in its (trivially computed) coefficients and not in its support from the ac_i's and p_m that make it up. Thus if all ac_i's and p_m are chosen to have the same tractable support (with the desired number of evenly-spaced mass points to achieve desired tractability) then the full will be equally tractable. Beyond this, even if p_m and the ac_i's are specified in an arbitrary manner, the resulting Laplace first-order condition can be trivially computed from the ILTs of each of these inputs and then either solved directly by applying the Laplace transform or approximated using a small number of evenly-spaced mass points for tractability. In either case, the complexity of computing and representing the system is significantly reduced. This is another example of how our approach directly applies well beyond the simple monopoly model we focused on in Section 2.

A.3 Symmetrically differentiated Nash-in-Price competition

FORTHCOMING IN NEXT DRAFT.

B Definitions, Theorems, and Pass-Through Properties Demand Functions

In the following we discuss properties of single-product utility functions on some finite interval $[0, \bar{q}]$ that may be written in the form

$$U(q) = \int_{-\infty}^{0} u(t)\, q^{-t} dt, \tag{20}$$

where the integral is defined by (21). Note that this integral may be interpreted as Laplace transform in terms of the variable $s \equiv \log q$. As long as we wish to consider utility bounded below, there is virtually no loss of generality associated with the utility specification (20). (In situations when utility unbounded below is desired, e.g. for constant demand elasticity smaller than one, we can instead use the bilateral specification (22)). Similarly, for economic purposes there is no loss of generality associated with $\bar{q}$ being finite, since its value may be chosen arbitrarily. Note that even though we refer to $U(q)$ as a utility function, the same mathematical theorems would apply even if we considered, say, cost functions instead.

Technical clarification (Integral definition). Here we define the integral (20) to be the Riemann-Stieltjes integral

$$U(q) = \int_{-\infty}^{0} q^{-t} du_I(t) \tag{21}$$

for some function $u_I(t)$, not necessarily nonnegative, such that the integral converges. If this function is differentiable, its derivative $u_I'(t)$ is the $u(t)$ that appears on the right-hand side of (20). If $u_I(t)$ is only piecewise differentiable, then $u(t)$ is not an ordinary function, but involves Dirac delta functions (i.e. point masses) at the points of discontinuity of $u_I(t)$.

Note that in certain parts of the paper we need a more general definition of (20) than (21). In those cases we use the Schwartz-Sobolev distribution theory instead of the Riemann-Stieltjes integral theory.

Proposition (Uniqueness). *For each function $U(q)$ that may be represented in the form (20) in the sense (21), there exists just one normalized*[41] *function $u_I(t)$ such that (21) holds.*

Proof. This follows from Theorem I.6.3 of of Widder (2010). □

Proposition (Polynomial functions). *Any polynomial utility function may be written in the form (20).*

Proof. If we choose $u_I(t)$ appearing in (21) to be piecewise constant with a finite number N of points of discontinuity $\{t_j, j = 1, 2, ..., N\}$, the integral becomes

$$U(q) = \sum_{j=1}^{N} a_j q^{-t_j},$$

where a_j is the (signed) magnitude of the discontinuity at point t_j, i.e. the magnitude of the mass that $u(t)$ has at point t_j. If we choose t_j to be nonpositive integers, $U(q)$ will be a polynomial of q. By appropriate choices of N and a_j, any polynomial of q may be expressed in this way. □

Proposition (Arbitrarily precise approximations). *An arbitrary utility function $\tilde{U}(q)$ continuous on an interval interval $[0, \bar{q}]$ may be approximated with an arbitrary precision by utility functions of the form (20), in the sense of uniform convergence on $[0, \bar{q}]$.*

Technical clarification (Uniform convergence). In other words, for any continuous $\tilde{U}(q)$ there exists a sequence $\{U_j(q), j \in \mathbb{N}\}$ of functions of the form (20) such that for any $\epsilon > 0$, all elements of the sequence after some position n_ϵ satisfy $\sup_{q\in[0,\bar{q}]} |\tilde{U}(q) - U_j(q)| < \epsilon$.

[41] Normalization here means that $u_I(0+) = 0$ and $u_I(t) = (u_I(t+) + u_I(t-))/2$. See Section I.6 of Widder (2010).

Proof. Given that polynomials are included in the specification (20), the theorem follows from the Weierstrass approximation theorem, which states that polynomials are dense in the space of continuous functions on a compact interval. For a constructive proof of the theorem due to Bernstein, see e.g. Section VII.2 of Feller (2008). □

Proposition (Analyticity). *All functions of the form (20) are analytic.*

Proof. This follows from Theorem I.5a of Widder (2010). □

Comment (Conditions for the existence of inverse Laplace transform). The Laplace representation (20) of a given utility function $U(q)$ exists under various conditions. Theorem 18b in Section VII.18 of Widder (2010) states general necessary and sufficient conditions on $U(q)$ for the existence of $u_I(t)$ such that (21) is satisfied; *almost all* utility functions we may encounter in economic applications *do satisfy these conditions.*[42] Sections VII.12-17 of Widder (2010) provides conditions that guarantee that $u_I(t)$ exists and has certain properties, such as being of bounded variation, nondecreasing, or belonging to the functional space L^p. Additional conditions may be found in Chapter 2 of the book by Arendt et al. (2011), which contains recent developments in the theory.

The following proposition goes beyond the theory of the Riemann-Stieltjes integral and instead discusses Laplace transform of generalized functions based on the Schwartz-Sobolev distribution theory. Zemanian (1965) provides the definition of the space of distributions $\mathcal{D}'$ used below. We denote by $\mathbb{C}_{\bar{s}}^{-}$ the half complex plane $\{s|\text{Re } s < \bar{s}\}$, where $\bar{s}$ is a real number smaller than $\log \bar{q}$, i.e. than the logarithm the upper bound $\bar{q}$ of the domain of $U(q)$ considered in (20).

Proposition (Uniqueness, analyticity, asymptotics - generalized functions). *A function $U(q)$ such that $U_{[s]}(s) \equiv U(e^s)$ in the domain $\mathbb{C}_{\bar{s}}^{-}$ is analytic (i.e. holomorphic) and bounded*[43] *by a polynomial of s may be expressed in the form (20) with u representing a distribution (i.e. an element of $\mathcal{D}'$). This distribution is unique. Conversely, for any Laplace-transformable distribution u, the integral (20) viewed as a function of $s \equiv \log q$ in the domain $\mathbb{C}_{\bar{s}}^{-}$ is analytic and bounded by a polynomial of s.*

[42] To handle rare useful cases that do not satisfy these conditions or their counterpart related to bilateral Laplace transform, we use Laplace transform based on Schwartz-Sobolev distribution theory instead of the Riemann-Stieltjes integral theory.

[43] Bounded in the sense of its absolute value being no greater than the absolute value of a polynomial of s.

Proof. The three sentences of the proposition are implied by the following statements in Zemanian (1965): (1) Theorem 8.4-1 and Corollary 8.4-1a, (2) Theorem 8.3-1a, (3) Theorem 8.3-2 and the text following Corollary 8.4-1a. □

B.1 Bilateral Laplace transform

The discussion above generalizes to single-product utility functions that may be written in the form

$$U(q) = \int_{-\infty}^{\infty} u(t)\, q^{-t} dt. \tag{22}$$

This specification allows for, e.g., constant demand elasticity smaller than 1. (We plan to add more details on the bilateral Laplace transform in a future draft.)

B.2 Complete monotonicity of the demand specification

Many demand curves have economic properties that may seem unexpected, but which are easily understood in terms of Laplace transform. To develop the related theory, we start with a standard definition of completely monotone functions and then discuss relations between complete monotonicity, the form of Laplace inverse demand, and economic consequences for the pass-through rate, as well as for supply chain models.[44]

Definition (Completely monotone function). *A function $f(x)$ is completely monotone iff for all $n \in \mathbb{N}$ its nth derivative exists and satisfies*

$$(-1)^n f^{(n)}(x) \geq 0.$$

It turns out that many commonly used demand functions are such that the consumer surplus is completely monotone as a function of negative log quantity. For this reason, we make the following definition.

Definition (Complete monotonicity of the demand specification).[45] *We say that the demand function (or utility function) satisfies the complete monotonicity criterion iff*

[44]Brockett and Golden (1987) also discuss relations between complete monotonicity and a type of Laplace transform. The Laplace transform used there is in terms of quantity q, whereas in our discussion, it is in terms of the logarithm of quantity. These two transforms are distinct and should not be confused.

[45]In principle, it is possible to empirically test whether an empirical demand curve satisfies the complete monotonicity criterion. The relevant empirical test has been developed by Heckman et al. (1990). It would just have to be translated from the duration analysis context to our demand theory context.

consumer surplus is a completely monotone function of $-s$, i.e. for all $n \in \mathbb{N}$,

$$CS^{(n)}_{[s]}(s) \geq 0,$$

or equivalently

$$U^{(n)}_{[s]}(s) - U^{(n+1)}_{[s]}(s) \geq 0.$$

Strict complete monotonicity criterion then refers to these inequalities being strict.

Proof of Definition Equivalence. With the marginal utility of the outside good normalized to one and $U(0)$ is set to zero, we have $CS(q) = -qP(q) + \int_0^q P(q_1)\,dq_1 = -qU'(q) + \int_0^q U'(q_1)\,dq_1 = U(q) - qU'(q)$. This translates into $CS_{[s]}(s) = U_{[s]}(s) - U'_{[s]}(s)$. The equivalence for any $n \in \mathbb{N}$ then follows by differentiation.

Proposition. *A (single-product) utility function is bounded below and satisfies the complete monotonicity criterion iff the Laplace consumer surplus $cs(t)$ is nonnegative and supported on $(-\infty, 0)$, i.e. $CS(q) = \int_{-\infty}^0 cs(t)\,q^{-t}dt$ for some $cs(t) \geq 0$.*[46]

Proof. This follows from Bernstein's theorem on completely monotone functions, formulated e.g. as Theorem IV.12a of Widder (2010) or Theorem 1.4 of Schilling et al. (2012). □

Proposition. *The complete monotonicity criterion for demand functions implies pass-through decreasing with quantity in the case of constant-marginal-cost monopoly. The only exception is Bulow-Pfleiderer demand, for which pass-through is constant.*

Proof. Constant marginal cost monopoly pass-through may be expressed as

$$\rho = \frac{CS'_{[s]}(s)}{CS''_{[s]}(s)}.$$

For a completely monotone problem, Laplace consumer surplus $cs(t)$ is nonnegative. For this reason, the inverse of ρ may be expressed as a weighted average of t with nonnegative

[46]Clarification to be added.

weight

$$w(t,s) \equiv \frac{t\,cs(t)\,e^{-st}}{\int_{-\infty}^{0} t\,cs(t)\,e^{-st}dt}$$

as follows

$$\frac{1}{\rho} = \frac{CS''_{[s]}(s)}{CS'_{[s]}(s)} = -\frac{\int_{-\infty}^{0} t^2 cs(t)\,e^{-st}dt}{\int_{-\infty}^{0} t\ cs(t)\,e^{-st}dt} = -\int_{-\infty}^{0} t\,w(t,s)\,dt.$$

In response to an increase in s, the weight gets shifted towards more negative t,[47] and $1/\rho$ decreases. We conclude that ρ is decreasing in q. Only if $t\,cs(t)$ is supported at one point will there be no shift in weight and ρ remains constant. That case corresponds to Bulow-Pfleiderer demand. □

Proposition. *The following demand functions satisfy the complete monotonicity criterion: Pareto/constant elasticity ($\epsilon > 1$), Bulow-Pfleiderer ($\epsilon > 1$), logistic distribution, log-logistic distribution ($\gamma > 1$), Gumbel distribution ($\alpha > 1$), Weibull distribution ($\alpha > 1$), Fréchet distribution ($\alpha > 1$), gamma distribution ($\alpha > 1$), Laplace distribution*[48]*, Singh-Maddala distribution ($a > 1$), Tukey lambda distribution ($\lambda < 1$), Wakeby distribution ($\beta > 1$), generalized Pareto distribution ($\gamma < 1$), Cauchy distribution.*

Proof. The complete monotonicity properties follow by straightforwardly recognizing that in these cases $t\,p(t)$ is nonnegative and supported on $(-\infty, 1)$, with the corresponding Laplace inverse demand functions $p(t)$ listed in Appendix D.2.[49] Note that for most of the inverse demand functions listed in the proposition, it is also possible to prove complete monotonicity using Theorems 1–6 of Miller and Samko (2001). □

Corollary. The demand functions listed in the previous proposition lead to constant-marginal-cost pass-through decreasing in quantity, with the exception of Pareto/constant elasticity and Bulow-Pfleiderer, which lead to constant pass-through.

