RM 252

Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition

Author(s): E. Glen Weyl and Michal Fabinger

Source: *Journal of Political Economy* , Vol. 121, No. 3 (June 2013), pp. 528-583

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/10.1086/670401

Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition

E. Glen Weyl

University of Chicago

Michal Fabinger

University of Tokyo

We extend five principles of tax incidence under perfect competition to a general model of imperfect competition. The principles cover (1) the independence of physical and economic incidence, the (2) qualitative and (3) quantitative manner in which taxes are split between consumers and producers, (4) the determinants of tax pass-through, and (5) the integration of local incidence to determine the overall division of surplus. We show how these principles can be used to simplify and generalize the analysis of a range of economic questions such as the optimal procurement of new markets and the welfare effects of third-degree price discrimination.

The working paper version of this article went, for a period, under the name "A Restatement of the Theory of Monopoly." While many colleagues have provided valuable and in some cases detailed comments, special debts are due to Kevin Murphy (who originally inspired the work), Tony Atkinson and Faruk Gul (for suggesting the subtitle and title of the article, respectively), Luciano de Castro and Michael Salinger (for excellent formal conference discussions), Mark Armstrong, David Atkin and Dave Donaldson (for suggesting new results), Jesse Shapiro (for suggesting ways of reframing the paper), and above all Jeremy Bulow (for more things than can be listed here). Weyl is grateful to the Milton Fund, the Harvard Real Estate Academic Initiative, and the Marion Ewing Kauffman Foundation for their financial support, which made possible the excellent research assistance provided by Miguel Espinosa, Ahmed Jaber, Rosen Kralev, Andrew Kramer, Stephanie Lo, Yali Miao, Toan Phan, Daichi Ueda, and William Weingarten. All errors are our own.

[*Journal of Political Economy,* 2013, vol. 121, no. 3]

> We may prepare the way for using, as we go, illustrations drawn from the incidence of taxation to throw side-lights on the problem of value. For indeed a great part of economic science is occupied with the diffusion throughout the community of economic changes which primarily affect some particular branch of production or consumption; and there is scarcely any economic principle which cannot be aptly illustrated by a discussion of the shifting of the effects of some tax.(Alfred Marshall, *Principles of Economics* [1890, bk. V, chap. 9])

I. Introduction

Following Marshall (1890), standard treatments of a range of topics in perfectly competitive markets are typically taught and analyzed in relationship to tax incidence. For example, Chetty (2009) surveys recent work in public finance that builds on the fact that incidence is often a "sufficient statistic" for various welfare analyses to reduce the number of structural assumptions needed to reach welfare conclusions. Virtually all of this work, however, assumes perfect competition, whereas much of contemporary economic analysis assumes that firms have market power. In this article, we show how the principles of incidence and their use as an analytic tool extend to imperfectly competitive models. We survey, where possible, and extend, where necessary, five fundamental principles of tax incidence under perfect competition to successively more general imperfectly competitive settings: monopoly, symmetric imperfect competition, and, finally, general imperfectly competitive models. We then apply these to economic problems outside of the traditional public finance problems in which they are most familiar, ranging from the effects of third-degree price discrimination to strategic effects in oligopoly.

To motivate incidence reasoning, consider the following application, which would typically be viewed as a problem in industrial organization or mechanism design rather than public finance. Suppose, as we do in Section VI.A, that an authority can create a market and wants to select the provider(s) of a concession to maximize the social surplus this concession generates. Dasgupta and Maskin (2000) describe an auction that quite generally screens different arrangements for the profits they create. However, Borenstein (1988) argues that this procedure does not directly screen consumer surplus and therefore will not in general allocate the concession efficiently from a social perspective. Under what circumstances will the allocation be socially efficient?

Clearly a sufficient condition is that all arrangements have the same ratio of consumer, and thus social, to producer surplus. This is one of the focal incidence quantities we analyze below. Because we consider this quantity over a wide range of settings, our logic simultaneously implies that, in order for the allocation of licenses by a private market to maximize total surplus,

- if different perfectly competitive arrangements are being considered, it is the (appropriately averaged) ratio of the demand to the supply elasticities that would have to be homogeneous across offerings;
- if different constant marginal cost monopolists are being considered, demand curves must have the same (average) curvature, and in particular, it would be sufficient for them to be linked by the Weyl and Tirole (2012) stretch parameterization of demand;
- if different symmetric oligopoly settings are being considered, then an industry conduct parameter measuring the degree of monopolization ($1/n$ in the Cournot model or $1 - \sum_{j \neq i}[\partial q_j/\partial p_i]/[\partial q_i/\partial p_i]$ in the differentiated products Nash-in-prices model) needs to be homogeneous across arrangements, though different arrangements need not involve the same game (one arrangement might involve monopolistic competition and another homogeneous products with conjectural variations).

All these results follow from the same simple formula: the stated conditions are those under which the four parameters entering this formula (or a ratio of them) are homogeneous across markets. Furthermore, as we discuss in Section VI, because many applied problems of this kind all depend on similar principles of incidence, they can be treated simultaneously rather than separately. We begin in Section II by largely recapitulating these principles of incidence under perfect competition, framing them in terms of five fundamental points:

1. *Economic vs. physical incidence*: Welfare effects of taxation are independent of who physically pays the tax.
2. *Split of tax burden*: Revenue mechanically raised is paid for by reductions in welfare split between the two sides of the market.
3. *Local incidence formula*: The ratio of the tax borne by consumers to that borne by producers, the *incidence*, I, equals $\rho/(1-\rho)$, where the *pass-through rate* ρ is the rise in price to consumers for each infinitesimal unit of specific tax imposed.
4. *Pass-through*: In turn, ρ is equal to $1/[1 + (\epsilon_D/\epsilon_S)]$, where ϵ_D and ϵ_S are, respectively, the elasticities of demand and supply.
5. *Global incidence*: The incidence of a finite (i.e., noninfinitesimal) tax change is obtained by replacing the pass-through rate by its quantity-weighted average over the range of the tax change. In particular, the ratio of total consumer and producer surplus discussed

above (the incidence of a tax that eliminates the market entirely) equals the quantity-weighted average of the pass-through rate over taxes ranging from zero to the minimum tax that kills the market.

We then consider successively more general models of imperfect competition and study how each of the principles extends or needs to be modified. We begin with monopoly in Section III and note four principal modifications. First, even the first unit of a tax is now more than fully borne by the two sides as there is a deadweight burden arising from the monopoly's distortion of prices. Second, by the envelope theorem, the monopoly fully bears the tax while the consumer still bears ρ per unit quantity, so $I = \rho$ now. Thus ρ now quantifies the deadweight burden. Building on this, we observe, third, that under constant marginal cost, every unit of quantity brought in to compete in the market now has a ratio of deadweight loss reduction to profit reduction, or *social incidence*, $SI = \rho$. This, too, can be integrated up to relate the ratio of deadweight loss to producer surplus to the (now markup-weighted) average pass-through.[1] Fourth, pass-through now depends not only on the relative elasticity of supply and demand but also on the curvature of demand.

Next, in Section IV, we consider a general model of symmetric imperfect competition in which firms set the elasticity adjusted Lerner index $[(p - mc)/p]\epsilon_D = \theta$, a *conduct parameter.* The conduct parameter is equal to one under monopoly, is equal to zero under perfect competition, and is typically greater than one when firms noncooperatively price complementary goods. We show how this nests many standard symmetric models of imperfect competition as alluded to above.

Symmetric imperfect competition largely interpolates between the behavior under perfect competition and that under monopoly. In particular $I = \rho/[1 - (1 - \theta)\rho]$ and $SI = \theta\rho/[1 + (1 - \theta)\rho]$. The formula for pass-through similarly interpolates between the monopoly and competition formulas in the case in which θ is independent of market conditions, as is the case under both the Cournot competition and complements model and under the (quasi-linear version of the) Dixit and Stiglitz (1977) model of monopolistic competition. However, two new elements emerge if θ depends on market conditions. First, if, as is typically the case with differentiated products Nash-in-prices competition derived from discrete-choice behavior among consumers, θ rises as prices rise/quantities fall, then pass-through is higher than the interpolation suggests. Second, when one is taking averages for nonlocal incidence, θ also must be averaged over the relevant range.

[1] Throughout the article, we use the term "markup" to refer to the absolute markup $p - mc$ rather than the relative markup $(p - mc)/p$, to which it is sometimes referred.

Finally, in Section V, we consider our most general model with asymmetric firms, allowing for taxes or exogenous competition that falls heterogeneously on firms. Because of the notational complexity introduced by asymmetries, we do not extend all of our principles and instead focus on presenting the model and deriving the local and global forms of (possibly heterogeneous across firms) tax incidence.[2] Each firm now has its own idiosyncratic conduct parameter θ_i, defined as the ratio of the real (markups times changes in strategy) to the pecuniary (quantities times changes in price) effects induced by the firm's changing its quantity. In the symmetric case, as well as some asymmetric cases such as the Melitz (2003) model, this reduces to the more standard common conduct parameter θ.

However, in the more typical case in which θ is heterogeneous across firms, incidence now depends on the covariance between various variables. We use the independence of physical and economic incidence to characterize the effects of taxes in terms of the firms they induce to reduce quantities, saying that the tax "falls on" the firms induced by the tax to reduce quantities. While incidence on consumers depends only on the (quantity-weighted across firms) average pass-through rate, the incidence of taxes on firms is now heavier (1) the more the tax falls on firms with high θ_i, (2) the more it falls on firms with high pass-through, and (3) the more pass-through covaries with θ_i across firms. Firms and society benefit from taxes targeted at firms with negative θ_i. Such negative values of θ_i are typical, for example, among inferior products in the Shaked and Sutton (1982) model of vertically differentiated products competing in prices. These results have natural implications for global incidence: for example, firms gain a greater share of surplus to the extent that it is the large firms that have highest θ_i.

We then discuss, in Section VI, a number of other applications in which the logic of incidence guides, simplifies, unifies, or generalizes the analysis. In a supply chain or regulatory relationship with an imperfectly competitive industry, pass-through is central to optimal policy. The welfare effects of price discrimination are largely determined by comparing incidence properties in the two markets separated by discrimination. The determination of whether strategies are strategic complements or substitutes, which Bulow, Geanakoplos, and Klemperer (1985b) in turn imply a wide range of effects in oligopoly theory, is often equivalent to comparisons of pass-through or incidence to simple thresholds. We also briefly allude to similar characterizations of many issues in the analysis of platforms, mergers, product design, behavioral welfare analysis, demand systems, and the empirics of international trade. Such characterizations were

[2] Appendix C extends principle 3, and other extensions are available on request.

derived in the working paper version of this article but are omitted for the sake of brevity.

Section VII concludes the article by discussing potential directions for future research. Results requiring more involved mathematics or of less general interest are collected in appendices following the main text.

II. Perfect Competition

We begin by reviewing the analysis of incidence in a perfectly competitive market and showing how it can be extended by integration from local formulas to provide a global characterization. To facilitate comparison between the different settings we consider, we articulate the analysis in terms of five principles, reconsidered successively in each setting. Throughout we assume for simplicity that demand and supply are smooth and that excess supply declines in price so that there is a unique equilibrium.[3] We also assume, for the sake of simplicity and consistency with nearly all literature we are aware of, that all goods outside the industry of interest are supplied perfectly competitively, and thus the welfare of producers arising from consumer substitution to these goods may be ignored. The biases introduced by this assumption, which will become apparent in Section V, are discussed explicitly in the conclusion, Section VII.

The first and most basic principle of incidence, due to Jenkin (1871–72), is that of *physical neutrality*: it is irrelevant to the economic incidence of a tax whether it is physically paid by producers or consumers. Let p_c be the price paid by consumers, p_s be the price received by the suppliers, and D and S be, respectively, the quantities demanded and supplied as a function of p_c and p_s. Whether buyers or sellers pay the tax physically, equilibrium is given by

$$D(p_c) = S(p_s),$$

where $p_s = p_c - t$. Thus the equilibrium is identical in the two cases. Despite this equivalence, in many of the applications considered below, it is conventional to think of the producers as "directly" bearing the tax or cost increase and its being indirectly "shifted" or "passed through" to consumers. We therefore use $p \equiv p_c$ to refer to the price paid by the consumers, $p - t$ to denote the price received by the suppliers, and the *pass-through rate* $\rho \equiv dp/dt$ as the rate at which prices paid by consumers rise when the tax increases. Implicit to this formalism is an understanding that the amount the nominal price (to the producers) falls when a tax is levied on consumers is $1 - \rho$.