Proposition. *The following demand functions do* ***not*** *satisfy the complete monotonicity criterion: normal distribution, lognormal distribution, constant superelasticity (Klenow and Willis), Almost Ideal Demand System (either with finite or infinite surplus), log-logistic distribution ($\gamma < 1$), Fréchet distribution ($\alpha < 1$), Weibull distribution ($\alpha < 1$), Gumbel*

[47] In the same mathematical sense as in the definition of first order stochastic dominance.

[48] Each half of the distribution separately, or the full distribution smoothed by *arccosh* to ensure the existence of the derivatives.

[49] The text of Appendix D.2 is to be improved.

distribution ($\alpha < 1$), Pareto/constant elasticity ($\varepsilon > 1$), gamma distribution ($\alpha < 1$), Singh-Maddala distribution ($a < 1$), Tukey lambda distribution ($\lambda > 1$), Wakeby distribution ($\beta < 1$), generalized Pareto distribution ($\gamma > 1$).

Note (normal distribution of values). We found that the normal distribution of consumer values has properties very close to those satisfying the complete monotonicity criterion: constant-marginal-cost pass-through is increasing in price (as we show below), and low-order derivatives of $CS(s)$ with respect to $-s$ are positive. We concluded that the complete monotonicity criterion is not satisfied based on examining the sign on the tenth derivative of $CS(s)$. The absence of complete monotonicity is consistent with our expression to the corresponding Laplace inverse demand, which does not seem to satisfy $t\, cs(t) \geq 0$. In most economic applications, the difference from completely monotone problems is inconsequential because it manifests itself only in very high derivatives of $CS(s)$.

C Invariance under average-marginal transformations

In this appendix we derive the structure of classes of functions on $\mathbb{R}^+$ with m parameters that are invariant under average-marginal transformations. An m-dimensional functional form class is a subset of a space of functions that is homeomorphic to an m-dimensional manifold (moduli space), possibly with a boundary. The m shape parameters are the coordinates on this moduli space. For example, in the case of the functional form $\beta \exp(-(q-\alpha)^2)$, the moduli space would be a plane ($\mathbb{R}^2$) with coordinates α and β.

Definition. *We say that a functional form class $\mathcal{C}$ is* invariant under average-marginal transformations *if for any function $F(q) \in \mathcal{C}$, the class also contains any linear combination of $F(q)$ and $qF'(q)$. In other words, $F \in \mathcal{C} \Rightarrow \forall (a,b) \in \mathbb{R}^2 :\ aF + bqF' \in \mathcal{C}$.*

In economic terms, we interpret $F(q)$ as the average of the variable $qF(q)$, such as revenue or cost, and $F(q) + qF'(q)$ as its marginal counterpart.

Theorem. *Any real finite-dimensional functional form class invariant under average-marginal transformations is the set of linear combinations of*

$$(\log q)^{a_{jk}}\, q^{-t_j}, \quad a_{jk} = 0, 1, ..., n_j, \quad j = 1, 2, ..., N_1,$$

$$(\log q)^{b_k} \cos\left(\tilde{t}_j \log q\right) q^{-\hat{t}_j}, \quad b_{jk} = 0, 1, ..., n_j, \quad j = 1, 2, ..., N_2,$$

$$(\log q)^{c_k} \sin\left(\tilde{t}_j \log q\right) q^{-\hat{t}_j}, \quad c_{jk} = 0, 1, ..., n_j, \quad j = 1, 2, ..., N_2,$$

where $\{t_j, j = 1, 2, ..., N_1\}$, $\{\tilde{t}_j, j = 1, 2, ..., N_2\}$, *and* $\{\hat{t}_j, j = 1, 2, ..., N_2\}$ *are sets of real numbers. If we exclude functions oscillating as* $q \to 0_+$, *only the functions in the first row are allowed. In that case the most general form is the set of linear combinations of*

$$q^{-t_j}, \quad q^{-t_j} \log q, \quad q^{-t_j} (\log q)^2, \quad ... \quad , q^{-t_j} (\log q)^{n_j}, \quad j = 1, 2, ..., N_1.$$

Proof. For convenience we express the (infinitely differentiable) functions $F(q)$ on $\mathbb{R}^+$ in terms of functions $G(s)$ defined on $\mathbb{R}$, with the identification $s \equiv \log q$, $F(q) \equiv G(\log q)$. Consider a function $F(q) \in \mathcal{C}$ and its counterpart $G(s)$. In terms of G, the average-marginal form invariance requires that the counterpart of $aG + bG'$ belong to the class $\mathcal{C}$, if the of counterpart of G does so. For technical reasons, we will work with $G(s)$ multiplied by the characteristic function $1_S(s)$ of an arbitrarily chosen non-empty interval $S \equiv (s_1, s_2)$, i.e. with $G_S(s) \equiv G(s) 1_S(s)$. We denote by $\hat{G}_S(\omega)$ the Fourier transform of $G_S(s)$, which in turn may be expressed as the inverse Fourier transform $G_S(s) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \hat{G}_S(\omega) e^{-i\omega s} d\omega$.[50]

By iterating the defining property of average-marginal invariance, we know that the class $\mathcal{C}$ contains also counterparts of the derivatives $G^{(n)}(s)$. We will consider the first m of them, in addition to $G(s)$. For $n = 1, 2, ..., m$, we denote by $G_S^{(n)}(s)$ the truncation of $G^{(n)}(s)$ to the interval S, i.e. $G_S^{(n)}(s) \equiv G^{(n)}(s) 1_{s \in S}$. Inside the interval S,

$$G_S^{(n)}(s) = \int_{-\infty}^{\infty} (-i\omega)^n \hat{G}_S(\omega) e^{-i\omega s} d\omega, \quad \text{for } s \in S, \ n \in \{0, 1, 2, ..., m\}. \tag{23}$$

The $m+1$ functions $G_S(s), G_S^{(1)}(s), G_S^{(2)}(s), ..., G_S^{(m)}(s)$ span a vector space with dimensionality $m+1$ or less. Dimensionality equal to $m+1$ would contradict the assumption of having an m-dimensional functional form class, which implies that the set of functions $G_S(s), G_S^{(1)}(s), G_S^{(2)}(s), ..., G_S^{(m)}(s)$ must be linearly dependent on the interval S. As a result, there must exist a polynomial $T_0(.)$ (with real coefficients), such that

$$\int_{-\infty}^{\infty} T_0(-i\omega) \hat{G}_S(\omega) e^{-i\omega s} d\omega \tag{24}$$

is zero for any $s \in S$. This expression vanishes not only for $s \in S \equiv (s_1, s_2)$, but also for $s \in (-\infty, s_1)$ and $s \in (s_2, \infty)$. This is because the right-hand-side of (23) when extended to arbitrary $s \in \mathbb{R}$ represents the nth derivative of $G_S(s)$ in the sense of the Schwartz-Sobolev distribution theory, and given that $G_S(s)$ vanishes for $s \in (-\infty, s_1)$ and $s \in (s_2, \infty)$, so must

[50]The Fourier transform used in the proof is equivalent to the Laplace transform with imaginary s. Both transform may be thought of as parts of the holomorphic Fourier-Laplace transform.

its nth derivative. Given that the expression (24) is a generalized function[51] of s that gives zero when integrated against any test function[52] supported on $(-\infty, s_1-\epsilon] \cup [s_1+\epsilon, s_2-\epsilon] \cup [s_2+\epsilon, \infty)$ for any $\epsilon > 0$, we may write it as a linear combination of Dirac delta functions and a finite number of their derivatives located at s_1 and s_2. By computing its Fourier transform we find that $T_0(-i\omega)\hat{G}_S(\omega)$ must be of the form

$$T_1(\omega)\, e^{is_1\omega} + T_2(\omega)\, e^{is_2\omega}$$

with some polynomials $T_1(\omega)$ and $T_2(\omega)$. Consequently, $\hat{G}_S(\omega)$ may be written as

$$\hat{G}_S(\omega) = \frac{T_1(\omega)}{T_0(-i\omega)} e^{is_1\omega} + \frac{T_2(\omega)}{T_0(-i\omega)} e^{is_2\omega}.$$

The polynomial $T_0(-i\omega)$ may have a common factor with $T_1(\omega)$ or $T_2(\omega)$ or both. If we cancel these common factors, we may rewrite the expression as

$$\hat{G}_S(\omega) = \frac{T_3(\omega)}{T_5(\omega)} e^{is_1\omega} + \frac{T_4(\omega)}{T_6(\omega)} e^{is_2\omega} \tag{25}$$

for some polynomials T_3, T_4, T_5, and T_5, such that T_3 has no common divisors with T_5 and similarly for T_4 with T_6. Let us compute the inverse Fourier transform of the last expression for $\hat{G}_S(\omega)$ using the residue theorem. To perform the integration, we consider each of the two terms in (25) separately and specialize to $s \in S$. We close the integration contour by semicircles at infinity of the complex plane, correctly chosen so that their contribution to the integral vanishes. The integral value is then equal to the sum of the pole (residue) contributions, which give exponentials of s multiplied by polynomials of s. We see that for $s \in S$,

$$G_S(s) = \sum_{j=1}^{N} D_j(s)\, e^{-ist_j},$$

for some integer N, complex numbers t_j and polynomials $D_j(s)$. Since the interval S was chosen arbitrarily, not just $G_S(s)$, but also $G(s)$ itself must take this form. In the last expression the constants may be complex. Without loss of generality, we can assume that the first N_1 numbers t_j are real and the remaining ones have an imaginary part. By combining individual terms into real contributions so that $G(s)$ is real, we get

$$G(s) = \sum_{j=1}^{N_1} A_j(s)\, e^{-st_j} + \sum_{j=1}^{N_2} \left(B_j(s)\cos\tilde{t}_j s + C_j(s)\sin\tilde{t}_j s\right) e^{-\hat{t}_{R,j}s},$$

[51] By a generalized function we mean an element of the space $\mathcal{S}'(\mathbb{R})$ of distributions.

[52] A test function here refers to an element of the space $\mathcal{S}(\mathbb{R})$ of space of rapidly decreasing functions.

where $A_j(s)$, $B_j(s)$, and $C_j(s)$ are polynomials, and $N_1 + 2N_2 = N$. This form of $G(s)$ translates into the following form of $F(q)$:

$$F(q) = \sum_{j=1}^{N_1} A_j(\log q)\, q^{-t_j} + \sum_{j=1}^{N_2} \left(B_j(\log q)\cos\left(\tilde{t}_j \log q\right) + C_j(\log q)\sin\left(\tilde{t}_j \log q\right)\right) q^{-\hat{t}_j}. \quad (26)$$

If we wish to exclude the possibility of oscillations, e.g. in economic applications where we allow the functional form to be valid arbitrarily close to $q = 0$, we can set the polynomials B_j and C_j to zero and consider only functions of the form

$$F(q) = \sum_{k=1}^{N_1} A_j(\log q)\, q^{-t_j}.$$

An example of functional forms of this kind is $aq^{-t} + bq^{-u} + cq^{-u}\log q + dq^{-u}(\log q)^2$. The reader can easily verify that this is a four-dimensional functional form class invariant under average-marginal transformations. In general, it is now straightforward to check that the result (26) implies the statement of the theorem. □

D Demand Forms

D.1 Curvature properties

Note: This appendix was written before we developed the theory related to the complete monotonicity criterion discussed in Appendix B. For this reason a number of proofs described below is redundant and will be removed in a future draft.

Table 1 provides a taxonomy of the curvature properties of demand functions generated by common statistical distributions and the single-product version of the Almost Ideal Demand System. Following Caplin and Nalebuff (1991b,a), we define the the curvature of demand as

$$\kappa(p) \equiv \frac{Q''(p)Q(p)}{[Q'(p)]^2}.$$

Cournot (1838) showed that the pass-through rate of a constant marginal cost monopolist is

$$\frac{1}{2-\kappa}$$

and thus that a) that the comparison of κ to unity determines the comparison of pass-through to unity in this case and b) that if $\kappa'(p) > 0$ that pass-through rises with price (falls with quantity), and conversely if κ declines with price (rises with quantity). The comparison of κ to unity also determines whether a demand is log-convex and its sign whether demand is convex. The comparison of κ to 2 determines whether demand has declining marginal revenue, a condition also known as Myerson (1981)'s regularity condition.

For probability distribution F, the corresponding demand function $Q(p) = \sigma\left(1 - F\left(\frac{p-\mu}{m}\right)\right)$ where σ and m are stretch parameters (Weyl and Tirole, 2012) and μ is a position parameter. Note that in this case

$$\kappa(p) = -\frac{\frac{\sigma^2}{m^2} F''\left(\frac{p-\mu}{m}\right)\left(1 - F\left(\frac{p-\mu}{m}\right)\right)}{\frac{\sigma^2}{m^2}\left[F'\left(\frac{p-\mu}{m}\right)\right]^2} = -\frac{F''\left(\frac{p-\mu}{m}\right)\left(1 - F\left(\frac{p-\mu}{m}\right)\right)}{\left[F'\left(\frac{p-\mu}{m}\right)\right]^2}.$$

Note, thus, that neither global level nor slope properties of κ are affected by σ, m or μ. We can thus analyze the properties of relevant distributions independently of their values, as represented in the table and the following proposition.

The most prominent conclusion emerging from this taxonomy is that the vast majority of forms used in practice in computational, statistical models such as Berry et al. (1995) have monotonically increasing curvature and most have curvature below unity. This suggests two conclusions. The first, highlighted in the text, is that, to the extent we believe these forms are more realistic than tractable forms, they have properties systematically differing from the BP class and thus it is important to derive tractable forms capable of matching their central property of monotonically increasing in price/decreasing in quantity curvature.

A second possible conclusion is that, to the extent that in some cases these properties are *not* empirically relevant, such as in the data of Einav et al. (Forthcoming), standard forms rule out observed behavior and thus analysts may wish to consider more flexible forms along these dimensions, such as those we derive. For example, the Apt demand class we discussed in Subsection 4.1 allows flexibility over curvature and the slope of curvature, as well as the level and local elasticity of demand, unlike many of the common statistical forms. To the extent there are not strong theoretical reasons to believe in the restrictions imposed by standard statistically based forms (which, in many cases, there are) allowing such relaxation is important because in many contexts the properties of firm demand and equilibrium are inherited directly from the demand function, at least with constant marginal cost (Weyl and Fabinger, 2013; Gabaix et al., 2013; Quint, 2014). Which conclusion is most appropriate will obviously depend on the empirical context and the views of the analyst.

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.*

	$\kappa < 1$	$\kappa > 1$	Price-dependent	Parameter-dependent
$\kappa' < 0$			AIDS with $b < 0$	
$\kappa' > 0$	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 $\kappa < 2$		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

Proof. Characterization of the curvature level (comparisons of κ 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.[53] 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

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

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

[53] 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.

expression is the same as that of

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

as $x\to\infty$. Clearly this is greater than 2 for $0<\alpha<1$ so that for sufficiently large x, $\kappa>2$.

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

   $$\kappa(x)=\frac{1-\alpha}{\alpha x^{\alpha}}+1.$$

   Clearly for any $\alpha<1$ as $x\to 0$ this expression goes to infinity, so that for sufficiently small x, $\kappa>2$.

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:

   $$\kappa(x)=\frac{e^{x}(1-\alpha+x)\Gamma(\alpha,x)}{x^{\alpha}}. \tag{27}$$

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

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

   as $1-\alpha>0$ for $\alpha<1$. Thus clearly for small enough x, the Gamma distribution with shape $\alpha<1$ has $\kappa>2$.

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; we let ϕ denote the corresponding density. It is well-known that $\Phi''(x)=-x\phi(x)$. Thus

   $$\kappa(x)=\frac{x\left[1-\Phi(x)\right]}{\phi(x)}.$$

Taking the derivative and simplifying yields

$$\kappa'(x) = \frac{[1-\Phi(x)]\,(1+x^2) - x\phi(x)}{\phi(x)},$$

which clearly has the same sign as its numerator, as ϕ is a density and thus everywhere positive. But a classic strict lower bound for $\Phi(x)$ is $\frac{x}{1+x^2}\phi(x)$, implying $\kappa' > 0$.

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

$$\kappa'(x) = e^{-x} > 0.$$

Thus the logistic distribution has $\kappa' > 0$.

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

$$\kappa'(x) = e^{-x}.$$

Note that this is the same as for the logistic distribution; in fact κ for the Gumbel minimum distribution is 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

$$\kappa'(x) = e^{-x}\big(e^{2x}[e^{e^{-x}} - 1] - e^{e^{-x}}[e^x - 1]\big).$$

For $x < 0$ this is clearly positive as both terms are strictly positive: $1 > e^x$ and because $e^{-x} > 0$, $e^{e^{-x}} > 1$. For $x > 0$ we can rewrite κ' as

$$e^{e^{-x}}(e^x - 1) + e^{-x}\left(e^{e^{-x}} - 1\right),$$

which again is positive as $e^x > 1$ for $x > 0$ and $e^{e^{-x}} > 1$ by our argument above.

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 $x < 0$

$$\kappa'(x) = 2e^{-x} > 0.$$

So the Laplace distribution exhibits globally weakly increasing pass-through, strictly increasing for prices below the mode. The curvature for this distribution is $1 - 2e^{-x}$ as opposed to $1 - e^{-x}$ for Gumbel and Logistic. However these are very similar, again pointing out the similarities among curvature properties of 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

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

which can easily be shown graphically to be positive; we plan to show this analytically in the next draft. Thus this distribution, as well, has $\kappa' > 0$.

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

$$\kappa'(x) = (-x)^{\alpha-1}\alpha^2\left[1 - \alpha + e^{(-x)^{\alpha}}\left([1-\alpha]\left[(-x)^{\alpha} - 1\right] + [-x]^{2\alpha}\alpha\right)\right],$$

which has the same sign as

$$1 - \alpha + e^{(-x)^{\alpha}}\left([1-\alpha]\left[(-x)^{\alpha} - 1\right] + [-x]^{2\alpha}\alpha\right). \tag{28}$$

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

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

and its derivative is
$$\frac{e^{(-x)^{\alpha}}(-x)^{2\alpha}\alpha\left(1 + \alpha + [-x]^{\alpha}\alpha\right)}{x},$$

which is clearly strictly negative for $x < 0$. Thus expression (28) is strictly decreasing and approaches 0 as x approaches 0. It is therefore positive for all negative x, showing that again in this case $\kappa' > 0$.

7. Weibull distribution with shape $\alpha > 1$: As with the Fréchet distribution algebra from the earlier formula shows
$$\kappa'(x) = x^{\alpha-1}(\alpha - 1)\alpha^2,$$

which is clearly positive for $\alpha > 1$ as the range of this distribution is positive x. Thus the Weibull distribution with $\alpha > 1$ has $\kappa' > 0$.

8. Gamma distribution with shape $\alpha > 1$: Taking the derivative of Expression 27 yields:

$$\kappa'(x) = \frac{\alpha - 1 - x + \frac{e^x}{x^\alpha}\left(x^2 - 2x[\alpha-1] + [\alpha-1]\alpha\right)\Gamma(\alpha, x)}{x},$$

which has the same sign as

$$\alpha - 1 - x + \frac{e^x}{x^\alpha}\left(x^2 - 2x[\alpha-1] + [\alpha-1]\alpha\right)\Gamma(\alpha, x), \tag{29}$$

given that $x > 0$. 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 29 is strictly greater than

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

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

This establishes the second part of the proposition. Turning to our final two claims, algebra shows that the curvature for the Fréchet distribution is

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

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 expression goes to $-\infty$. So for sufficiently small $x > 0$, $\kappa(x) < 1$. On the other consider the second version of the expression. Its numerator is

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

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

$$-x^{-\alpha} \cdot -x^{\alpha}\left(1+\alpha\right)$$

as $x \to \infty$. Thus

$$\lim_{x\to\infty} \kappa(x) = \frac{1+\alpha}{\alpha} > 1$$

and thus for sufficiently large x and any $\alpha > 1$, this distribution has $\kappa > 1$.

Finally, consider our claim about AIDS. First note that for this demand function

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

as $b < 0$ and $p \le e^{-\frac{a}{b}} < e^{2-\frac{a}{b}}$. 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 $\kappa(p) > 1$. 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 $\kappa(p) < 1$.

$$\kappa'(p) = b^2 - \left(a - 2b + b\log[p]\right)^2,$$

which has the same sign as

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

Thus $\kappa' < 0$.

□

We now turn to two important distributions, which are typically used to model the income distribution, whose behavior is more complex and which, to our knowledge, have not been analyzed for their curvature properties. We focus only on the two that we believe to be most common (the first), best theoretically founded (both) and to provide the most accurate match to the income distribution (the second). Namely, we analyze the lognormal and double Pareto-lognormal (dPln) distributions, the latter of which was proposed by Reed (2003) and Reed and Jorgensen (2004). Other common, accurate models of income distributions which we have analyzed in less detail, appear to behave in a similar fashion.