[3] However, essentially all results can be extended to the case in which these fail. Some of these extensions were included in the working paper version of this article. Details are available on request.

Principle of incidence (Perfect competition) 1. Physical incidence of taxes is neutral in the sense that a tax levied on consumers, or a unit parallel downward shift in consumer inverse demand, causes nominal prices to consumers to fall by $1-\rho$, where the pass-through rate $\rho \equiv dp/dt$ is the rate at which prices to consumers rise when a tax is imposed on producers.

Under perfect competition, both consumers and producers take prices as given and choose quantities so as to maximize their welfare. Thus, as first argued by Dupuit (1844), if we let $CS(p) = \int_p^\infty D(x)\,dx$ denote the surplus of consumers, $PS(p-t) = \int_0^{p-t} S(x)\,dx$ denote that of producers, and Q denote the equilibrium quantity, then $dCS/dt = -\rho Q$ and $dPS/dt = -(1-\rho)Q$. Therefore, the total tax burden, equal to the equilibrium quantity, is split among consumers and producers with weights ρ and $1-\rho$. Thus Jenkin observed two further principles.

Principle of incidence (Perfect competition) 2. The total burden of the infinitesimal tax (beginning from zero tax) is shared between consumers and producers.

Principle of incidence (Perfect competition) 3. The economic incidence (or incidence for short) of the infinitesimal tax (beginning from zero tax), the ratio of what is borne by consumers to that borne by producers, is given by $I = \rho/(1-\rho)$.

In what follows we will often be concerned with changes in cost that do not directly generate revenue, and thus we will focus only on the impact of an increase in a tax or cost on the market participants rather than on the revenue raised by the tax-imposing authority. Thus, while principle 2 literally applies only to infinitesimal taxes, we will refer to the "total burden" of a tax increase as Qdt even for tax changes that do not begin from $t=0$ as this is still the total mechanical burden of the taxes not mediated through endogenous responses.[4] Applying this terminology, we can drop the parenthetic references to the tax beginning from zero, as we do throughout the rest of this article.[5]

Jenkin next asked what economic factors determined ρ. By the implicit function theorem, $D(p) = S(p-t)$ implies, assuming we begin at zero tax, that

$$D'(p)\rho = (\rho - 1)S'(p-t) \Rightarrow (S' - D')\rho = S'$$
$$\Rightarrow \rho = \frac{S'}{S' - D'} = \frac{1}{1 + (\epsilon_D/\epsilon_S)},$$

[4] One interpretation of this is that every additional infinitesimal unit of the tax is imposed by a different authority that does not internalize the negative externalities on other tax authorities, and in this sense, the total revenue the authority raises is precisely Qdt.

[5] We also drop references to a tax being infinitesimal where it does not create ambiguity and use the term "local" to refer to quantities pertaining to infinitesimal changes.

where $\epsilon_D \equiv -(D'p/Q)$ is the *elasticity of demand* and $\epsilon_S \equiv S'p/Q$ is the *elasticity of supply*. Thus, as the classic result goes, the inelastic side of the market bears the burden of taxation.

PRINCIPLE OF INCIDENCE (Perfect competition) 4. The pass-through rate is determined by the relative elasticity of supply and demand, $\rho = 1/[1 + (\epsilon_D/\epsilon_S)]$. The pass-through increases in the ratio of the elasticity of supply relative to that of demand.

Not discussed in Jenkin's analysis, or in the following literature as far as we have been able to determine, is how to integrate up these local changes into finite or global changes in taxes. Suppose that a tax increases by a finite amount from t_0 to $t_1 > t_0$. With $Q(t)$ the equilibrium quantity in the market as a function of the tax and $\rho(t)$ the pass-through rate as a function of t, the local analysis implies that

$$\Delta CS_{t_0}^{t_1} = -\int_{t_0}^{t_1} \rho(t) Q(t)\, dt$$

and

$$\Delta PS_{t_0}^{t_1} = -\int_{t_0}^{t_1} [1 - \rho(t)] Q(t)\, dt.$$

Now define the *quantity-weighted average pass-through rate between* t_0 *and* t_1 to be

$$\bar{\rho}_{t_0}^{t_1} \equiv \frac{\int_{t_0}^{t_1} \rho(t) Q(t)\, dt}{\int_{t_0}^{t_1} Q(t)\, dt}.$$

Then defining the *incidence between* t_0 *and* t_1 to be $I_{t_0}^{t_1} \equiv \Delta CS_{t_0}^{t_1} / \Delta PS_{t_0}^{t_1}$, we have

$$\bar{I}_{t_0}^{t_1} = \frac{\Delta CS_{t_0}^{t_1}}{\Delta PS_{t_0}^{t_1}} = \frac{-\int_{t_0}^{t_1} \rho(t) Q(t)\, dt}{-\int_{t_0}^{t_1} [1 - \rho(t)] Q(t)\, dt}$$

$$= \frac{\bar{\rho}_{t_0}^{t_1}}{1 - \bar{\rho}_{t_0}^{t_1}} \frac{\int_{t_0}^{t_1} Q(t)\, dt}{\int_{t_0}^{t_1} Q(t)\, dt} = \frac{\bar{\rho}_{t_0}^{t_1}}{1 - \bar{\rho}_{t_0}^{t_1}}.$$

Thus the formula for the incidence of a finite tax change is the same as that for a local tax change in which the pass-through rate is replaced by the quantity-weighted average pass-through rate over the range of the finite change. One noninfinitesimal change of common interest is that of raising the tax so high as to eliminate the market. Let $\bar{t}$ be the smallest tax, possibly infinite, at which $Q(\bar{t})$ is zero. We can call the *average quantity-weighted pass-through rate* $\bar{\rho} \equiv \bar{\rho}_0^{\bar{t}}$. Then the global incidence in the market $\bar{I} = CS/PS = \bar{\rho}/(1 - \bar{\rho})$.

PRINCIPLE OF INCIDENCE (Perfect competition) 5. The incidence of a finite tax change is the same as that of a local change in which the local pass-through rate is replaced by the quantity-weighted average pass-through rate over the range of the change. The global incidence of the market, the ratio of consumer to producer surplus, is the quantity-weighted average pass-through rate between zero tax and the smallest (perhaps infinite) tax that chokes off all trade.

This implies that the same factors that affect local incidence, relative elasticities of supply and demand, determine the global division of surplus. If demand is more elastic than supply globally, the surplus of the market's existence will accrue primarily to suppliers and conversely mutatis mutandis. Intuitively, to the extent that pass-through does not vary much as tax rates change, taxing a market hurts most that side of the market that benefits most from its existence. The fact that one side bears little of the cost of taxation indicates that it gains little from the market existing.

III. Monopoly

Analogously to the perfect competition case, we assume that the monopolist's profit function is concave in quantity and that her cost $c(q)$ and demand function, represented below by inverse demand $p(q)$, are smooth. The monopolist's revenues are $p(q)q$ with corresponding marginal revenue $mr(q) = p(q) + p'(q)q$ and her marginal cost $mc(q) = c'(q)$. Note that a (per-unit) tax on consumers does not affect p' or q and simply reduces p by the amount of the tax, while a tax on producers raises marginal cost by t uniformly. The monopolist chooses quantity to equate $mr(q) = mc(q)$. Thus, by the exact same argument as in the competitive case, the first principle of incidence extends to monopoly. To our knowledge this result is due to Jeremy Bulow and, while straightforward, has not appeared in print.

PRINCIPLE OF INCIDENCE (Monopoly) 1. Physical incidence of taxes is neutral in the sense that a tax levied on consumers, or a unit parallel downward shift in consumer inverse demand, causes nominal prices to consumers to fall by $1 - \rho$.

As far as we know, the logic of local incidence, principles 2 and 3, under monopoly is not discussed in previous literature. Because under monopoly consumer behavior obeys the same demand curve as under perfect competition, it continues to be the case that $dCS/dt = -\rho q$. However, a monopolist, unlike a price taker, chooses the market price to maximize producer surplus. Thus when, by the envelope theorem, we take the quantity sold as given in computing the impact of a change in taxes on profits under monopoly, we also hold the price fixed. Producer surplus, the monopolist's profits, is $[p(q) - t]q - c(q)$, and thus, by this logic, $dPS/dt = -q$.

We thus obtain two results, one qualitative and the other quantitative. Qualitatively, the burden of a tax is no longer simply shared between consumers and producers. Instead, the total burden on consumers and producers is greater than the revenue raised because the quantity sold is already distorted downward. While this qualitative point is likely widely understood, we do not know an explicit statement of it as such in the literature, and it is far from universally applied.

PRINCIPLE OF INCIDENCE (Monopoly) 2. The total burden of the tax is more than fully shared by consumers and producers. While the monopolist fully pays the tax out of her welfare, consumers also bear an excess burden.

Quantitatively, the size of the excess burden of the tax, per unit of revenue raised, is ρ and is borne by the consumers. Another way of expressing the relationship of pass-through to excess burden, which is useful in quantifying changes in social rather than just consumer welfare, is to consider an alternative shock to the market. For this we specialize to the case in which the firm has constant marginal cost c; similar but somewhat more complicated results relaxing this assumption are developed in Appendix A. Consider the exogenous entrance into the market of a quantity of the good, $\tilde{q}$, assumed to be produced at cost c. If we allow q to continue to denote the total quantity sold in the market, the firm's profits are now $[p(q) - c](q - \tilde{q})$ and its marginal revenue is now $p(q) + p'(q)(q - \tilde{q})$. Note that increasing $\tilde{q}$ has the same effect on the firm's incentives as changing costs by $p'(q)$:

$$\frac{dq}{d\tilde{q}} = p'(q)\frac{dq}{dt} = p'(q)\frac{\rho}{p'(q)} = \rho.$$

Thus the effect of increasing $\tilde{q}$ on the equilibrium quantity supplied in the market is ρ.

Let us define the markup function $m(q)$ for any q as $m(q) \equiv p(q) - c$.[6] Deadweight loss from monopoly may then be written as $\text{DWL}(q) = \int_q^{q^{**}} m(q_1)\,dq_1$, where q is the current (monopoly) quantity and q^{**} is the socially optimal quantity such that $m(q^{**}) = 0$. Thus $d\text{DWL}(q)/dq =$

[6] Even though we often write m without its argument, the reader should not interpret our notation for marginal cost (mc), marginal revenue (mr), and marginal surplus (ms) as markup m times another quantity.

$-m(q)$, and the change in deadweight loss that occurs as a result of an increase in $\tilde{q}$ is $d\mathrm{DWL}/d\tilde{q} = -\rho m$. On the other hand, the reduction in producer surplus from an increase in $\tilde{q}$ can again by computed by the envelope theorem by holding the optimally chosen q fixed: $dPS/d\tilde{q} = -m$. Thus we have that the *social incidence of competition*

$$SI \equiv \frac{d\mathrm{DWL}}{d\tilde{q}} \Big/ \frac{dPS}{d\tilde{q}} = \rho.$$

The motivation for considering the social incidence is analogous to that for considering the incidence of a tax. With constant marginal cost, the only source of firm profits is market power, and, of course, the only source of deadweight loss is also market power. Competition, through the perfectly competitive entry of $\tilde{q}$ goods into the market, undermines market power in the same way that a tax undermines production. The social incidence measures the relative rate at which competition (a "tax" on market power) erodes its two products: deadweight loss and profits.

Principle of incidence (Monopoly) 3. The incidence of a tax is $I = \rho$. If costs are linear (constant marginal), then the social incidence of competition $SI \equiv (d\mathrm{DWL}/d\tilde{q})/(dPS/d\tilde{q})$ is also equal to ρ.