We begin with the lognormal distribution, which is much more commonly used, and for which we have detailed, analytic results. However, while most of the arguments for the below proposition are proven analytically, some simple points are made by computational inspection.

Proposition 2. *For every value σ, there exist finite thresholds $\overline{y}(\sigma) > \underline{y}(\sigma)$ such that*

1. *If $y \geq \overline{y}(\sigma)$ then $\kappa' \leq 0$, and similarly with strict inequalities or if the directions of the inequalities both reverse.*

2. *If $y \geq \underline{y}(\sigma)$ then $\kappa \geq 1$, and similarly with strict inequalities or if the directions of the inequalities both reverse.*

Both $\overline{y}$ and $\underline{y}$ are strictly decreasing in σ.

Under the lognormal distribution, behavior depends critically on the amount of inequality or equivalently the standard deviation of the logarithm of the distribution: there is famously a one-to-one relationship between the Gini coefficient associated with a lognormal distribution and its logarithmic standard deviation. If inequality is not high, the behavior of curvature like a normal distribution occurs except at fairly high incomes levels; for a Gini of .34, for example, monotonicity of κ is preserved until the top 1% of the income distribution and

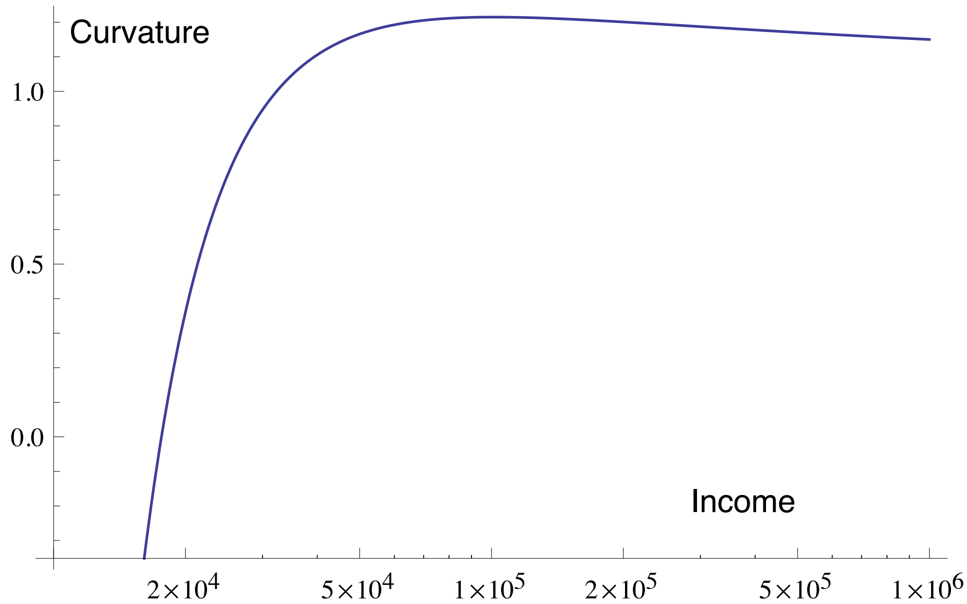

Figure 9: Curvature of a lognormal distribution calibrated to the US income distribution: parameters are $\mu = 10.5$ and $\sigma = .85$.

log-concavity outside of the top 30%. However, if inequality is sufficiently high, in particular if the Gini coefficient is above about .72, then the lognormal distribution has $\kappa > 2$ over some range and then κ converges back to 1 for very large incomes. This result is not discussed in the proposition, but can easily be seen by inspecting a graph of the expression for κ given in the proof of the proposition for various values of σ yielding Gini coefficients of various magnitudes around .72.

For intermediate levels of inequality between these, like that seen in nearly every country, the lognormal distribution has curvature that rises from $-\infty$ to above unity before gradually returning towards unity. For an example calibrated to the US income distribution (Figure 9), the crossing to above unity occurs at an income of about \$33k, between the mode and the median and the downward slope begins at about \$100k. Despite this, curvature never falls below unity again and in fact is at each quantile increasing in σ (again, not discussed in the proposition). Again taking the example of the US income-calibrated distribution, curvature peaks at about 1.21 and only falls to 1.20 by \$200k, eventually leveling out to about 1.1 for the extremely wealthy.[54] Thus, in practice, curvature is closer to flat at the top than significantly declining.

[54]Note, however, that in the true limit as $y \to \infty$, $\kappa \to 1$. However, in practice this occurs at such high income levels that the asymptote to a bit above 1 is a more realistic representation.

Proof. For a lognormal distribution with parameters (μ, σ), $F(x) = \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)$, so that

$$Q(p) = 1 - \Phi\left(\frac{\log(p) - \mu}{\sigma}\right), Q'(p) = -\frac{\phi\left(\frac{\log(p)-\mu}{\sigma}\right)}{\sigma p}$$

and

$$Q''(p) = -\frac{\phi'\left(\frac{\log(p)-\mu}{\sigma}\right)}{\sigma^2 p^2} + \frac{\phi\left(\frac{\log(p)-\mu}{\sigma}\right)}{\sigma p^2} = -\frac{\phi\left(\frac{\log(x)-\mu}{\sigma}\right)}{\sigma^2 p^2}\left(\sigma + \frac{\log(x) - \mu}{\sigma}\right).$$

where the second equality follows from the identities regarding the normal distribution from the previous proof and $y \equiv \frac{\log(p)-\mu}{\sigma}$. Thus

$$\kappa\left(p(y)\right) = \frac{(y+\sigma)\left[1 - \Phi(y)\right]}{\phi(y)}. \tag{30}$$

Note that we immediately see, as discussed above, that κ increases in σ at each quantile as the inverse hazard rate $\frac{1-\Phi}{\phi} > 0$; similarly, for any quantile associated with y, $\kappa \to \infty$ as $\sigma \to \infty$ so it must be that the set of y for which $\kappa > 1$ a) exists for sufficiently large σ and b) expands monotonically in σ. This implies that, if point 2) of the proposition is true, $\underline{y}$ must strictly decrease in σ. This also implies that for sufficiently large σ, $\kappa > 2$ for some y.

Now note that $\lim_{y\to\infty} \frac{y[1-\Phi(y)]}{\phi(y)} = 1$. To see this, note that both the numerator and denominator converge to 0 as $1 - \Phi$ dies super-exponentially in y. Applying L'Hôpital's rule:

$$\lim_{y\to\infty} \frac{y\left[1 - \Phi(y)\right]}{\phi(y)} = \lim_{y\to\infty} \frac{1 - \Phi(y) - \phi(y)y}{\phi'(y)} = \frac{y\phi(y) - \left[1 - \Phi(y)\right]}{y\phi(y)} = \frac{0}{0}.$$

where the first equality follows from the identity for ϕ' we have repeatedly been using, and from here on we no longer note the use of. Again applying L'Hôpital's rule:

$$\lim_{y\to\infty} \frac{y\left[1 - \Phi(y)\right]}{\phi(y)} = \lim_{y\to\infty} \frac{\phi(y) + y\phi'(y) + \phi(y)}{\phi(y) + y\phi'(y)} = \lim_{y\to\infty} \frac{2\phi(y) - y^2\phi(y)}{\phi(y) - y^2\phi(y)} = \lim_{y\to\infty} \frac{2 - y^2}{1 - y^2} = 1.$$

The same argument, but one step less deep, shows that $\lim_{y\to\infty} \frac{\sigma[1-\Phi(y)]}{\phi(y)} = 0$. Together these imply that $\lim_{y\to\infty} \kappa\left(p(y)\right) = 1$ and thus that, if $\kappa > 1$ at some point, it must eventual decrease to reach 1.

Similar methods may be used to show, as discussed in the text, that $\kappa \to -\infty$ as $y \to -\infty$. Furthermore we know from the proof for the normal distribution above that $\frac{y[1-\Phi(y)]}{\phi(y)}$ is monotone increasing and that $\frac{\sigma[1-\Phi(y)]}{\phi(y)}$ is monotone decreasing. The latter point implies that the set of y for which κ is decreasing must be strictly increasing in σ and thus that, if point 1) of the proposition is true, then $\overline{y}$ must strictly decrease in σ.

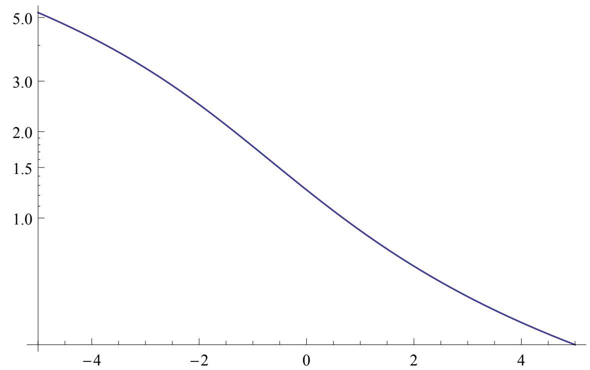

Figure 10: The figure shows the value, in logarithmic scale, of the left-hand side of inequality 31.

All that remains to be shown is that κ's comparison to unity and the sign of κ' obey the threshold structure posited. Note that we only need to show the cut-off structure for κ' and that this immediately implies the structure for κ, given the smoothness of all functions involved, because if κ increases up to some threshold and then decreases monotonically while reaching an asymptote of unity, it must lie above unity above some threshold. Otherwise, if it ever crossed below unity, it would have to be increasing in some region to asymptote to unity at very large p, violating the threshold structure for κ'. Furthermore, the same logic implies that that the region where $\kappa > 1$ must be strictly larger than the region where $\kappa' < 0$ (that $\overline{y} > \underline{y}$) as κ must rise strictly above unity before sloping strictly down towards it.

We drop arguments wherever possible in what follows to ease readability. We use the symbol $\propto$ to denote expressions having the same sign, not proportionality as is typical.

$$\kappa' = \frac{(1-\Phi)\,\phi - (y+\sigma)\phi^2 - (y+\sigma)(1-\Phi)\phi'}{\phi^2} = \frac{1-\Phi-(y+\sigma)\left[\phi - y(1-\Phi)\right]}{\phi} \propto$$

$$1-\Phi-(y+\sigma)\left[\phi - y(1-\Phi)\right] \propto \frac{1-\Phi}{\phi - y(1-\Phi)} - y - \sigma.$$

where the last sign relationship follows by the common inequality that $\phi(y) > y\left[1-\Phi(y)\right]$. Thus $\kappa' > 0$ if and only if

$$\frac{1-\Phi}{\phi - y(1-\Phi)} - y > \sigma. \tag{31}$$

Figure 10 shows that the left-hand side of this inequality is strictly decreasing. We have not found a simple means to prove this formally, but it is clearly true by inspection of the figure. Thus the left-hand side of inequality 31 must cross σ at most once and this must be from above to below.

It only remains to show that this expression does, in fact, make such as single crossing for all values of σ. It suffices to show that the small y limit of the left-hand side of inequality 31 is ∞ and that its large y limit is 0. We show these in turn.

The first claim is easy: clearly $-y\,(1-\Phi) \to \infty$, while $1-\Phi$ is finite, as $y \to -\infty$. Thus the first term approaches 0 and the second ∞ as $y \to -\infty$.

The second claim is more delicate. The expression is the same as

$$\frac{(1-\Phi)\,(1+y^2) - y\phi}{\phi - y\,(1-\Phi)}.$$

This asymptotes to the indefinite expression $\frac{0}{0}$ as $y \to \infty$ as it is well-known that $\lim_{y\to\infty} \frac{\phi}{y(1-\Phi)} = 1$. Applying L'Hôpital's rule yields

$$\lim_{y\to\infty} \frac{(1-\Phi)\,(1+y^2) - y\phi}{\phi - y\,(1-\Phi)} = \lim_{y\to\infty} \frac{-\phi\,(1+y^2) + 2y\,(1-\Phi) - \phi - y\phi'}{\phi' - (1-\Phi) + y\phi} =,$$

, applying now familiar tricks,

$$\lim_{y\to\infty} 2\frac{\phi - y(1-\Phi)}{1-\Phi} = \frac{0}{0}.$$

Again, we apply L'Hôpital's rule:

$$\lim_{y\to\infty} 2\frac{\phi - y(1-\Phi)}{1-\Phi} = \lim_{y\to\infty} 2\frac{\phi' - (1-\Phi) + y\phi}{\phi} = \lim_{y\to\infty} -\frac{1-\Phi}{\phi} = 0.$$

□

Even the slight decline in the lognormal distribution's curvature at very high incomes is an artifact of its poor fit to incomes distributions at very high incomes. It is well-known that at very high incomes the lognormal distribution fits poorly; much better fit is achieved by distributions with fatter (Pareto) tails, especially in countries with high top-income shares like the contemporary United States (Atkinson et al., 2011). A much better fit is achieved by the dPln distribution (Reed, 2003). Figure 11's left panel shows curvature as a function of income for the parameters Reed estimates (for the 1997 US income distribution). Curvature monotonically increases up the income distribution.

However it levels off at quite moderate income (it is essentially flat beyond \$100k) and at a lower level (≈ 1.04) than under the log-normal calibration, except at exorbinate incomes, where the lognormal distribution has thin tails. Thus it actually has a *thinner* tail, except at the very extreme tail, than the log-normal calibration, paradoxically. This is because Reed calibrated only to the mid-section of the US income distribution, given that the survey he used is notoriously thin and inaccurate at higher incomes; this led him to estimate a very

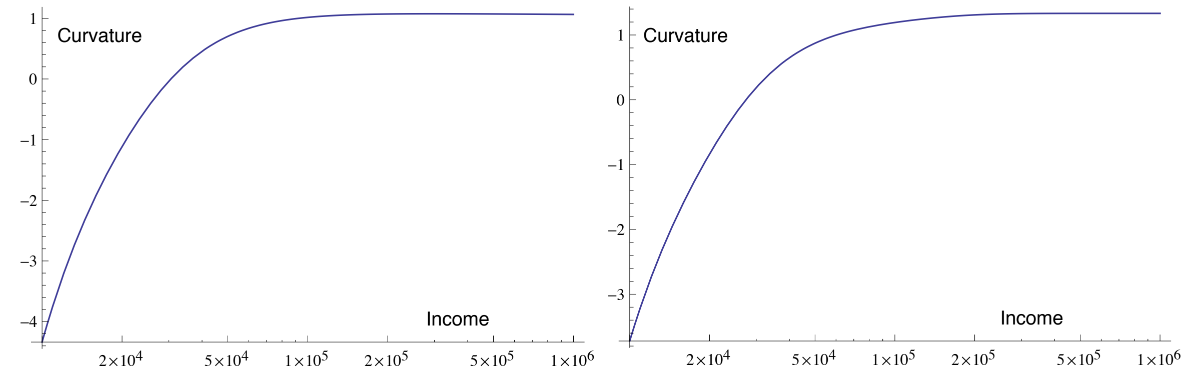

Figure 11: Curvature of the double Pareto-lognormal distribution lognormal under parameters estimated by (Reed, 2003) (left) and by updated by us (right); parameters in the former case are $\alpha = 22.43, \beta = 1.43, \mu = 10.9, \sigma = .45$ ad in the latter case are $\alpha = 3, \beta = 1.43, \mu = 10.9, \sigma = .5$. The x-axis has a logarithmic scale in income.

high (thin-tailed) Pareto coefficient in the upper tail of 22.43. Consensus economic estimates, for example Diamond and Saez (2011), suggest that that 1.5-3 is the correct range for the Pareto coefficient of the upper tail of the income distribution in the 2000's.

We therefore construct our own calibration consistent with that finding. To be conservative we set the upper tail Pareto coefficient to 3, maintain $\beta = 1.43$ to be consistent with Reed and because the lower-tail is both well-measured in his data and has not changed dramatically in the last decade and a half (Saez, 2013). We then adjust μ and σ in the unique way, given these coefficients, to match the latest US post-tax Gini estimates (.42), using a formula derived by Hajargasht and Griffiths (2013), and average income ($53k). This yields the plot in the right panel of Figure 11. There curvature continues to monotonically increase at a significant rate up to quite high incomes: at $50k it is .87, at $100k it is 1.19 and by $200k it has leveled off at 1.31, near its asymptotic value of $1 + \frac{1}{\alpha} = \frac{4}{3}$. It is this last calibration that we use to represent the dPln calibration US income distribution in the text.

Moreover, the monotone increasing nature of curvature is not only true in the US data. While we have not been able to prove any general results about this four-parameter class, we have calculated similar plots to Figure 11 for every country for which a dPln income distribution has been estimated, as collected by Hajargasht and Griffiths. In every case curvature is monotone increasing in income, though in some cases it levels off at a quite low level of income (typically when the Gini is high relative to the upper tail estimate). Even this leveling off seems to us likely to be a bit of an artifact, arising from the lack of reliable top incomes tax data in many of the developing countries on which Hajargasht and Griffiths focus. In any case, it appears that a "stylized fact" is that a reasonable model of most country's income distributions has curvature that is significantly below unity among the poor, rises above unity for the rich and monotone increasing over the full range so long as top income inequality is significant relative to overall inequality.

D.2 Laplace inverse demand functions

The following table contains Laplace inverse demand functions corresponding to inverse demand functions used in the literature. Although for most Laplace inverse demand functions we include only a few terms, closed-form expressions for all terms exist and will be included in future versions of this draft. A few other Laplace inverse demand functions will be added. Here p_a refers to a mass-point of magnitude p_a at location a. In the alternative notation on the lower lines, $\delta(x-a)$ refers to a mass-point of magnitude 1 at location a, i.e. to a Dirac delta function centered at a.

Constant elasticity / Pareto: $q(P) = \left(\frac{P}{\beta}\right)^{-\epsilon} \quad P(q) = \beta q^{-1/\epsilon}$

$p(t): \quad p_{\frac{1}{\epsilon}} = \beta$

$p(t): \quad \beta\delta\left(t - \frac{1}{\epsilon}\right)$

Constant pass-through / BP: $q(P) = \left(\frac{P-\mu}{\beta}\right)^{-\epsilon} \quad P(q) = \mu + \beta q^{-1/\epsilon}$

$p(t): \quad p_0 = \mu, \quad p_{\frac{1}{\epsilon}} = \beta$

$p(t): \quad \beta\delta\left(t - \frac{1}{\epsilon}\right) + \mu\delta(t)$

Gumbel distribution: $q(P) = \exp\left(-\exp\left(\frac{P-\alpha}{\beta}\right)\right) \quad P(q) = \alpha + \beta\log(-\log(q))$

$p(t): \quad p_0 = \mu, \quad p(t) = -\frac{\beta}{t} \quad \text{for} \quad t < 0$

$p(t): \quad \alpha\delta(t) - \frac{\beta 1_{t<0}}{t}$

Weibull distribution: $q(P) = e^{-\left(\frac{P}{\beta}\right)^{\alpha}} \quad P(q) = \beta(-\log(q))^{\frac{1}{\alpha}}$

$p(t): \quad \frac{(-1)^{\frac{1}{\alpha}}\beta t^{-\frac{1}{\alpha}-1}}{\Gamma\left(-\frac{1}{\alpha}\right)} \quad \text{for} \quad t < 0$

$p(t): \quad \frac{(-1)^{\frac{1}{\alpha}}\beta 1_{t<0} t^{-\frac{1}{\alpha}-1}}{\Gamma\left(-\frac{1}{\alpha}\right)}$

Fréchet distribution: $q(P) = 1 - e^{-\left(\frac{P-\mu}{\beta}\right)^{-\alpha}} \quad P(q) = \mu + \beta(-\log(1-q))^{-1/\alpha}$

$p(t): \quad p_0 = \mu, \quad p_{\frac{1}{\alpha}} = \beta, \quad p_{\frac{1}{\alpha}-1} = -\frac{\beta}{2\alpha}, \quad p_{\frac{1}{\alpha}-2} = \frac{\beta}{8\alpha^2} - \frac{5\beta}{24\alpha}, \quad \ldots$

$p(t): \quad \left(\frac{\beta}{8\alpha^2} - \frac{5\beta}{24\alpha}\right)\delta\left(t - \frac{1}{\alpha} + 2\right) + \beta\delta\left(t - \frac{1}{\alpha}\right) - \frac{\beta\delta\left(t - \frac{1}{\alpha} + 1\right)}{2\alpha} + \mu\delta(t) + \ldots$

Logistic distribution: $q(P) = \left(\exp\left(\frac{P-\mu}{\beta}\right) + 1\right)^{-1} \quad P(q) = \mu - \beta\log\left(\frac{1}{1-q} - 1\right)$

$p(t): \quad p_0 = \mu, \quad p_0^{(1)} = -\beta, \quad p_{-1} = -\beta, \quad p_{-2} = -\frac{\beta}{2}, \quad p_{-3} = -\frac{\beta}{3}, \quad p_{-4} = -\frac{\beta}{4}, \quad \ldots$

$p(t): \quad -\beta\sum_{j=1}^{\infty}\frac{\delta(j+t)}{j} + \mu\delta(t) - \beta\delta'(t)$

Log-logistic distribution: $q(P) = \left(\left(\frac{P}{\sigma}\right)^{\gamma} + 1\right)^{-1} \quad P(q) = \sigma\left(\frac{q}{1-q}\right)^{-1/\gamma}$

$p(t): \quad p_{\frac{1}{\gamma}} = \sigma, \quad p_{\frac{1}{\gamma}-1} = -\frac{\sigma}{\gamma}, \quad p_{\frac{1}{\gamma}-2} = \frac{\sigma}{2\gamma^2} - \frac{\sigma}{2\gamma}, \quad p_{\frac{1}{\gamma}-3} = -\frac{\sigma}{6\gamma^3} + \frac{\sigma}{2\gamma^2} - \frac{\sigma}{3\gamma}, \quad \ldots$

$p(t): \quad \left(\frac{\sigma}{2\gamma^2} - \frac{\sigma}{2\gamma}\right)\delta\left(t - \frac{1}{\gamma} + 2\right) + \sigma\delta\left(t - \frac{1}{\gamma}\right) - \frac{\sigma\delta\left(t - \frac{1}{\gamma} + 1\right)}{\gamma} + \ldots$