The analysis of pass-through under monopoly is originally due to Cournot (1838), but our exposition here is more similar to the Marshallian treatment of Bulow and Pfleiderer (1983).[7] Monopoly optimization is given by $mr(q) = mc(q) + t$, and thus

$$mr'\frac{dq}{dt} = mc'\frac{dq}{dt} + 1 \Rightarrow \frac{dq}{dt} = \frac{1}{mr' - mc'}$$
$$\Rightarrow \rho = \frac{dp}{dt} = p'\frac{dq}{dt} = \frac{p'}{mr' - mc'}.$$

Marginal revenue, $mr = p + p'q$, has two terms: the price p and the negative of the marginal consumer surplus $ms \equiv -p'q$ that consumers earn when quantity expands. We can thus write

$$\rho = \frac{1}{\frac{p' - ms'}{p'} - \frac{mc'}{p'}} = \frac{1}{1 - \frac{ms'q \cdot p}{q \cdot ms \cdot p'}\frac{q \cdot ms}{q \cdot p} - \frac{mc'q \cdot p}{p'q \cdot mc}\frac{q \cdot mc}{q \cdot p}}$$
$$= \frac{1}{1 + \frac{\epsilon_D}{\epsilon_{ms}}\frac{ms}{p} + \frac{\epsilon_D}{\epsilon_S}\frac{mc}{p}},$$

[7] However, our discussion and interpretation are largely our own.

where ϵ_S is the elasticity of the inverse marginal cost curve ("the supply function") and $\epsilon_{ms} = ms/ms'q$ is the elasticity of the inverse marginal surplus function. The expression for pass-through may be further simplified using

$$\frac{ms}{p} = -\frac{p'q}{p} = \frac{1}{\epsilon_D}$$

and Lerner's (1934) rule

$$\frac{p - mc}{p} = \frac{1}{\epsilon_D} \Rightarrow \frac{mc}{p} = \frac{\epsilon_D - 1}{\epsilon_D}.$$

This yields

$$\rho = \frac{1}{1 + \frac{\epsilon_D - 1}{\epsilon_S} + \frac{1}{\epsilon_{ms}}}.$$

Two changes from the competitive formula appear. First, ϵ_D has been replaced by $\epsilon_D - 1$. Under monopoly the elasticity of demand is never below unity, and thus, intuitively, the appropriate elasticity of demand is relative to unity rather than zero. However, this neither qualitatively changes any effects (as $\epsilon_D > 1$ always at the optimum) nor introduces any new determinants of pass-through.

Thus the second change, the new inverse elasticity of marginal surplus term, is conceptually more significant. As we discuss more extensively through functional form examples in Fabinger and Weyl (2012), ϵ_{ms} measures the curvature of the logarithm of demand because

$$(\log D)' = \frac{D'}{D} = \frac{1}{p'q} = -\frac{1}{ms},$$

so

$$(\log D)'' = \frac{ms'}{ms^2}\frac{1}{p'} = -\frac{1}{\epsilon_{ms}}\frac{1}{ms}\left(-\frac{1}{p'q}\right) = -\frac{1}{\epsilon_{ms}}\frac{1}{ms^2}.$$

Therefore, log-concave demand always has $1/\epsilon_{ms} > 0$ and log-convex $1/\epsilon_{ms} < 0$. This connection between ϵ_{ms} and demand curvature was emphasized particularly by Seade (1985).[8] He and Bulow and Pfleiderer (1983) noted that pass-through exceeds unity under linear cost if and only if ϵ_{ms} is negative. This characterization extends to many symmetric imperfectly competitive models as we show in the next section. As Bagnoli and Bergstrom (2005) argue and we discuss extensively in Section VI, this leads many qualitative comparative statics in imperfectly competitive models to turn on whether demand is log concave or log convex. Another relevant threshold is that if demand is concave, then $1/\epsilon_{ms} > 1$, whereas if it is convex, $1/\epsilon_{ms} < 1$.[9]

A statistical way of viewing $1/\epsilon_{ms}$, proposed by Gabaix et al. (2013), is to notice that if α is the Pareto tail index for the demand, viewed as a probability distribution of consumer values, then $\epsilon_{ms} = -\alpha$.[10] Therefore, for the generalized Pareto/constant pass-through class of demand functions proposed by Bulow and Pfleiderer, which include linear, exponential, and constant elasticity as special cases, $\epsilon_{ms} = 1$ for linear, $1/\epsilon_{ms} \to 0$ for exponential, and $\epsilon_{ms} = -\epsilon$ for constant elasticity ϵ. A final way to think of $1/\epsilon_{ms}$ is in relationship to risk aversion: $1 - (1/\epsilon_{ms})$ is the Arrow-Pratt measure of relative risk aversion of the inverse demand function p if this function were to be viewed as a utility function.

Economists have typically seen ϵ_{ms}, compared to ϵ_D and ϵ_S, as variously difficult to estimate empirically and form intuitions about (see, e.g., Farrell and Shapiro 2010a). This attitude strikes us as overly pessimistic. Suppose, for example, that consumer willingness to pay were proportional to income. Then $1/\epsilon_{ms}$ corresponds to the well-known curvature properties of income distributions in the segment of the population representing the marginal consumer of the product. Such properties were used by Saez (2001) to calibrate models of optimal income taxation. In particular, $\alpha \in [1.5, 3] \Rightarrow \epsilon_{ms} \in [-3, -1.5]$ appears to fit well in the upper tail for most countries and a much less convex distribution (lognormal) appears to fit lower and middle-range incomes. Of course, this income example is very specific, though commonly used: consumers' willingness to pay for most products is not simply proportional to income. In Fabinger and Weyl (2012), we discuss more extensively how to calibrate $1/\epsilon_{ms}$ in various settings.

[8] He labels $E \equiv 1 - (1/\epsilon_{ms})$.

[9] The inverse elasticity of marginal surplus

$$\frac{1}{\epsilon_{ms}} = \frac{ms'q}{ms} = \frac{(p''q + p')q}{p'q} = 1 + \frac{p''q}{p'}.$$

Given that $q > 0 > p'$, if $p'' < (>)\ 0$, then this second term is $> (<)\ 0$ and $1/\epsilon_{ms} > (<)\ 1$.

[10] Gabaix et al. (2013) instead use $\gamma = 1/\alpha$ as their tail index. In the discussion that follows we use the Pareto tail index as it is more common in the economics literature, especially in public finance.

PRINCIPLE OF INCIDENCE (Monopoly) 4.

$$\rho = \frac{1}{1 + \frac{\epsilon_D - 1}{\epsilon_S} + \frac{1}{\epsilon_{ms}}}.$$

The general features of pass-through under competition carry over, but under monopoly it is also the case that the more positively (log-) curved demand is, the higher pass-through is.

Precisely the same logic as under perfect competition allows the extension of local to finite and global incidence, using the appropriate monopolistic incidence formulas. The logic may also be extended to calculate finite changes in deadweight loss using social incidence. Because the arguments are analogous, we do not repeat them here but simply note that the relevant average pass-through rate is now the *markup-weighted average pass-through* taken over values of $\tilde{q}$:

$$\tilde{\rho}_{\tilde{q}_0}^{\tilde{q}_1} \equiv \frac{\int_{\tilde{q}_0}^{\tilde{q}_1} \rho(\tilde{q}) m(\tilde{q}) d\tilde{q}}{\int_{\tilde{q}_0}^{\tilde{q}_1} m(\tilde{q}) d\tilde{q}},$$

and the *markup-weighted average pass-through* relevant for *global social incidence* $\widetilde{SI} \equiv \mathrm{DWL}/PS$ is

$$\tilde{\rho} \equiv \tilde{\rho}_0^{q^{**}}.$$

The validity of the global formula follows from noting that when $\tilde{q} = q^{**}$, there is neither any deadweight loss nor any producer surplus.

PRINCIPLE OF INCIDENCE (Monopoly) 5. Incidence of finite or global tax changes follows the incidence formula in which pass-through is replaced by quantity-weighted average pass-through over the appropriate range. If the firm has constant marginal cost, finite social incidence of competition similarly follows the local incidence formula in which pass-through is replaced by markup-weighted average pass-through over the range of competition, and for the global incidence formula, the range of competition is that between no competition and the socially optimal quantity being produced in competition.

Thus the pass-through rate, ρ, differently averaged, determines both the ratio of consumer to producer surplus and the ratio of deadweight loss to producer surplus. This is intuitive because a monopolist bears the full cost of any tax she faces. Therefore, any burden borne by consumers is exactly the excess burden. The pass-through therefore measures the local incidence to consumers, rather than to producers, and to society compared to producers. By the same arguments as under competition, global incidence is simply an averaging of local incidence.

IV. Symmetric Imperfect Competition

We now consider incidence under symmetric, imperfect competition. There are n firms in the industry, distinguished by index i, each producing a single product. These goods may be distinct from the consumers' point of view. However, the demand system is assumed to be fully symmetric. Our notation corresponds to finite n, but the discussion applies, mutatis mutandis, also to the case of a continuum of firms, which we assume to be of measure one.

For any firm i the cost associated with producing quantity q_i is given by the same cost function $c(q_i)$. The marginal cost is denoted $mc(q_i) \equiv c'(q_i)$. The market-clearing price at which firm i sells its product depends in general on the quantity produced and sold by each of the firms. If we let all quantities q_j with $j \in \{1, 2, \ldots, n\}$ be equal to the same number q, the corresponding price will be denoted $p(q)$. Note that $p(q)$ is a function of just one scalar argument and is independent of i. Its derivative $p'(q)$ captures the response of the price of any of the goods to a simultaneous infinitesimal symmetric increase of all quantities.

We define the elasticity of market demand as $\epsilon_D = -(p/qp')$, not to be confused with the elasticity of the residual demand that any of the firms faces. When, as above, we allow for exogenous competition $\tilde{q}$,[11] ϵ_D denotes the elasticity of demand for goods by the *welfare-relevant firms* $-p/[(q - \tilde{q})p']$, not the elasticity of the total demand for the good. In all cases we focus on a unique symmetric equilibrium, either because it is the only equilibrium or because we are interested in changes local to this focal equilibrium.

Rather than specify a particular model of firm interactions, which are not crucial for most of the questions of incidence, we instead follow Genesove and Mullin's (1998) variation on Bresnahan (1989) in postulating that the elasticity-adjusted Lerner index $[(p - mc - t)/p]\epsilon_D$ is set equal to a *conduct parameter* θ. Where it is not explicitly relevant, we do not explicitly write $-t$ and instead take it as an implicit part of mc.

In some cases θ may depend on the total level of production q (not just the $q - \tilde{q}$ produced by the firms under welfare consideration) and thus implicitly on the interventions t and $\tilde{q}$. However, we are not aware of any standard models of imperfect competition in which θ as defined above

[11] In this multifirm case, exogenous competition $\tilde{q}$ means that a quantity $\tilde{q}$ of each of the goods produced by the n firms is exogenously supplied to the market.

is directly affected by the interventions (taxes or exogenous competition) we consider.[12]

This model nests a surprisingly wide range of forms of imperfect competition, far broader than those considered by Bresnahan or Genesove and Mullin. In fact, we are not aware of any commonly used complete information symmetric models it does not include. To illustrate this, we briefly describe how three very different and broad models fall within this framework: quantity choices with symmetric conjectural variations in the spirit of Bowley (1924) with Cournot's model of competition as a special case, symmetrically differentiated Nash-in-prices competition or collaboration (complementary monopoly) including as a special case Cournot's model of perfect complements, and partial equilibrium monopolistic competition with the quasi-linear version of the Dixit and Stiglitz (1977) model as a special case.