Laplace distribution ($q < \frac{1}{2}$): $q(P) = \frac{1}{2}\exp\left(\frac{\mu - P}{\beta}\right) \quad P(q) = \mu - \beta\log(2q)$

$p(t): \quad p_0 = \mu - \beta\log(2), \quad p_0^{(1)} = -\beta$

$p(t): \quad \delta(t)(\mu - \beta\log(2)) - \beta\delta'(t)$

Laplace distribution ($q > \frac{1}{2}$): $q(P) = 1 - \frac{1}{2}\exp\left(\frac{P-\mu}{\beta}\right) \quad P(q) = \mu + \beta\log(2(1-q))$

$p(t): \quad p_0 = \beta\log(2) + \mu, \quad p_{-1} = -\beta, \quad p_{-2} = -\frac{\beta}{2}, \quad p_{-3} = -\frac{\beta}{3}, \quad p_{-4} = -\frac{\beta}{4}, \quad \ldots$

$p(t): \quad \delta(t)(\beta\log(2) + \mu) - \beta\sum_{j=1}^{\infty}\frac{\delta(j+t)}{j}$

Normal distribution: $q(P) = \operatorname{erfc}\left(\frac{P-\mu}{\sqrt{2}\sigma}\right) \quad P(q) = \mu - \sqrt{2}\sigma\operatorname{erfc}^{-1}(2-q)$

$p(t): \quad p_0^{(1)} = -\sqrt{\frac{\pi}{2}}\sigma, \quad p_0^{(2)} = -\frac{1}{2}\sqrt{\frac{\pi}{2}}\sigma, \quad p_0^{(3)} = \frac{1}{24}\left(-\sqrt{2}\pi^{3/2} - 2\sqrt{2\pi}\right)\sigma, \quad \ldots$

$p(t): \quad -\sqrt{\frac{\pi}{2}}\sigma\delta'(t) - \frac{1}{2}\sqrt{\frac{\pi}{2}}\sigma\delta''(t) + \frac{1}{24}\left(-\sqrt{2}\pi^{3/2} - 2\sqrt{2\pi}\right)\sigma\delta^{(3)}(t) + \ldots$

lognormal distribution: $q(P) = \operatorname{erfc}\left(\frac{\log(P)-\mu}{\sqrt{2}\sigma}\right) \quad P(q) = \exp\left(\mu - \sqrt{2}\sigma\operatorname{erfc}^{-1}(2-q)\right)$

$p(t): \quad p_0^{(1)} = \sqrt{\frac{\pi}{2}}\left(-e^{\mu}\right)\sigma\delta'(t), \quad p_0^{(2)} = \frac{1}{4}\pi e^{\mu}\sigma^2 - \frac{1}{2}\sqrt{\frac{\pi}{2}}e^{\mu}\sigma, \quad \ldots$

$p(t): \quad \left(\frac{1}{4}\pi e^{\mu}\sigma^2 - \frac{1}{2}\sqrt{\frac{\pi}{2}}e^{\mu}\sigma\right)\delta''(t) - \sqrt{\frac{\pi}{2}}e^{\mu}\sigma\delta'(t) + \ldots$

Almost Ideal Demand System: $q(P)=\frac{\alpha+\beta\log(P)}{P}\quad P(q)=-\frac{\beta W\left(-\frac{qe^{-\frac{\alpha}{\beta}}}{\beta}\right)}{q}$

$p(t):\quad p_0=e^{-\frac{\alpha}{\beta}},\ \ p_{-1}=\frac{e^{-\frac{2\alpha}{\beta}}}{\beta},\ \ p_{-2}=\frac{3e^{-\frac{3\alpha}{\beta}}}{2\beta^2},\ \ p_{-3}=\frac{8e^{-\frac{4\alpha}{\beta}}}{3\beta^3},\ \ p_{-4}=\frac{125e^{-\frac{5\alpha}{\beta}}}{24\beta^4},\ \ ...$

$p(t):\quad \frac{125e^{-\frac{5\alpha}{\beta}}\delta(t+4)}{24\beta^4}+\frac{8e^{-\frac{4\alpha}{\beta}}\delta(t+3)}{3\beta^3}+\frac{3e^{-\frac{3\alpha}{\beta}}\delta(t+2)}{2\beta^2}+e^{-\frac{\alpha}{\beta}}\delta(t)+\frac{e^{-\frac{2\alpha}{\beta}}\delta(t+1)}{\beta}+\ldots$

Constant superelasticity: $q(P)=\left(\epsilon\log\left(\frac{\theta-1}{\theta P}\right)+1\right)^{\frac{\theta}{\epsilon}}\quad P(q)=\frac{(\theta-1)e^{\frac{1}{\epsilon}-\frac{q^{\epsilon/\theta}}{\epsilon}}}{\theta}$

$p(t):\quad p_0=e^{\frac{1}{\epsilon}}-\frac{e^{\frac{1}{\epsilon}}}{\theta},\ \ p_{-\frac{\epsilon}{\theta}}=\frac{e^{\frac{1}{\epsilon}}}{\theta\epsilon}-\frac{e^{\frac{1}{\epsilon}}}{\epsilon},\ \ p_{-\frac{2\epsilon}{\theta}}=\frac{e^{\frac{1}{\epsilon}}}{2\epsilon^2}-\frac{e^{\frac{1}{\epsilon}}}{2\theta\epsilon^2},\ \ p_{-\frac{3\epsilon}{\theta}}=\frac{e^{\frac{1}{\epsilon}}}{6\theta\epsilon^3}-\frac{e^{\frac{1}{\epsilon}}}{6\epsilon^3},\ \ ...$

$p(t):\quad \left(\frac{e^{\frac{1}{\epsilon}}}{2\epsilon^2}-\frac{e^{\frac{1}{\epsilon}}}{2\theta\epsilon^2}\right)\delta\left(t+\frac{2\epsilon}{\theta}\right)+\delta(t)\left(e^{\frac{1}{\epsilon}}-\frac{e^{\frac{1}{\epsilon}}}{\theta}\right)+\left(\frac{e^{\frac{1}{\epsilon}}}{\theta\epsilon}-\frac{e^{\frac{1}{\epsilon}}}{\epsilon}\right)\delta\left(t+\frac{\epsilon}{\theta}\right)+\ldots$

Cauchy distribution: $q(P)=\frac{\tan^{-1}\left(\frac{a-P}{b}\right)}{\pi}+\frac{1}{2}\quad P(q)=a+b\tan\left(\pi\left(\frac{1}{2}-q\right)\right)$

$p(t):\quad p_1=\frac{b}{\pi},\ \ p_0=a,\ \ p_{-1}=-\frac{\pi b}{3},\ \ p_{-3}=-\frac{\pi^3 b}{45},\ \ p_{-5}=-\frac{2\pi^5 b}{945},\ \ p_{-7}=-\frac{\pi^7 b}{4725},\ \ ...$

$p(t):\quad a\delta(t)+\frac{b\delta(t-1)}{\pi}-\frac{1}{3}\pi b\delta(t+1)-\frac{1}{45}\pi^3 b\delta(t+3)-\frac{2}{945}\pi^5 b\delta(t+5)-\frac{\pi^7 b\delta(t+7)}{4725}+\ldots$

Singh Maddala distribution: $q(P)=\left(\left(\frac{P}{b}\right)^a+1\right)^{-\tilde{q}}\quad P(q)=b\left(q^{-\frac{1}{\tilde{q}}}-1\right)^{\frac{1}{a}}$

$p(t):\quad p_{\frac{1}{a\tilde{q}}}=b,\ \ p_{-\frac{a-1}{a\tilde{q}}}=-\frac{b}{a},\ \ p_{-\frac{2a-1}{a\tilde{q}}}=\frac{b}{2a^2}-\frac{b}{2a},\ \ p_{-\frac{3a-1}{a\tilde{q}}}=-\frac{b}{6a^3}+\frac{b}{2a^2}-\frac{b}{3a},\ \ ...$

$p(t):\quad \left(\frac{b}{2a^2}-\frac{b}{2a}\right)\delta\left(\frac{2a-1}{a\tilde{q}}+t\right)+b\delta\left(t-\frac{1}{a\tilde{q}}\right)-\frac{b\delta\left(\frac{a-1}{a\tilde{q}}+t\right)}{a}+\ldots$

Tukey lambda distribution: $q(P)=P^{(-1)}(P)\quad P(q)=\frac{(1-q)^\lambda-q^\lambda}{\lambda}$

$p(t):\quad p_{-\lambda}=-\frac{1}{\lambda},\ \ p_0=\frac{1}{\lambda},\ \ p_{-1}=-1,\ \ p_{-2}=\frac{\lambda}{2}-\frac{1}{2},\ \ p_{-3}=-\frac{\lambda^2}{6}+\frac{\lambda}{2}-\frac{1}{3},\ \ ...$

$p(t):\quad \left(-\frac{\lambda^2}{6}+\frac{\lambda}{2}-\frac{1}{3}\right)\delta(t+3)+\frac{\delta(t)}{\lambda}+\left(\frac{\lambda}{2}-\frac{1}{2}\right)\delta(t+2)-\frac{\delta(t+\lambda)}{\lambda}-\delta(t+1)+\ldots$

Wakeby distribution: $q(P)=P^{(-1)}(P)\quad P(q)=\mu-\frac{\gamma\left(1-q^{-\delta}\right)}{\delta}+\frac{\alpha\left(1-q^{\beta}\right)}{\beta}$

$p(t):\quad p_0=\frac{\alpha}{\beta}-\frac{\gamma}{\delta}+\mu,\ \ p_{-\beta}=-\frac{\alpha}{\beta},\ \ p_\delta=\frac{\gamma}{\delta}$

$p(t):\quad \delta(t)\left(\frac{\alpha}{\beta}-\frac{\gamma}{\delta}+\mu\right)-\frac{\alpha\delta(t+\beta)}{\beta}+\frac{\gamma\delta(t-\delta)}{\delta}+\ldots$

E Details of Applications

E.1 Antràs-Chor

In this subappendix we include additional details about our method of solving the AC model and illustrate a quadratic solution involving only a single implicit solution to a fully explicit equation..

WE PLAN TO ADD THE ADDITIONAL DETAILS IN THE NEXT DRAFT.

We now consider the solution of the restricted AC model in the case. As in the relaxed solution, consider the optimal choice of a path for β subject to producing a total quantity $\hat{q}$. Note that $q(j;\beta)$ is a strictly increasing function of j for any path of β achieving $\hat{q}$ by definition. Thus it is equivalent, instead of solving for the optimal restricted β for each j, to solve for the optimal $\beta^{\star\star}$ for each $q(j;\beta)\in[0,\hat{q}]$ and then invert the resulting $q(j;\beta^{\star\star})$ function to recover the value optimal β at each j. This method preserves the separability we used in the relaxed problem and thus greatly simplifies the restricted problem. Wherever it does not create confusion we suppress as many arguments as possible, especially the dependence on β, to preserve notational economy.

By the same arguments as in the restricted case, the cost of production $\hat{q}$ is $C(\hat{q};\beta)$ where

$$C(\hat{q};\beta) = \int_0^{\hat{q}} [1-\beta(q)]\, MR(q)dq,$$

where $\beta(q)$ is a notationally-abusive contraction of $\beta(j(q;\beta))$. However, to actually produce $\hat{q}$, we need

$$\int_0^1 S([1-\beta(q(j))]\, MR(q(j)))\, dj = \hat{q},$$

where $S = MC^{-1}$, the supply curve, exists because of our assumption that MC is strictly monotone increasing. Changing variables so that both integrals are taken over j:

$$C(\beta) = \int_0^1 [1-\beta(q(j))]\, MR(q(j))\, S([1-\beta(q(j))]\, MR(q(j)))\, dj.$$

Thus the firm solves a Lagrangian version of this problem that is separable in each j, or equivalently q:

$$\max_{\beta} \int_0^1 \lambda S([1-\beta(q(j))]\, MR(q(j))) - ([1-\beta(q)]\, MR(q) S([1-\beta(q(j))]\, MR(q(j))))\, dj - \lambda\hat{q}.$$

At each q this is a simple maximization problem. The firm chooses the value of β maximizing

$$\lambda S([1-\beta(q)]\, MR(q)) - [1-\beta(q)]\, MR(q) S([1-\beta(q)]\, MR(q)),$$

the difference between the total value of the production by that firm and the total cost of that production. Clearly both terms are decreasing in β given that $MR > 0$ in any range where the firm would consider producing, so given that the firm chooses between only two values of β, $\beta_I > \beta_O$, the firm will strictly choose in-sourcing if and only if

$$MR(q) > \frac{\lambda [S([1-\beta_O]\, MR(q)) - S([1-\beta_I]\, MR(q))]}{[1-\beta_O]\, S([1-\beta_O]\, MR(q)) - [1-\beta_I]\, S([1-\beta_I]\, MR(q))}. \tag{32}$$

If the sign here is equality (which generically occurs on a set of measure 0 so long as the functions are nowhere constant relative to one another) then the firm is indifferent and if the inequality is reversed the firm strictly chooses in-sourcing. As λ rises, the firm will in-source less and produce more; thus varying λ over all positive numbers traces out all potentially optimal solutions. Note that this could easily be extended to a situation where the firm has any simple restricted choice of β, not just two values.

Furthermore, once $\beta(q)$ is set, we can easily recover the optimal $\beta^{\star\star}$ for each j by noting that the optimal value of $\beta^{\star\star}$ at $\tilde{j}$ is the optimal value at $\tilde{q}$ satisfying the production equation

$$\int_0^{\tilde{j}} S\left(\left[1-\beta^{\star\star}\left(q(j)\right)\right] MR\left(q^{\star\star}(j)\right)\right) dj = \tilde{q}.$$

This implies the differential equation $q'(j) = S\left(\left[1-\beta^{\star\star}\left(q(j)\right)\right] MR\left(q^{\star\star}(j)\right)\right)$ and thus the inverse differential equation $j'(q) = \frac{1}{S\left(\left[1-\beta^{\star\star}(q)\right] MR\left(q^{\star\star}\right)\right)}$ which together with the boundary condition $j(0)=0$ yields $j(q)$ and thus $\beta^{\star\star}$ at each j.

It remains only to pin down the optimal value of λ. To do this, denote the set of q on which Inequality 32 is satisfied $B_I(\lambda)$ and on which it is reversed $B_O(\lambda)$.[55] Total production is

$$q_\lambda = \int_{j\in(0,1):q(j)\in B_I(\lambda)} S\left(\left(1-\beta_I\right) MR\left(q(j)\right)\right) dj + \int_{j\in(0m1):q(j)\in B_O(\lambda)} S\left(\left(1-\beta_O\right) MR\left(q(j)\right)\right) dj,$$

while total cost $C_\lambda =$

$$\int_{B_I(\lambda)\cap(0,q_\lambda)} \left[1-\beta_I\right] MR\left(q\right) dq + \int_{B_O(\lambda)\cap(0,q_\lambda)} \left[1-\beta_O\right] MR\left(q\right).$$

Profit is

$$R\left(q_\lambda\right) - C_\lambda$$

and the first-order condition for its maximization is

$$MR\left(q_\lambda\right)\frac{\partial q_\lambda}{\partial\lambda} - \frac{\partial C_\lambda}{\partial\lambda} = 0 \implies MR\left(q_\lambda\right) = \frac{\frac{\partial C_\lambda}{\partial\lambda}}{\frac{\partial q_\lambda}{\partial\lambda}} = \lambda,$$

because λ is defined as the shadow cost of relaxing the constraint on production.

Now we consider obtaining as close as possible to an explicit solution. Note that, to do so, we must be able to characterize S, B_O and B_I explicitly. S is the inverse of MC and thus MC must admit an explicit inverse. To characterize B_O and B_I explicitly requires solving Inequality 32 with equality to determine the relevant thresholds, which, as we will see, requires marginal revenue to have an explicit inverse.

One of the simplest forms satisfying these conditions and yet yielding our desired non-monotonicity is $P(q) = p_0 + p_{-t}q^t + p_{-2t}q^{2t}$ and $MC(q) = mc_{-t}q^t$, where $t, p_0, p_{-t}, mc_{-t} >$

[55]We ignore the generically 0-measure set on which it is an equality.

$0 > p_{-2t}$. In this case $S(p) = \left(\frac{p}{mc_{-t}}\right)^{\frac{1}{t}}$. Thus the equality version of Inequality 32 becomes

$$MR(q) = \frac{\lambda\left(\left[\frac{(1-\beta_O)MR(q)}{mc_{-t}}\right]^{\frac{1}{t}} - \left[\frac{(1-\beta_I)MR(q)}{mc_{-t}}\right]^{\frac{1}{t}}\right)}{(1-\beta_O)\left[\frac{(1-\beta_O)MR(q)}{mc_{-t}}\right]^{\frac{1}{t}} - (1-\beta_I)\left[\frac{(1-\beta_I)MR(q)}{mc_{-t}}\right]^{\frac{1}{t}}} \implies$$

$$\implies MR(q) = \frac{\lambda\left[(1-\beta_O)^{\frac{1}{t}} - (1-\beta_I)^{\frac{1}{t}}\right]}{(1-\beta_O)^{\frac{1+t}{t}} - (1-\beta_I)^{\frac{1+t}{t}}} \equiv \lambda k,$$

where k is the relevant collection of constants. Note that this is an extremely simple threshold rule in terms of marginal revenue. Given that we have chosen a form of marginal revenue that admits an inverse, it is simple to solve out for the threshold rule in terms of quantities; this is why we needed marginal revenue to have an inverse solution.

$$p_0 + (1+t)p_{-t}q^t + (1+2t)p_{-2t}q^{2t} = \lambda k \implies$$

$$q = \left(\frac{-p_{-t}(1+t) \pm \sqrt{p_{-t}(1+t)^2 + 4\left(p_0 - k\lambda\right)p_{-2t}(1+2t)}}{2p_{-2t}(1+2t)}\right)^{\frac{1}{t}}.$$

Between these two roots, in-sourcing is optimal; outside them, outsourcing is optimal.[56]

This provides closed-form solutions as a function of λ, but λ remains to be determined. This is, unfortunately, where things start to get a bit messier. The integral determining q_λ can be explicitly taken, but only in terms of the less-standard Appell Hypergeometric function. The equation for $MR\left(q_\lambda\right) = \lambda$ therefore cannot be solved explicitly for λ. However, it is a single explicit equation. Once λ has been determined, optimal sourcing is determined in closed-form as described above. We plot this and the relaxed optimal β, in Figure 12, in the same format as in the text for the case when $p_0 = .2, p_{-t} = 2, p_{-2t} = -4, mc_{-t} = .5, t = .5, \beta_I = .8, \beta_O = .3$. Clearly we obtain similar, non-monotone results, but now these require only a single call of Newton's method to solve an otherwise explicit equation, as opposed to the two-dimension search we required to solve the case presented in the text.

We do not discuss second-order conditions here, but they can easily be derived and checked to hold for this example as well as for the example in the text. A grossly sufficient condition is that marginal revenue is declining over the solution range, as is the case in both of these examples.

[56] Actually if $\lambda k < p_0$ then the lower root should be interpreted as 0.

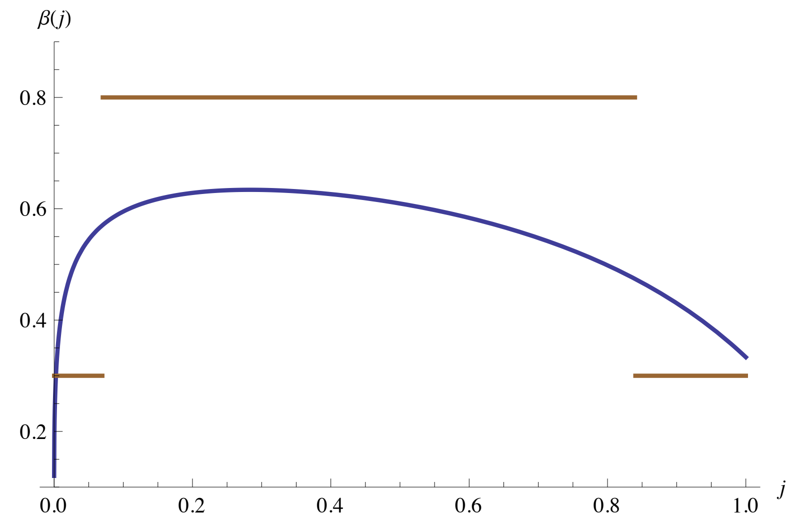

Figure 12: Relaxed and restricted solutions to the AC model when $P(q) = .2 + 2q^{\frac{1}{2}} - 4q$, $MC(q) = \frac{q^{\frac{1}{2}}}{2}$, $\beta_O = .3$ and $\beta_I = .8$.

E.2 Stole-Zweibel

FORTHCOMING IN THE NEXT DRAFT.

E.3 Monopolistic competition

The following brief discussion will be extended in the next draft.

E.3.1 Flexible Krugman model

Here we briefly discuss the logic behind the observation in the main text. It is convenient to express the model's equations in terms of the equilibrium level of marginal cost $MC^{\star}$.[57] Output optimally designated for the domestic market and the export market will satisfy $R'(q_d) = MC^{\star}$ and $R'(q_x) = \tau\, MC^{\star}$, respectively, and therefore may be solved for in closed form in terms of $MC^{\star}$ for tractable specifications of the revenue function (or consumer preferences).[58] The same is true for wages, since $w = MC^{\star}/L'(q_d + \tau q_x)$.

For a chosen $MC^{\star}$ we may compute the level of fixed cost f consistent with it using the

[57] The case of a single country corresponds to the Dixit-Stiglitz model. It may be obtained from our two-country discussion by setting $\tau \to \infty$ and $q_x = 0$. In this case one does not have to express the model's equations in terms of the equilibrium level of marginal cost $MC^{\star}$ as we do below. Instead, for tractable functions $R\,(q)$ and $L\,(q)$ one can solve for equilibrium quantity $q^{\star}$ in closed form (in terms of the fixed cost of production f) from an equation that combines profit maximization and free entry: $(L(q) + f)R'(q) = R(q)L'(q)$.