1. *Homogeneous products oligopoly.*—This is the model in the spirit of Bresnahan, and our analysis here is taken from him. When firm i chooses its quantity q_i, it assumes that an infinitesimal change dq_i would make each other firm change its quantity by $[R/(n-1)]dq_i$ in response. In this framework, Cournot competition corresponds to $R = 0$ and homogeneous product Bertrand competition to $R = -1$. Let us denote the total industry quantity by $Q \equiv \sum_{i=1}^{n} q_i$. The price at which firm i sells its product is independent of i and is given by the market inverse demand function $P(Q) \equiv p(Q/n)$. The profit of firm i is $P(Q)q_i - c(q_i)$. The associated first-order condition evaluated at symmetric quantities $q_1 = q_2 = \cdots = q_n = Q/n \equiv q$ may be written as

$$(1+R)qP'(nq) + P(nq) - mc(q) = 0$$
$$\Rightarrow \frac{p - mc}{p}\epsilon_D = \frac{1+R}{n} = \theta,$$

since

$$\epsilon_D \equiv -\frac{p}{qp'} = -\frac{P}{qnP'}.$$

If R is a constant, as in the Cournot, Bertrand, or Delipalla and Keen (1992) conjectures models, then θ is also constant.

2. *Symmetrically differentiated Nash-in-prices.*—The quantity $q_i(p_i, \mathbf{p}_{-i})$ each symmetric firm sells depends on its own price p_i and on the prices of the other $n-1$ firms. At symmetric prices p, $q(p) \equiv q_i(p, p, \ldots, p)$ for any i is the inverse of $p(q)$. By Lerner's rule applied to the residual demand of each firm, at a symmetric equilibrium each chooses

[12] A potential exception is models of (implicit) collusion in which the fall in profitability from an increase in taxes might directly affect θ.

$$\frac{p - mc}{p} = -\frac{q}{p} \Big/ \frac{\partial q_i}{\partial p_i}.$$

The elasticity of market demand is given by

$$\epsilon_D = -\frac{p}{q} \sum_j \frac{\partial q_i}{\partial p_j}.$$

Thus the equilibrium condition is

$$\begin{aligned} \frac{p - mc}{p} \epsilon_D &= \frac{\sum_j \partial q_i / \partial p_j}{\partial q_i / \partial p_i} = 1 + \frac{\sum_{j \neq i} \partial q_i / \partial p_j}{\partial q_i / \partial p_i} \\ &= 1 + \frac{\sum_{j \neq i} \partial q_j / \partial p_i}{\partial q_i / \partial p_i} \equiv 1 - A = \theta, \end{aligned} \tag{1}$$

where the third equality follows by symmetry and

$$A \equiv -\sum_{j \neq i} \frac{\partial q_j}{\partial p_i} \Big/ \frac{\partial q_i}{\partial p_i}$$

is the *aggregate diversion ratio* (Shapiro 1996) from any individual firm to the rest of the industry. Intuitively, A may be thought of as the fraction of sales lost by one firm when it increases its price that are captured by other firms in the industry. This ratio is always positive, and thus $\theta < 1$ if goods are substitutes but is actually negative if goods are complements; thus in this case $\theta > 1$. In Cournot's collaboration model, where goods are perfect complements, $\partial q_j / \partial p_i = \partial q_i / \partial p_i$ for any i and j as the quantity of each good sold depends only on the sum of the prices. Therefore, $A = -(n - 1)$ and $\theta = n$. In this special case, again, θ is independent of q. This is true more generally if the differentiated products demand system is linear as in this case A is constant and equal to the ratio of two linear coefficients.

3. *Monopolistic competition.*—There is a continuum of measure one of symmetric firms selling different varieties of a product.[13] We seek to model monopolistic competition in a single industry, as in Atkeson and Burstein (2008), rather than a whole economy. Thus we assume that consumers have quasi-linear utility, $U(\int_i u(q_i) di) - \int_i p_i q_i di$, where u is strictly concave and both u and U are smooth.

[13] This assumption of a continuum rules out the "indirect" effects emphasized by Atkinson and Stiglitz (1980) in their analysis of incidence under monopolistic competition.

If prices are symmetric, utility maximization given concavity of u requires that quantities q are symmetric across goods prices $p = U'(u(q))u'(q)$. We assume that $U'(u(q))u'(q)$ strictly declines in q (aggregate inverse demand for the goods is downward sloping) and thus associate $U'(u(q))u'(q)$ with $p(q)$.

The inverse demand function facing each firm i, when the quantity chosen by all other firms is q, is $U'\left(\int_j u(q)dj\right)u'(q_i)$, which we denote simply by $U'u'(q_i)$. The optimal quantity is thus given by $p + U'u''q = mc(q)$ or $(p - mc)/p = -U'u''q/p$. On the other hand,

$$\epsilon_D = -\frac{p}{qp'} = -\frac{p}{q}\frac{1}{U''(u')^2 + U'u''}.$$

Thus

$$\theta = \frac{p - mc}{p}\epsilon_D = \frac{U'u''}{U''(u')^2 + U'u''}.$$

If U is concave, the goods are substitutes and $\theta < 1$; if U is convex, the goods are complements and $\theta > 1$. In the quasi-linear version of the Dixit-Stiglitz model, $u(q) = q^\beta$, where $\beta \in (0, 1)$ and U is flexible as it must be renormalized for quasi-linearity. A natural specification, proposed by Atkeson and Burstein with a finite number of firms per industry, is that $U(x) = x^\eta$ with $\eta > 0$, which we now assume.[14] In particular, calculations show that $\theta = (1 - \beta)/(1 - \beta\eta)$;[15] if $\eta \geq 1/\beta$, then industry inverse demand is weakly upward sloping, so we rule out this case even though $\eta = 1/\beta$ is perhaps the most immediate quasi-linear version of the Dixit-

[14] It is a little-remarked fact that in the (quasi-linear) Dixit-Stiglitz model, firms' goods are complements as long as $\eta > 1$, and thus it is Pareto improving for the industry to be monopolized rather than to be monopolistically competitive. In the Atkeson and Burstein model, η in our notation is $(1 - \eta)/(1 - \rho)$ in their notation; because they have a continuum of industries, the quasi-linearity assumption holds exactly despite the fact that their model in principle involves income effects.

[15] The calculations are as follows:

$$\frac{U'u''}{U''(u')^2 + U'u''} = \frac{\dfrac{u''}{(u')^2}}{\dfrac{U''}{U'} + \dfrac{u''}{(u')^2}} = \frac{\dfrac{\beta(\beta-1)q^{\beta-2}}{\beta^2 q^{2\beta-2}}}{\dfrac{\beta(\beta-1)q^{\beta-2}}{\beta^2 q^{2\beta-2}} + \dfrac{\eta(\eta-1)u^{\eta-2}}{\eta u^{\eta-1}}}$$
$$= \frac{\dfrac{\beta-1}{\beta q^\beta}}{\dfrac{\beta-1}{\beta q^\beta} + \dfrac{\eta-1}{u}} = \frac{1}{1 - \dfrac{\beta(\eta-1)}{1-\beta}} = \frac{1-\beta}{1-\beta\eta}.$$

Stiglitz model. Clearly here θ is a constant. When $\eta > (<)\ 1$ so that goods are complements (substitutes), $\theta > (<)\ 1$.

We now proceed with our analysis using this general model. Assuming no direct dependence of θ on the physical incidence of taxes, the same argument as before (applied to the new equilibrium conditions) extends the independence of economic and physical incidence.

Principle of incidence (Symmetric imperfect competition) 1. In the same sense as in our discussion of monopoly and perfect competition, economic incidence under symmetric imperfect competition is independent of physical incidence.

Again, we can apply the envelope theorem to consumers to state that $dCS/dt = -\rho Q$. No simple envelope theorem applies to firms, however, as they are neither price takers nor joint profit-maximizing price setters. We must therefore compute the incidence on firms. Symmetric-across-firms profits per firm are $[p(q) - t]q - c(q)$. We have $\rho = dp/dt = p'(dq/dt)$, so $dq/dt = \rho/\rho'$. Thus the impact on per-firm producer surplus ps is

$$\begin{aligned}\frac{dps}{dt} &= (\rho - 1)q + \frac{\rho}{p'}(p - t) - mc\frac{\rho}{p'}\\ &= \left(\rho - 1 - \rho\frac{p - t - mc}{p}\frac{p}{-p'q}\right)q\\ &= -[1 - \rho(1 - \theta)]q.\end{aligned}$$

Aggregating across firms, we conclude that $dPS/dt = -[1 - (1 - \theta)\rho]Q$. Note that this formula is a linear combination of the formulas under monopoly and perfect competition, with a weight θ on monopoly and $1 - \theta$ on the perfectly competitive case. This extension of the monopoly formula was first derived by Atkin and Donaldson (2012), extending the working paper version of this article, and inspired this result.

Principle of incidence (Symmetric imperfect competition) 2. If $\theta < (>)\ 1$, firms less than (more than) fully bear the cost of the tax.[16] As long as $\theta > 0$, the tax has an excess burden, as the burden on consumers more than fully completes the burden on producers.

We can also extend the social incidence, again under the assumption of constant marginal cost c, to imperfect competition using the notion

[16] Kimmel (1992) notes that incidence on producers may actually be negative: they may gain from a tax increase. To see this, plug in the expression for ρ from principle 4 into the denominator of the incidence formula of principle 3 to obtain

$$1 - \frac{1 - \theta}{1 + \dfrac{\theta}{\epsilon_\theta} + \dfrac{\epsilon_D - \theta}{\epsilon_S} + \dfrac{\theta}{\epsilon_{ms}}},$$

of exogenous competition. Suppose that a per-firm exogenous quantity $\tilde{q}$ enters the market. Following the logic of the monopoly section, firms now perceive industry marginal revenue $p + (q - \tilde{q})p'$ and equate $p + \theta(q - \tilde{q})p' = c$. Thus an increase in $\tilde{q}$ is equivalent to a reduction in cost/tax of $-\theta p'$, so $dq/d\tilde{q} = \theta\rho$. Defining the markup function as $m(q) \equiv p(q) - c$, we have, in analogy with the monopoly case,

$$\frac{d\mathrm{DWL}}{dq} = -[p(q) - c] = -m(q).$$

As with the effect of tax on profits, we cannot apply any simple envelope theorem under imperfect competition. Profits per firm are $(q - \tilde{q})[p(q) - c]$, so

$$\begin{aligned}\frac{dps}{d\tilde{q}} &= \theta\rho(q - \tilde{q})p' + m(q)(\theta\rho - 1)\\ &= m(q)\left[\theta\rho\frac{(q - \tilde{q})p'}{p(q)}\frac{p(q)}{p(q) - c} + \theta\rho - 1\right]\\ &= -m(q)[1 + (1 - \theta)\rho].\end{aligned}$$

Thus, while $I = \rho/[1 - (1 - \theta)\rho]$, $SI = \theta\rho/[1 + (1 - \theta)\rho]$. Two things are worth noting.

1. With ρ held fixed, $\partial I/\partial\theta < 0$, as long as $I > 0$. That is, the less competitive conduct is, the more of taxation is borne by firms relative to consumers. Especially in its global interpretation as below (firms capture a greater share of surplus the less competitive conduct is), this is intuitively obvious. Perhaps slightly less obvious is that this continues to hold (tax burden continues to fall more on firms) as θ increases past one. While this result holds fixed the pass-through rate, which, as we will see

which has the same sign as

$$\frac{\epsilon_D - \theta}{\epsilon_S} + \theta\left(1 + \frac{1}{\epsilon_\theta} + \frac{1}{\epsilon_{ms}}\right).$$

Kimmel argues that with constant marginal cost ($\epsilon_S = \infty$) and Cournot conduct, the industry gains from a tax if and only if $1/\epsilon_{ms} < -1$, which follows from this formula. This cannot be the case globally as it implies infinite consumer surplus, but in Fabinger and Weyl (2012), we discuss some empirical cases in which it occurs locally. Negative values of ϵ_θ (as we show are common under differentiated products Nash-in-prices competition) will make this result more likely, while decreasing returns to scale make it less likely, especially for small θ (which is required for $(1/\epsilon_\theta) + (1/\epsilon_{ms}) < -1$ as stability requires $\theta[(1/\epsilon_\theta) + (1/\epsilon_{ms})] > -1$). We thus consider this result unlikely to be empirically relevant in many symmetric industries and defer our in-text discussion of potentially profit-enhancing taxes to the case asymmetries. We thank Joe Farrell for bringing this issue to our attention.

below, is partly determined by θ, the same comparative statics can be shown to hold as long as θ is independent of q, holding fixed more primitive quantities such as elasticities or the underlying demand and supply curves.

2. Again with ρ held fixed, $\partial SI/\partial\theta > 0$, again as long as $SI > 0$. That is, the less competitive conduct is, the more "exogenous competition" $\tilde{q}$ reduces deadweight loss compared to its impact on profits. Again, the global version is perhaps more intuitive, though likely less obvious than that for I: the more competitive an industry is, the greater the ratio of profits to the deadweight loss of market power (even with constant marginal cost and thus zero competitive profits). When $\theta > (1+\rho)/\rho > 1$, as may be the case with complementary goods, exogenous competition actually benefits producers (as well as society). Furthermore, as $\theta \to 0$, the ratio of deadweight loss to profits goes to zero. Again, these results can be shown taking more "primitive" quantities than the pass-through rate as fixed, but this analysis is omitted here.

Principle of incidence (Symmetric imperfect competition) 3. The incidence of a tax is $I = \rho/[1-(1-\theta)\rho]$, while under constant marginal cost the social incidence of competition is $SI = \theta\rho/[1+(1-\theta)\rho]$. With pass-through held fixed, incidence of taxation falls more on firms the greater θ is, and social incidence of competition falls more heavily on deadweight loss than on firms the greater θ is.

Under imperfect competition, quantity is chosen according to $p(q) - \theta ms(q) = mc(q)$, where ms is the marginal surplus per firm. Following the logic of the monopoly case,

$$\rho = \frac{1}{1 - \dfrac{d\theta}{dq}\dfrac{ms}{p'} - \theta\dfrac{ms'q\cdot p}{q\cdot ms\cdot p'}\dfrac{q\cdot ms}{q\cdot p} - \dfrac{mc'q\cdot p}{p'q\cdot mc}\dfrac{q\cdot mc}{q\cdot p}}$$
$$= \frac{1}{1 + \dfrac{d\theta}{dq}q + \theta\dfrac{\epsilon_D}{\epsilon_{ms}}\dfrac{ms}{p} + \dfrac{\epsilon_D}{\epsilon_S}\dfrac{mc}{p}}.$$

Now $(p-mc)/p = \theta/\epsilon_D$, so $mc/p = (\epsilon_D - \theta)/\epsilon_D$. Defining ϵ_θ as $\theta/[q(d\theta/dq)]$, we can now write

$$\rho = \frac{1}{1 + \dfrac{\theta}{\epsilon_\theta} + \dfrac{\epsilon_D - \theta}{\epsilon_S} + \dfrac{\theta}{\epsilon_{ms}}}. \tag{2}$$

This formula nests and generalizes both the homogeneous products conjectural analysis of Delipalla and Keen (1992) and the differentiated products Nash-in-prices analysis of Anderson, de Palma, and Keider (2001), illustrating why they reach essentially the same conclusion. The exception is one term, θ/ϵ_θ, to which we return shortly. Otherwise in both cases the denominator of this expression is a linear combination of that of monopoly and perfect competition, with the weight on monopoly being θ and on perfect competition $1-\theta$. When θ is invariant to changes in q, as was the case for many models discussed above (of both Cournot's models and the Dixit-Stiglitz model, e.g.), this additional term is absent because $1/\epsilon_\theta = 0$. In these cases, the qualitative results from the monopoly section continue to hold, and the relative importance of the distinctively monopolistic factor (the elasticity of marginal surplus) compared to the competitive factors (the relative elasticities of supply and demand) is determined by θ. Again if costs are linear, the comparison of ρ to unity is determined by the sign of ϵ_{ms}, regardless of the magnitude of θ.

Perhaps more interesting is the case in which θ depends on q. In this case if θ rises with q $(\epsilon_\theta > 0)$, pass-through is smaller than indicated by the linear combination as higher prices create more competitive conduct, thereby offsetting the impetus for a price increase. If θ falls with q $(\epsilon_\theta < 0)$, pass-through is greater than indicated by the linear combination as higher prices create less competitive conduct, thereby exacerbating the initial impetus for the price increase. We now consider an example of the second case: discrete choice-based models of differentiated products Nash-in-prices competition.

As analyzed above in the context of equation (1), under differentiated products Nash-in-prices competition, $\theta = 1 - A$, where A is the aggregate diversion ratio. A standard microfoundation of differentiated products demand is discrete choice (Anderson, de Palma, and Thisse 1992): every one of a continuous distribution of consumers may buy at most one unit of at most one good. The firms in the market compete to make this one sale to as many consumers as possible. To keep notation simple, we now focus on the case of duopoly, but our discussion can easily be extended to the case of any number of symmetric firms, as shown in Appendix B.

There are two firms, 1 and 2, each selling a single good. A given consumer i has utility u_1^i from consuming good 1 and utility u_2^i from consuming good 2; no consumer can consume more than one good. The values (u_1, u_2) are drawn from a distribution with a differentiable full-support density $f(u_1, u_2)$ that is symmetric in its arguments: $f(u_1, u_2) = f(u_2, u_1)$. At prices (p_1, p_2), individuals purchasing from firm 1 earn a weakly higher utility from this than from purchasing good 2 $(u_1 - p_1 \geq u_2 - p_2)$ or making no purchase at all $(u_1 \geq p_1)$. The measure of such customers may we written as

$$q_1(p_1, p_2) = \int_{p_1}^{\infty} \int_{-\infty}^{u_1 - p_1 + p_2} f(u_1, u_2)\, du_2\, du_1.$$

Therefore, at symmetric prices $p_1 = p_2 \equiv p$,[17]

$$A = \frac{\partial q_1/\partial p_2}{-(\partial q_1/\partial p_1)} = \frac{\int_{p_1}^{\infty} f(u_1, u_1 - p_1 + p_2)du_1}{\int_{-\infty}^{p_2} f(p_1, u_2)du_2 + \int_{p_1}^{\infty} f(u_1, u_1 - p_1 + p_2)du_1} = \frac{\int_{p}^{\infty} f(u_1, u_1)du_1}{\int_{-\infty}^{p} f(p, u_2)du_2 + \int_{p}^{\infty} f(u_1, u_1)du_1}. \tag{3}$$

The measure of "product switchers," $\int_p^{\infty} f(u_1, u_1)du_1$, is manifestly decreasing in p; as emphasized in Jaffe and Weyl (2010), when prices are higher, goods inside the industry are less attractive than those outside and thus exert less competitive pressure. The measure of "market exiters," $\int_{-\infty}^{p} f(p, u_2)du_2$, is necessarily increasing in p if $\int_{-\infty}^{p} f_1(p, u_2)du_2 \geq 0$ at this price p. This condition is certainly satisfied if demand is locally weakly concave in own price at $p_1 = p_2 = p$,[18] since

$$\frac{\partial^2}{\partial^2 p_1} q_1(p_1, p_2) = -\int_{-\infty}^{p_2} f_1(p_1, u_2)du_2.$$

In this case we conclude that $(1/A) - 1$ is increasing in p, and consequently, $\theta = 1 - A$ rises in response to an increase in the price p.

The conduct parameter θ increases with price under much more general conditions, as discussed in Appendix B for duopoly and for more general oligopoly. In the special case of valuations u_i independent and drawn from the same distribution with cumulative distribution function $G(u)$, we make the following observation: If the reversed distribution function (i.e., $\tilde{G}(\tilde{u}) \equiv 1 - G(-\tilde{u})$) for valuations generates a single-product monopoly demand with $1/\epsilon_{ms}$ decreasing in price for prices be-

[17] Note that all equalities in (3) will hold even for $f(u_1, u_2)$ that is not symmetric in its arguments (i.e., $f(u_1, u_2) \neq f(u_2, u_1)$ in general) provided that we replace A with the diversion ratio d_1^2 between good 1 and good 2 defined in Sec. V.

[18] A simple example is provided by a uniform distribution with rectangular support $[u_{\min}, u_{\max}] \times [u_{\min}, u_{\max}]$, with $p \in (u_{\min}, u_{\max})$. Although this distribution does not satisfy our differentiability and full-support assumptions, it may be obtained as a limit of a sequence of distributions that do satisfy them.

low some $-\underline{p}$,[19] then the duopoly or oligopoly $\theta = 1 - A$ is increasing in p for $p \in (\underline{p}, \infty)$. As discussed in Fabinger and Weyl (2012), these conditions are globally satisfied for the normal (Gaussian), logistic, and type I extreme value (Gumbel) distributions as well as for their reversals.

It is possible to construct examples in which θ decreases with p over some range of prices, but they are not as natural and thus, we suspect, are less likely to be of empirical relevance.[20]

Principle of incidence (Symmetric imperfect competition) 4.

$$\rho = \frac{1}{1 + \dfrac{\theta}{\epsilon_\theta} + \dfrac{\epsilon_D - \theta}{\epsilon_S} + \dfrac{\theta}{\epsilon_{ms}}}.$$

The new term, θ/ϵ_{ms}, leads to lower (higher) pass-through when higher prices/lower quantities lead the industry to be more (less) competitive.

Finally, to extend incidence globally, we must average not just pass-through but also the product of pass-through and θ. We define the relevant weighted averages $\overline{\theta\rho}_{t_0}^{t_1}$, $\widetilde{\theta\rho}_{\tilde{q}_0}^{\tilde{q}_1}$, $\overline{\theta\rho}$, and $\widetilde{\theta\rho}$ implicitly, analogously to the previous sections. Clearly in the cases in which θ is independent of q, it may be taken out of the average. For example,

$$\overline{\theta\rho} \equiv \frac{\displaystyle\int_0^\infty \theta(t)\rho(t)Q(t)\,dt}{\displaystyle\int_0^\infty Q(t)\,dt},$$

where $\theta(t)$, analogously to the above, is shorthand for $\theta(q(t))$. Using this definition and following the example, we get $\bar{I} = \bar{\rho}/(1 - \bar{\rho} + \overline{\theta\rho})$.[21]

Principle of incidence (Symmetric imperfect competition) 5. Incidence of finite and global tax as well as social incidence of competition changes under constant marginal cost may be obtained from the corresponding local formulas by replacing ρ and $\rho\theta$ by their weighted averages over the appropriate range of tax changes or competition changes. The weight is given by quantity in the case of incidence and by markup in the case of social incidence.

[19] If we let the auxiliary single-good monopoly problem have constant marginal cost, then $1/\epsilon_{ms}$ locally decreasing in price $\tilde{p}$ is equivalent to the pass-through rate locally increasing in price $\tilde{p}$, since $\rho(\tilde{p}) = \epsilon_{ms}/(\epsilon_{ms} + 1)$.

[20] It suffices to make the integral $\int_{-\infty}^{p} f_1(p, u_2)\,du_2$ sufficiently strongly negative in this range without making $f(p, p)$ large, which is achievable for a small enough range of prices p. In this case the ratio of the measure of product switchers and the measure of market exiters will rise in response to an increase in p, and so will A.

[21] Note that specializing this result to the case of constant marginal cost and Cournot competition provides a simple proof of Anderson and Renault's (2003) surplus bounds under Cournot competition and also of the univariate Prékopa (1971)–Borell (1975) theorem relating the log curvature of densities to those of survival functions.

V. General Model

We now consider the most general model of this article, allowing for asymmetric, imperfectly competitive firms. This generalizes all our previous results but requires significant new concepts and notation. Therefore, rather than extend all the principles to this more general context, we develop an analogy to our previous model and show how this can be used to derive the basic incidence formula. The other principles can be developed similarly, and a treatment is available on request. See Reny, Wilkie, and Williams (2012) for an elegant exposition of principle 4 in the homogeneous products with conjectures case and Appendix C for a more general discussion that highlights the analogies with our analysis in the symmetric case.

Each firm, i, which may be part of a finite or continuous set of firms, produces quantity q_i and earns profits $p_i(\mathbf{q})q_i - c_i(q_i)$, where $\mathbf{q}$ is the (possibly infinitely dimensional) vector of quantities produced by all firms. Both p_i and c_i are assumed to be smooth, and in the case in which there is a continuum of firms, p is assumed to be representable as an integral of a smooth function over the firms' quantities. In most of the development below we act as if there are a finite number of firms, making clear the continuous analogue only when necessary.[22] As above, we again focus on a single equilibrium.