[58] As mentioned in the main text, a convenient choice of numéraire allows us to keep the revenue function $R\,(.)$ independent of economic circumstances.

free-entry condition: $R(q_d)+R(q_x) = wL(q_d+\tau q_x)+wf$. The equilibrium number (measure) $N^\star$ of firms in each economy then satisfies $N^\star = L_E/(L(q_d+q_x)+f)$, where L_E is the labor labor endowment one of the two economies.[59] Other variables of interest, e.g. trade flows or welfare, are then simply functions of the ones discussed above. A future draft of this paper will contain a more detailed explanation.

Krugman model with non-iceberg and iceberg international trade costs. Although the Krugman model with non-iceberg trade costs is not our main focus here, we mention it for completeness. Let us assume the presence of non-iceberg international trade costs that require hiring labor $L_T(q_x)$ in order for q_x to reach its destination in the other country.[60] The export FOC is now $R'(q_x) - wL_T'(q_x) = \tau MC^\star$, while the free entry condition becomes $R(q_d) + R(q_x) = wL(q_d + \tau q_x) + wL_T(q_x) + wf$. The resulting number (measure) of firms is $N^\star = L_E/((L(q_d+q_x)+f)+L_T(q_x))$. The model may be solved explicitly along the same lines in terms of chosen $MC^\star$ and w, with f and τ treated as derived quantities.

E.3.2 Flexible Melitz model

Single country. The firm profit maximization condition and the free entry condition are

$$R'(q) = wL'(q;a), \tag{33}$$

$$R(q_c) = wL(q_c;a_c) + wf. \tag{34}$$

A convenient solution strategy is to choose q_c and then calculate f_e as a derived quantity. For a chosen q_c we can find a_c explicitly by combining (33) and (34) into $R'(q_c)(L(q_c;a_c)+f) = R(q_c)L'(q_c;a_c)$, since $L(q;a)$ is assumed to be tractable also with respect to a. Wages are then given recovered from (34): $w = R(q_c)/(L(q_c;a_c)+f)$.

Now we need to show how to calculate the fixed cost of entry f_e and the measure of firms. The fixed cost of entry consistent with the chosen cutoff quantity is given simply by the unrestricted entry condition:

$$w\delta_e f_e = \bar{\Pi} = \int_{q \geq q_c} (R(q) - wL(q;a) - wf)\, dG(a(q)).$$

Here $a(q)$ is the firm's productivity parameter as an explicit function of the optimally chosen

[59] L_E may be exogenous, as in the original Krugman model, but even for endogenous labor supply it is possible to obtain fully explicit solutions to the model in terms of the parameter $MC^\star$.

[60] In a symmetric equilibrium it does not matter how this labor is split between the countries, as long as symmetry of the model is maintained. For asymmetric countries, we could assume that the transport requires labor from both countries. The model may be solved in terms of marginal costs of serving each market.

quantity q that results from using (33). For Pareto G, and L and R tractable from the point of view of q (but not necessarily having a linear term) and $L(q;a)$ linear in a, there exist closed-form expressions for this integral in terms of special functions, which are straightforward to derive, especially if one uses symbolic manipulation software such as Mathematica. If the shape parameter of the Pareto distribution is a negative integer, the integrals actually reduce to simple power functions.

If M_e denotes the measure of firms that enters each period (in one country), then the measure of operating firms is $M = G(a_c) M_e/\delta_e$. The total labor used in the economy is given by $L_E = M_e f_e + Mf + M\bar{L}$, where $\bar{L} = G(a_c)^{-1} \int_{q \geq q_c} L(q;a)\, dG(a(q))$ is the labor on average hired for the variable cost of production. Under the same assumptions, the integral again has an explicit form in terms of special functions. We see that in these cases we can get fully explicit expressions for f_e and M in terms of chosen q_c and L_E.

Other quantities of interest, such as trade flows or welfare, may be found in an analogous fashion. In a future draft of this paper we will provide a more detailed discussion.

Two countries with non-iceberg and iceberg international trade costs. Our solution strategy is treat MC_{1n} and MC_{2x} as given and to express other variables of the model in terms to these two chosen parameters. In particular, we will show how to derive explicit expressions for the fixed cost of exporting f_x and cost of entry f_e. The (variable-cost) labor requirement $L(q;a)$ is assumed to be a tractable combination of equidistant powers of a, with coefficients that in general depend on q. Firms' profit maximization leads to the set of equations:

$$MC_n = R'(q_n) \tag{35a}$$

$$MC_n = wL'(q_n;a) \tag{35b}$$

$$MC_x = R'(q_d) \tag{35c}$$

$$MC_x = \tfrac{1}{\tau} R'(q_f) - \tfrac{1}{\tau} wL'_T(q_f) \tag{35d}$$

$$MC_x = wL'(q_d + \tau q_f; a) \tag{35e}$$

$$R(q_{1n}) - wL(q_{1n};a_1) = f \tag{35f}$$

$$R(q_{2d}) + R(q_{2f}) - wL(q_{2d} + \tau q_{2f}; a_2) - wL_T(q_{2f}) = f + f_x \tag{35g}$$

Here q_n is the quantity sold by a non-exporting firm, while q_d and q_f represent quantities that reach domestic and foreign customers of an exporting firm, respectively. In addition to exporting cost $wL_T(q_f)$, we allow for an iceberg trade cost factor $\tau \geq 1$.

For a chosen MC_{1n}, we can calculate q_{1n} from (35a). The corresponding a_1 may be found by solving a linear equation that results from combining (35b) and (35f) in a way that eliminates wages. Wages then may be recovered by substituting back to (35b).

For a chosen MC_{2x}, we can derive q_{2d} from (35c) and q_{2f} from (35d). The value of a_2 is then determined by (35e). We find q_{2n} by solving (35a) and (35b) with MC_{2n} eliminated, and then in turn use one of these to find MC_{2n}. This means that we know the marginal cost at the cutoffs. Fixed cost of exporting f_e is then identified from (35g).

For a given marginal cost, we can find the corresponding quantities and productivity parameters a by a similar method from (35a-35e), this time treating w as known. We denote the resulting functions $q_n(MC_n)$, $q_d(MC_x)$, $q_x(MC_x)$, $a_n(MC_n)$, and $a_x(MC_x)$. Using these functions we can now determine the entry labor requirement f_e from the unrestricted entry condition:

$$w\delta_e f_e = \bar{\Pi} = \int\limits_{y\in S_n} \Pi\left(q_n(y); a_n(y)\right) dG\left(a_n(y)\right) + \int\limits_{y\in S_x} \Pi\left(q_x(y); a_x(y)\right) dG\left(a_x(y)\right),$$

where Π is the profit function (revenue minus cost), $G(a)$ is the cumulative distribution function of a, and the integration ranges are $S_n \equiv (MC_{2n}, MC_{1n})$ and $S_x \equiv (0, MC_{1n})$. Under various assumptions these integrals may be evaluated in closed form, often involving special functions. If a measure M_e of firms enters each period (in one of the countries), then the equilibrium measure of operating firms is $M = M_e G(a_1)/\delta_e$ and that of exporting firms is $M_x = M_e G(a_2)/\delta_e$. These measures may be calculated from the labor market clearing condition $M_e f_e + Mf + M_x f_x + (M - M_x)\bar{L}_n + M_x\bar{L}_x = L_E$, where

$$\bar{L}_n \equiv \tfrac{1}{G(a_1)-G(a_2)} \int\limits_{y\in S_n} L\left(q_n(y); a_n(y)\right) dG\left(a_n(y)\right),\ \bar{L}_x \equiv \tfrac{1}{G(a_2)} \int\limits_{y\in S_x} L\left(q_x(y); a_x(y)\right) dG\left(a_x(y)\right).$$

Under the same assumptions as before, these integrals may be evaluated in closed form. Again, other variables of interest, such as trade flows or welfare, may be obtained in a similar way.

E.3.3 Flexible Melitz/Melitz-Ottaviano model with non-separable utility

While a significant part of the international trade literature relies on separable utility functions, there exist realistic economic phenomena what are more easily modeled with non-separable utility. An instantly classic alternative to the Melitz model that uses non-separable utility is the model of Melitz and Ottaviano, which assumes that with greater selection of heterogeneous-good varieties available to consumers, the marginal gain from an additional variety decreases relative to the gains from increased quantity. Trade liberalization leads to tougher competition, which results not only in higher productivity, but also in the decrease of markups charged by a given firm.

Here we briefly discuss a generalization of the flexible Melitz model where the utility function is allowed to be non-separable. This generalized model contains as special cases both the Melitz model and the Melitz and Ottaviano model.[61] The utility is of the form

$$U_\Omega \equiv F\left(U_\Omega^{(1)}, U_\Omega^{(2)}, ..., U_\Omega^{(m)}\right), \quad U_\Omega^{(i)} \equiv \int_\Omega U^{(i,\omega)}\left(q_\omega\right)\, d\omega.$$

In order to preserve tractability, we assume that $U^{(i,\omega)}\left(q_\omega\right)$ are linear combinations[62] of equally-spaced powers of q_ω and that the set of exponents does not depend on i or ω. For example, we could specify $U_\Omega \equiv U_\Omega^{(1)} + \kappa_1 (U_\Omega^{(1)})^{\xi_1} + \kappa_2 (U_\Omega^{(2)})^{\xi_2}$, $U_\Omega^{(1)} \equiv \int_\Omega q_\omega^{\gamma_1}\, d\omega$, and $U_\Omega^{(2)} \equiv \int_\Omega q_\omega^{\gamma_2}\, d\omega$, with $(\gamma_1+1)/(\gamma_2+1)$ equal to the ratio of two small integers. The choice $\kappa_1 = \kappa_2 = 0$ corresponds to the Melitz model, while the choice $\xi_1 = 2$, $\xi_2 = 1$, $\gamma_1 = 1$, and $\gamma_2 = 2$ gives the Melitz and Ottaviano model, which is based on a non-homothetic quadratic utility. Our general specification allows also for homothetic non-separable utility functions that feature market toughness effects analogous to those in the Melitz and Ottaviano model.

It is straightforward to verify that just like the flexible Melitz model with separable utility, this more general version leads to tractable optimization by individual firms, as well as for tractable aggregation under the same conditions. The reason for the tractability of the firm's problem is simple: the firm's first-order condition will have the same structure as previously, a linear combination of equidistant powers (with additional dependence of the coefficients of the linear combination on *aggregate* variables of the type $\int_\Omega q_\omega^\gamma\, d\omega$ for some constants γ). Given that the nature of the firm's problem is unchanged, it follows that being able to explicitly aggregate over heterogeneous firms does not require any additional functional form assumptions relative to the separable utility case.

F Extension to Multi-Equation Models

FORTHCOMING IN NEXT DRAFT.

[61]In addition to the heterogeneous-good varieties explicitly considered here, the Melitz and Ottaviano model includes a homogeneous good. In our discussion, the homogeneous good is absent, but adding it to the model is straightforward.

[62]Of course, without loss of generality we could assume that $U^{(i,\omega)}\left(q_\omega\right)$ are power functions and let the function F combine them into any desired linear combinations. However, for clarity of notation it is preferable to keep the number m of different expressions $U_\Omega^{(i)}$ small.