As with the symmetric imperfectly competitive model, rather than directly modeling firm conduct, we specify a firm-specific conduct parameter. Let $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{m} \equiv \mathbf{p} - \mathbf{mc}$, respectively, denote the vector of firm prices, quantities, and markups, where $mc_i(q_i) \equiv c_i'(q_i)$. Additionally, assume that firms have a single-dimensional strategic variable σ_i that determines its actions; this may be price, quantity, or some supply function in the spirit of Klemperer and Meyer (1989). Each firm takes the other firms' strategies as fixed when changing its strategy. As we will see below, however, this does not rule out traditional conjectural variations models. We assume that all $\mathbf{q}$ and $\mathbf{p}$ are smooth functions of $\boldsymbol{\sigma}$, which is assumed to live on the real line, and always obey the demand system $\mathbf{q}(\boldsymbol{\sigma}) = \mathbf{q}(\mathbf{p}(\boldsymbol{\sigma}))$. We use d rather than ∂ notation in denoting derivatives of $\mathbf{q}$ and $\mathbf{p}$ with respect to σ_i to capture the fact that these may include conjectural variations in some interpretations.

We then let each firm's *conduct parameter*

$$\theta_i \equiv \frac{\mathbf{m} \cdot \dfrac{d\mathbf{q}}{d\sigma_i}}{-\mathbf{q} \cdot \dfrac{d\mathbf{p}}{d\sigma_i}} = \frac{\mathbf{m} \cdot \dfrac{d\mathbf{q}}{d\sigma_i}}{-\mathbf{q} \cdot \left(\dfrac{d\mathbf{q}}{d\sigma_i} \dfrac{\partial \mathbf{p}}{\partial \mathbf{q}} \right)}. \tag{4}$$

[22] With asymmetries, a countably infinite analogue is also possible, but we are not aware of any standard model that uses it and thus never explicitly employ it.

As before, when we introduce taxes $\mathbf{t}$,

$$\theta_i = \left[(\mathbf{m} - \mathbf{t}) \cdot \frac{\partial \mathbf{q}}{\partial \sigma_i}\right] \Big/ \left(-\mathbf{q} \cdot \frac{d\mathbf{p}}{d\sigma_i}\right).$$

We begin by discussing the definition of θ_i and then provide several examples analyzing the value of θ_i in specific models.

The numerator of θ_i is the set of all *real* or *nonpecuniary* effects of firm i's changing its strategy. This includes both the profits she earns by selling more units at her own markup and the real (nonpecuniary) externalities she exerts on other firms by altering their quantities. When only σ_i is taken as directly manipulable, it is socially optimal to set this numerator, and thus θ_i, to zero. Furthermore, note that in a perfectly competitive market, $\theta_i = 0$ for all firms as all firms' markups are equal to zero. In both of these senses, $\theta_i = 0$ corresponds to perfectly competitive conduct. However, note that it does not necessarily imply that firm i has zero markup. If other firms have positive markups and firm i's increasing quantity reduces the quantity of other firms, then firm i would have to have a positive markup to satisfy $\theta_i = 0$. For example, if all of firm i's marginal sales were taken from other firms, firm i would have to have the same markup as the diversion-weighted average for other firms.

On the other hand, the denominator of θ_i is the *pecuniary* effects that firm i's changing its strategy has through changing prices. These again include both the firm's own self-benefiting market power $(-q_i[dp_i/d\sigma_i])$ and the pecuniary externalities the firm creates for other firms. Such effects are irrelevant and should be ignored from a social perspective as, following the classic logic of pecuniary effects, every benefit brought by a price change is offset by a harm to individuals on the other side of the market. However, if firm i were seeking to maximize industry profits by choosing σ_i, she would set $\theta_i = 1$ as the total impact on industry profits of a change in σ_i is $\mathbf{m} \cdot (d\mathbf{q}/d\sigma_i) + \mathbf{q} \cdot (d\mathbf{p}/d\sigma_i)$. In this sense, $\theta_i = 1$ corresponds to monopolistic/perfectly collusive conduct and $\theta_i > 1$ implies overweighting pecuniary effects (relative to a monopolist's behavior) and thus acting even less competitively than a monopolist.

As a final general comment, note that if all firms are symmetric, as in the previous section, and $\theta_i = \theta$, then increasing all q_i symmetrically raises the symmetric, per-firm quantity by dq and the price by dp satisfying $mdq = \theta q dp$, where m, p, and q are all symmetric, per-firm magnitudes. Thus

$$\theta = -\frac{m}{q}\frac{dq}{dp} = -\frac{m}{p}\frac{p(dq/dp)}{q} = \frac{p - mc}{p}\epsilon_D,$$

our definition of θ from the previous section. Hence our generalized definition of θ_i collapses to our symmetric definition when firms are symmetric.

We now consider asymmetric versions of the examples of the previous section and fit them into this model.

1. *Homogeneous products oligopoly.*—At the market level, everything is the same as in the previous section, but now each firm may have a different cost function, set of conjectures, and therefore equilibrium quantity. An individual firm's profits are $P(Q)q_i - c_i(q_i)$. Firms' strategies are denoted in units of quantity, in keeping with Telser (1972), but we allow for conjectural variations; see Jaffe and Weyl (forthcoming) for a more detailed exposition. In particular, taking the derivative with respect to σ_i and assuming that firm i anticipates a reaction in amount r_i^j from firm j, we see that firm i's first-order condition is

$$P'(Q)\left(1 + \sum_{j \neq i} r_i^j\right) q_i + P(q_i) - mc_i(q_i) = 0$$
$$\Rightarrow \frac{m_i}{-\,q_i(1 + \sum_{j \neq i} r_i^j)P'} = 1. \tag{5}$$

First consider the case of Cournot in which $r_i^j = 0$ for all $j \neq i$. Then $d\mathbf{q}/d\sigma_i$ is zero except for an entry of 1 in position i, and thus the numerator of the left-hand side of the final equality (5) is that of θ_i in equation (4). On the other hand, $\mathbf{q} \cdot (d\mathbf{p}/d\sigma_i) = QP'$, and thus the denominator of the final expression in (5) is equal to firm i's share $s_i \equiv q_i/Q$ multiplied by $-\mathbf{q} \cdot (d\mathbf{p}/dq_i)$. Thus equation (5) implies that under Cournot, $\theta_i = s_i$.[23]

With conjectural variations, $d\mathbf{q}/d\sigma_i$ has r_i^j in entry $j \neq i$ and 1 in its ith entry. Thus

$$\mathbf{m} \cdot \frac{d\mathbf{q}}{d\sigma_i} = m_i + \sum_{j \neq i} r_i^j m_j.$$

The derivative $d\mathbf{p}/d\sigma_i$ has in each entry $(1 + \sum_{j \neq i} r_i^j)P'$, so

$$-\mathbf{q} \cdot \frac{d\mathbf{p}}{d\sigma_i} = -Q\left(1 + \sum_{j \neq i} r_i^j\right)P'.$$

[23] Note that while the above logic is left unchanged by including a tax, as long as this is included in the definition of m_i, including exogenous competition makes the relevant $s_i = (q_i - \tilde{q}_i)/(Q - \tilde{Q})$. Thus increasing exogenous competition causes θ_i to fall for firms whose effective (after exogenous competition) share it causes to fall and rise for firms whose effective share it causes to rise. Also, θ_i directly depends on interventions in the models that follow. However, given that the derivative of θ_i does not enter into our analysis below, we omit further discussion of this point.

Thus

$$\theta_i = s_i\left(1 + \sum_{j\neq i} r_i^j \frac{m_j}{m_i}\right).$$

As anticipated, θ_i is higher if firms have positive conjectures and lower if they have negative conjectures. Conjectures about firms with relatively large markups matter more than those with small markups. The firm acts monopolistically (like a cartel member) if

$$1 + \sum_{j\neq i} r_i^j \frac{m_j}{m_i} = \frac{1}{s_i}.$$

2. *Differentiated Nash-in-prices.*—Now we have $\sigma_i = p_i$, and following the logic of the previous section, each firm sets $m_i = -q_i/(\partial q_i/\partial p_i)$. We have

$$\mathbf{m}\cdot\frac{d\mathbf{q}}{d\sigma_i} = \sum_j m_j \frac{\partial q_j}{\partial p_i}.$$

On the other hand, $d\mathbf{p}/d\sigma_i$ has entries 0 everywhere but at position i, where it has entry 1. Thus

$$\begin{aligned}\theta_i &= \frac{\sum_j m_j(\partial q_j/\partial p_i)}{-q_i} = \frac{m_i - \sum_{j\neq i} d_i^j m_j}{-q_i[1/(\partial q_i/\partial p_i)]}\\ &= \frac{m_i - \sum_{j\neq i} d_i^j m_j}{m_i} = 1 - \sum_{j\neq i} d_i^j \frac{m_j}{m_i},\end{aligned}$$

where $d_i^j \equiv -(\partial q_j/\partial p_i)/(\partial q_j/\partial p_i)$ is the *diversion ratio between good i and good j* as the effect of an increase in q_i is equivalent to a change in price of size $1/(\partial q_i/\partial p_i)$. This quantity is familiar from antitrust analysis: $d_i^j m_j$ is Farrell and Shapiro's (2010a) *upward pricing pressure* from good j to good i were the firms to merge.

As in the symmetric case, diversion plays a key role in determining θ, except that now the proper weighting is given by relative markups. Note that θ_i may be negative (a firm may be pricing below the socially optimal level) if it has a small markup relative to the firms from which it diverts sales. For instance, suppose that there are only two firms. Then for firm 1,

$$\theta_1 = 1 - d_1^2\frac{m_2}{m_1} = 1 + \frac{\frac{\partial q_2}{\partial p_1}}{\frac{\partial q_1}{\partial p_1}}\frac{q_2}{q_1}\frac{\frac{\partial q_1}{\partial p_1}}{\frac{\partial q_2}{\partial p_2}} = 1 + \frac{\frac{\partial q_1}{\partial p_2}}{\frac{\partial q_2}{\partial p_2}}\frac{q_2}{q_1},$$

where the third equality follows by Slutsky symmetry. The ratio

$(\partial q_1/\partial p_2)/(\partial q_2/\partial p_2)$ may be arbitrarily close to minus one if firm 2 has very little substitution to the outside good. Thus if q_2 is large relative to q_1, θ_1 may be arbitrarily large in magnitude and negative, and thus taxes that fall on (cause reductions in the quantity of) firm 1 may be socially desirable despite firm 1 potentially having positive and even large markup.

A simple set of situations in which this holds is the duopoly case on which Shaked and Sutton (1982) focus in their model of "vertical" product differentiation, where firms are characterized by a quality level u_i and consumers are characterized by an income level y. Consumers with income y gain utility $u_i y$ from consuming a product of quality u_i. Suppose that there are two products, one with higher utility than the other ($u_2 > u_1$), and that some consumer strictly prefers good 1 so that $q_1 > 0$. Then Shaked and Sutton show that any consumer who is indifferent between good 2 and no purchase will strictly prefer purchasing good 1. Thus there is no substitution between good 2 and the outside good; the only substitution out of good 2 is to good 1, and $(\partial q_1/\partial p_2)/(\partial q_2/\partial p_2) = -1$. Thus θ_1 is negative as long as $q_2 > q_1$, which Shaked and Sutton show occurs whenever y is uniformly distributed and the goods have symmetric constant marginal costs; but it can easily be shown that this occurs in many, perhaps most, other cases of interest. Intuitively, if firm 1 is a copycat that diverts lots of sales from good 2 but is highly elastic to outside goods because of its low quality, $\theta_1 < 0$. Other models with vertically differentiated goods and Nash-in-prices competition (such as a Hotelling model in which one good has a vertical quality advantage) behave similarly. On the other hand, if all goods are complements, $\theta_i > 1$ as $d_i^j < 0$ for all i, j.

3. *Monopolistic competition.*—As in the symmetric case, there is a measure one continuum of firms, but now each firm's good may have a different (strictly concave) utility index, $u_i(q_i)$. As before, each firm sets $m_i = -U' u_i'' q_i$. In the continuum model,

$$\mathbf{m} \cdot \frac{d\mathbf{q}}{d\sigma_i} = \int_j m_j \frac{dq_j}{d\sigma_i} \, dj$$

and, mutatis mutandis, for other dot products. As in the traditional quantity interpretation of monopolistic competition, the quantity is the choice variable $dq_j/dq_i = 1$ for $j = i$ and zero otherwise. Thus $\mathbf{m} \cdot (d\mathbf{q}/d\sigma_i) = m_i di$.[24] We have $dU'/d\sigma_i = U'' u_i' \, di$, so for $j \neq i$, $dp_j/d\sigma_i = U'' u_i' u_j' \, di$, and for $j = i$, $dp_j/d\sigma_i = U' u_i''$. Thus

[24] Our notation here in terms of differentials is not precise or standard from a formal perspective. However, it is more convenient and intuitive for readers not familiar with functional analysis. A more formal analysis is available on request.

$$\int_j q_j \frac{dp_j}{d\sigma_i} dj = (U'' u_i' \overline{q_j u_j'} + U' u_i'' q_i) di,$$

where $\overline{q_j u_j'}$ is the average value of $q_j u_j'$ over all other firms. Thus

$$\theta_i = \frac{-U' u_i'' q_i di}{-(U'' u_i' \overline{q_j u_j'} + U' u_i'' q_i) di} = \frac{U' u_i'' q_i}{U'' u_i' \overline{q_j u_j'} + U' u_i'' q_i}.$$

As in the symmetric case, this is < (>) 1 if $U'' > (<) 0$. A natural example is the quasi-linear version of Melitz's (2003) extension of the Dixit-Stiglitz model to firms with heterogeneous but constant marginal costs. In this model each firm has $u_i(q) = q^\beta$ and $c_i(q_i) = q_i/\gamma_i$, where γ_i is a firm's productivity.[25]

We assume, as in the version of the Dixit-Stiglitz model we solved, that $U(x) = x^\eta$. We observe that in terms of

$$x \equiv \int_j u_j dj = \int_j (q_j)^\beta dj,$$

the average $\overline{q_j u_j'}$ may be expressed as

$$\overline{q_j u_j'} = \int_j q_j u_j' dj = \beta \int_j (q_j)^\beta dj = \beta x.$$

The equation for the conduct parameter then becomes

$$\theta_i = \frac{1}{\beta \dfrac{U'' x}{U'} \dfrac{u_i'}{u_i'' q_i} + 1} = \frac{1}{\beta(\eta - 1) \dfrac{1}{\beta - 1} + 1} = \frac{1 - \beta}{1 - \beta\eta},$$

which is constant common across firms and exactly the same as in the basic Dixit-Stiglitz model. Thus in the quasi-linear Melitz model, all firms have the same conduct parameter as they would in a quasi-linear Dixit-Stiglitz model with the same demand parameters.

With asymmetric firms our definition of the relevant taxes and pass-through rates also must change. We now allow for the tax $\boldsymbol{\tau}$ to fall heterogeneously on firms, assuming by way of normalization that a unit of the tax has total quantity-weighted size one: $(\boldsymbol{\tau} \cdot \mathbf{q})/Q = 1$, where $Q = \mathbf{1} \cdot \mathbf{q}$. The size of the tax imposed is denoted by t_τ, so the total tax is $t_\tau \boldsymbol{\tau}$.[26] Pass-through is now a vector, dependent on the $\boldsymbol{\tau}$ considered: $\boldsymbol{\rho}_\tau \equiv d\mathbf{p}/dt_\tau$.

To form a basis for all $\boldsymbol{\tau}$, we assume that either $d\mathbf{p}/d\sigma_i$ or $d\mathbf{q}/d\sigma_i$ are of full rank, as is true in all (nondegenerate) models considered above and

[25] Note that we drop fixed costs as they are not relevant to the static analysis we perform.

[26] The tax $\boldsymbol{\tau}$ is allowed to vary with the level of t, as it may have to maintain the normalization.

all others we are aware of. We define $\boldsymbol{\tau}_i$ to be such that $-\boldsymbol{\rho}_{\tau_i}$ points in the same direction as $d\mathbf{p}/d\sigma_i$ and $-(d\mathbf{q}/dt_{\tau_i})$ points in the same direction as $d\mathbf{q}/d\sigma_i$; we use the negative convention as taxes and quantities typically move in opposite directions.[27] By linear independence, any $\boldsymbol{\tau}$ is a linear combination of $\boldsymbol{\tau}_i$. Thus we collect the coefficients of the linear combination and label them $\boldsymbol{\lambda}^\tau$. Intuitively, $\boldsymbol{\lambda}^\tau$ tells us whom the tax falls on, not physically but in terms of the induced changes in firms' economic strategies (and thus quantities and prices). This effectively extends principle of incidence 1: it states that the incidence of the tax depends not on who pays it but on the real changes in strategies it induces and the pass-through rates associated with the taxes directly affecting those strategies.

To determine the costs of taxation borne by consumers, we can again employ the definition of consumer surplus:

$$\frac{dCS}{dt_\tau} = -\boldsymbol{\rho}_\tau \cdot \mathbf{q} = -\rho_\tau Q.$$

Here $\rho_\tau \equiv \boldsymbol{\rho}_\tau / Q$ is the *quantity-weighted average pass-through rate* (*across firms*), from which we omit a bar to avoid confusion with the previous quantity weighting over levels of the tax rather than across firms.

For the impact on producer surplus, we use the decomposition from above:

$$\begin{aligned}
\frac{dPS}{dt_\tau} &= \sum_i \lambda_i^\tau \frac{dPS}{dt_{\tau_i}} \\
&= \sum_i \lambda_i^\tau \left[\frac{d\mathbf{p}}{dt_{\tau_i}} \cdot \mathbf{q} + \frac{d\mathbf{q}}{dt_{\tau_i}} \cdot (\mathbf{m} - \mathbf{t}) \right] - Q \\
&= \sum_i \lambda_i^\tau (\boldsymbol{\rho}_{\tau_i} \cdot \mathbf{q} - \theta_i \boldsymbol{\rho}_{\tau_i} \cdot \mathbf{q}) - Q \\
&= Q\left[\sum_i \lambda_i^\tau (\rho_{\tau_i} - \theta_i \rho_{\tau_i}) - 1 \right] \\
&= -Q[1 - (1-\theta)\rho_\tau + \mathrm{Cov}(\lambda_i^\tau \rho_{\tau_i}, \theta_i)],
\end{aligned}$$

where $\theta \equiv \sum_i \theta_i / n$ and $\mathrm{Cov}(\lambda_i^\tau \rho_{\tau_i}, \theta_i)$ represents the covariance between the product of (quantity-weighted average) pass-through and the targeting of the tax λ_i^τ in terms of the firms it causes to reduce quantity, on the one hand, and the conduct parameters θ_i on the other hand.

[27] Note that these conditions are always compatible as $d\mathbf{p}/dq_i = (d\mathbf{q}/dq_i)(\partial\mathbf{p}/\partial\mathbf{q})$ and are often equivalent (whenever $\partial\mathbf{p}/\partial\mathbf{q}$ is of full rank). However, under the Cournot models they are not equivalent as $\mathbf{p}$ is unidimensional (constant across firms) in the case of competition and $\mathbf{q}$ is unidimensional in the case of collaboration.

Note, first, how this expression collapses in the special cases relevant to those we have thus far considered:

1. *Perfect competition.*—If $\theta_i = 0$ for all firms, then $dCS/dt_\tau = -\rho_\tau Q$ and $dPS/dt_\tau = -(1-\rho_\tau)Q$. Thus we can allow for multiple products and heterogeneously applied taxes by simply replacing the pass-through rate with the quantity-weighted average pass-through rate, as is typically done for perfectly competitive aggregation.[28]
2. *Monopoly.*—If all firms have $\theta_i = 1$, then $dPS/dt_\tau = -Q$, just as under monopoly. Thus a perfect cartel with multiple products has the same incidence expression (using the quantity-weighted average pass-through) as does a monopoly producing a single product.[29]
3. *Symmetric oligopoly.*—If all firms have the same $\theta_i = \theta$ as in the quasi-linear Melitz model, even if firms are otherwise heterogeneous, the covariance term drops out and we recover the symmetric oligopoly expression $dPS/dt_\tau = -[1-(1-\theta)\rho_\tau]Q$, again weighting by quantities in the pass-through.

The truly novel term is thus the covariance. This states that firms benefit to the extent that taxes tend to fall on (quantities tend to fall for) firms with small θ_i. Note that this controls for any harm to consumers, which depends only on the overall pass-through, and thus it is also in society's interest for taxes to fall on firms with low θ_i. These firms have socially, relatively undistorted strategies, and thus it is less harmful if their prices rise. In fact it actually causes a social gain to impose taxes that are borne by firms that have negative θ_i even when all firms have positive markups, as in the Shaked and Sutton example above. Targeting firms with low average pass-through ρ_{τ_i} is also desirable. Similarly, the more pass-through is concentrated among firms with low conduct parameters, the smaller the burden of taxation on firms.

Global interpretations of the covariance logic above are natural. The same principle of quantity-weighted averaging for integration continues to apply and must now also be applied to the covariance term. For global incidence, $\bar{t}_\tau$ now must be such that the quantities in all markets, not just the one market, are zero. Industries in which firms with the largest quantities have the highest conduct parameters (large θ_i) will require higher taxes on those firms to eliminate all quantity from the market. This will cause a large covariance and thus a heavy incidence on firms. Thus the ratio of consumer to producer surplus will, all else being equal,

[28] This is the reason why under perfect competition we did not use the space to include multiple products or heterogeneously applied taxes: aggregation is too familiar to yield additional insights.

[29] Thus in an analogous way our results here can be extended to multiple, each multi-product, firms, though we do not develop this extension here for the sake of brevity.

be small in industries in which large firms have the least competitive conduct. Similarly, industries in which firms with high pass-through have high conduct parameters will tend to have lower incidence and thus large producer relative to consumer surplus. These are intuitive because producer surplus is reduced (relative to consumer surplus) by a strong representation, in either quantity or weight due to pass-through, of firms with low conduct parameters.

VI. Applications

In this section we discuss a few of the many applications of incidence to substantiate the claim of the preceding analysis to relevance beyond public finance.

A. *Procuring New Markets*

Consider the analysis of Borenstein (1988): a public authority seeks to select the provider(s) of a concession to maximize the social surplus this creates.[30] Suppose that each of the concession operators will charge a uniform price if she is selected to be among the operators, each (oligopolistic) group of operators has a single potential proposal, and the members have private information on both the consumer surplus *CS* and profits *PS* this proposal will generate. The authority is unwilling or unable to monitor prices ex post to avoid monopoly distortions and thus must simply choose the operator generating most surplus. We discuss relaxing these assumptions at the end of this subsection.

Solving for the optimal mechanism in this multidimensional context is beyond the scope of our analysis here, so instead we focus on deriving some analytical principles directly from the logic of incidence.[31] To begin, note that social surplus is the sum of consumer surplus and producer surplus, or $(1+\bar{I})PS$. Because only *PS* affects the incentives of the various potential operators to seek the concession, the reasoning of Jehiel and Moldovanu (2001) suggests that it will typically be impossible to use a mechanism to screen for anything other than *PS* created by

[30] In addition to purely public settings considered by Borenstein, similar trade-offs arise when platforms, such as supermarkets (Armstrong and Zhou 2011) or websites (Edelman, Ostrovsky, and Schwarz 2007), allow product sellers to display their wares or advertisements for these prominently in exchange for payment, because, as Gomes (2012) argues, the platform has an incentive to internalize the consumer surplus generated by these products in order to profit from consumers on other offerings such as fixed fees for using the platform. In these literatures our assumptions of no discrimination, ex post monitoring, or project selection are maintained.

[31] For more on two distinctive approaches to multidimensional mechanism design, see Rochet and Stole (2003) and Veiga and Weyl (2012). The logic of the results given here is closely related to that of the latter paper for obvious reasons.

various proposals. If the planner views $\bar{I}$ as symmetrically distributed across groups of firms (conditional on PS), then she wishes to select the firm with the highest PS if and only if $E[(1+\bar{I})PS|PS]$ is ranked in the same way PS is.[32] A grossly sufficient condition for this is that $\bar{I}$ is distributed independently of PS. Clearly if $\bar{I}$ is constant across all competing groups, this is satisfied, implying the bulleted results discussed in the introduction.[33] However, the greater the noise in $\bar{I}$, the less value there will be in ensuring that the highest PS is selected relative to other goals (e.g., revenue maximization).

More generally, an auction should perform reasonably well as long as there is not a strong negative correlation between $\bar{I}$ and PS. If such correlation were too strong, the planner might want to be unresponsive to PS (as argued by Borenstein) because of the resulting adverse selection, randomizing among symmetric proposals. Especially in such cases, the authority would seek information that would allow it to handicap the auction to favor operators with high expected $\bar{I}$. Factors indicated by the above analysis are that proposals with constant or increasing returns should be favored over those with capacity constraints, those generating highly convex demand should be favored over ones with more concave demand, those with low conduct parameters should be favored over those with high conduct parameters, and those in which the largest and highest pass-through firms have low conduct parameters should be favored over the reverse.[34] We suspect that principles of incidence would also play an important role in the design of the optimal mechanism.

This importance of incidence is likely to carry over if our assumptions of a single, fixed proposal for each cluster and no ex post monitoring are relaxed. For example, Weyl and Tirole (2012) show that when rewards for creating a new market can be based only on equilibrium ex post prices and quantities, the factor that cannot be screened is $\bar{I}$, so again its statistical relationship to factors that can be screened is crucial to optimal policy. Other models allow investments by individuals in changing their projects while maintaining the fact that only PS can be

[32] Technically, only the ranking of the top proposal matters, but in many circumstances a full ranking is useful for the same reasons as in social choice theory (different alternatives may be available at different times).

[33] One special case of those results is Armstrong, Vickers, and Zhou's (2009) result that if concessions are monopolistic and all firms have linear demand and constant returns (which yields constant pass-through of 1/2), then the rankings of profits and social surplus are identical and a simple auction is optimal.

[34] One application of this logic is Nocke and Whinston's (forthcoming) result that marginal cost reductions (through mergers) for small firms in Cournot oligopoly are more desirable than for large firms, because they have lower conduct parameters, from the above analysis. Additionally, while we have assumed uniform pricing, if some firms were able to discriminate more effectively than others, discriminatory firms should be penalized (below their willingness to pay to enter the market) as they will appropriate a greater fraction of the social surplus they create.

observed or screened ex post, analogously to Holmström and Milgrom (1991), or would allow $\bar{I}$ and PS to both be observable ex post, but the set of projects available for proposal would be unobservable, as in Armstrong and Vickers (2010) and Nocke and Whinston (forthcoming). In the former case, the more substitutable investments are in improving the unobservable $\bar{I}$ and the observable PS, the less responsive the award (or at least profits) to PS should be. In the latter case, again both the statistical relationship between $\bar{I}$ and PS and ex ante information about $\bar{I}$ are relevant. Armstrong and Vickers (2010) and Nocke and Whinston (forthcoming) analyze these in the closely related case of mergers that we discuss in Section VI.E, assuming that proposals are selected among those permitted to maximize PS. As Armstrong and Vickers emphasize, minimum social surplus for a project to be acceptable should rise when there is less correlation between $\bar{I}$ and PS or more variance in $\bar{I}$. As Nocke and Whinston emphasize, marginal cost reductions (caused by mergers) for firms with higher I should be favored even conditional on social surplus generated to offset the private bias toward selecting based only on PS (see n. 34).

B. *Supply Chains and Optimal Taxation*

A canonical model of supply chains proposed by Spengler (1950) has an "upstream" firm choosing its price, which is then taken by a (downstream) firm that charges prices to consumers. Natural extensions considered by many authors allow for multiple stages in the supply chain and imperfect competition at each stage.[35]

First, consider a supply chain consisting of several layers of imperfectly competitive firms supplying a necessary input to a downstream sector, which may then supply end consumers or another downstream sector. We focus on the case in which each layer is symmetrically imperfectly competitive, but again this may be relaxed. There is a symmetric-at-symmetric prices demand $q(p_0)$ for the product. This, combined with a supply-side structure, determines an equilibrium pass-through rate ρ_0 of the retailers as a function of the per-unit cost p_1 they are charged by the upstream firms. The upstream firms thus face effective (symmetric-at-symmetric prices) demand $q(p_0(p_1))$ with elasticity $\epsilon_D \rho_0 p_1/p_0$, where ϵ_D is the direct demand elasticity downstream. Thus, when θ_1 is the conduct parameter of the level 1 sector, equilibrium in a symmetric upstream market is given by

[35] Analogous settings arise when firms sequentially choose how much of a homogeneous good to produce, as in the classic von Stackelberg (1934) model, extended by Anderson and Engers (1992) to the case in which this occurs in many stages. The pass-through of quantities at each stage to the final market plays a role analogous to that of cost pass-through along a supply chain. Details are available on request.

$$\frac{p_1 - mc}{p_0}\epsilon_D = \frac{\theta_1}{\rho_0}.$$

As a result, the comparison of markups between the upstream and downstream firms is given by the comparison of θ_1/ρ_0 to θ_0, where θ_0 is the conduct parameter downstream. This reasoning continues up the supply chain, with the aggregate pass-through of all levels beneath determining the incentives faced at each level. This implies that the pass-through from the nth to the $(n-1)$th level will depend on the derivative of the pass-through from the $(n-1)$th to the nth level and thus on the $(2+n)$th derivative of demand, in principle allowing the identification from markup data of very high-order properties of demand, extending the logic of Villas-Boas and Hellerstein (2006). Conversely, if constant pass-through is assumed, many of these effects disappear, strong predictions are implied, and the model is highly overidentified.

A slightly modified application of this reasoning involves a two-stage chain in which the first-stage firm (usually interpreted as a national government taxing foreign trade or the regulator of an imperfectly competitive industry) internalizes the welfare of final consumers or the taxed downstream industry to greater or lesser extents. While the logic of incidence provides a useful framework for any set of weights, we focus on the case considered by Brander and Spencer (1981) and Laffont and Tirole (1993) when consumer surplus is fully internalized and the welfare of the imperfectly competitive firms is entirely neglected.

The government charges a specific tax t. If D is the equilibrium demand of consumers, the marginal loss to consumers of the product is ρD and to the government on inframarginal tax is $-t\rho D'$, while the marginal revenue gain to the state is D. Thus the optimum requires

$$1 = \rho\left(1 + \frac{t\epsilon_D}{p}\right) \Rightarrow \frac{t^*}{p} = \frac{1-\rho}{\rho\epsilon_D}.$$

Note that this formula in no way depends on the existence of imperfect competition; it applies equally well to the setting in which the foreign firms are perfectly competitive. It thus unifies the monopoly analysis of Brander and Spencer (1981) with the classic analysis of terms of trade reasons for taxing imports as in Johnson (1953–54). It may also be easily extended to asymmetric industries. The only difference is that with an imperfectly competitive foreign sector, it is possible that $\rho > 1$, and thus a negative tax (subsidy) on imports may, in principle, be optimal. Thus the two theories are one, at least this far.

However, the externality of a tax on the foreign sellers is strictly greater with conduct parameters above zero: rather than $D(1-\rho)$, the burden on the foreign industry is $[1-(1-\theta)\rho]D$. Thus there will be stronger incentives for international trade agreements to limit such taxes between countries where firms have higher conduct parameters and in models in which firms exercise this power, as shown by Ossa (2011). This is essentially the case in which the weight on the regulated or foreign industry's welfare is nonzero because trade negotiations lead it to be internalized.

C. *Third-Degree Price Discrimination*

A recent literature has revisited classical questions in the theory of monopoly price discrimination using an approach closely related to that employed here. Aguirre, Cowan, and Vickers (2010) return to one of the oldest questions in industrial organization, posed by Pigou (1920): When does explicit third-degree price discrimination by a monopolist raise output or welfare?[36]

Consider two markets, strong (S) and weak (W). Without discrimination, prices are constrained to be identical. With discrimination, prices in S exceed those in W by Δ. Aguirre et al. propose a natural continuous path from no discrimination to discrimination: we require that $p_S \leq p_W + \delta$. Assume that profits in each market, π_S and π_W, are concave in price. Then for any $\delta \in [0, \Delta]$, the monopolist will choose $p_S = p_W + \delta$. Her first-order condition is thus $\pi_S'(p_W + \delta) + \pi_W'(p_W) = 0$. For $\delta < \Delta$, $\pi_W' < 0 < \pi_S'$, but these both converge to zero as δ goes to Δ.

A firm facing exogenous quantity $\tilde{q}$ earns profits $[q(p) - \tilde{q}](p - c)$. Her first-order condition is thus $q'(p)(p - c) + q(p) - \tilde{q}$, while the first-order condition in the high market in the price discrimination problem is $q_S'(p)(p - c) + q_S(p) + \pi_W'(p - \delta)$. In effect, the downward pressure on prices from the constraint against discrimination in the low market enters in the same way as exogenous quantity. Moving toward discrimination is therefore equivalent to moving exogenous quantity from the strong market to the weak market.

Thus Aguirre et al. (2010) show that discrimination leads to higher output if an average of pass-through in the weak market exceeds that in the high market. Similarly, the change in social welfare in each market from the change in quantity is $\int m dq$, so a comparison of an average of the markup times the pass-through over the relevant range in the two markets determines the welfare effect of discrimination. The connections of pass-through to demand curvature make it clear how this result immediately implies the famous prior results of Pigou (1920), Robinson (1933), Schmalensee (1981), and Varian (1985) on the connections between demand curvature and the effects of discrimination.

[36] While we focus on the social welfare and output effects, Cowan (2012) also uses incidence elegantly to study the effects on consumer surplus. Extensions of his logic similar to those we propose below for output and welfare are possible for consumer surplus.

The logic of incidence can be used to extend both of these results to symmetric imperfect competition. Suppose that, with or without discrimination, each market is governed by the same conduct $\theta = [(p - c)/p]\epsilon_D$, where ϵ_D is either independent or pooled depending on whether discrimination is allowed or not. Without discrimination,

$$\epsilon_D = \frac{q_S \epsilon_{D_S} + q_W \epsilon_{D_W}}{q_S + q_W}.$$

Thus

$$\begin{aligned}\epsilon_D &= -\frac{p(q_S' + q_W')}{q_S + q_W} \\ &= -\frac{pq_S'}{q_S - \dfrac{q_W' q_S - q_S' q_W}{q_S' + q_W'}} \\ &= -\frac{q_W' p}{q_W + \dfrac{q_W' q_S - q_S' q_W}{q_S' + q_W'}}.\end{aligned}$$

Therefore, the argument that the prohibition on discrimination acts as equal and offsetting exogenous quantity competition in the two markets in an amount $(q_W' q_S - q_S' q_W)/(q_S' + q_W')$ holds generally. Because quantity pass-through in the two markets is $\theta\rho$, if θ is constant in p, precisely the same results, interpreted in terms of averages of pass-through or of demand curvature (as this is a simple transformation of pass-through), hold under imperfect competition.[37]

Similarly, the result may also be shown to hold when θ is not the same in the two markets. Then, if θ is higher in the strong market, it is clearly more likely, ceteris paribus, that discrimination is harmful as it is more likely that averaged over the relevant range $\theta_S \rho_S m_S > \theta_W \rho_W m_W$. This is a generalization of the result of Holmes (1989), who argues that when discrimination is in favor of individuals for whom competition is more intense (in differentiated products Nash-in-prices competition), discrimination is more likely to be harmful because $\theta = 1 - A$ and A, the aggregate diversion ratio, measures the degree of competition for customers within the market.

D. Strategic Effects

Bulow et al. (1985b) highlight the importance of strategic effects (whether one firm raising or lowering its price or quantity causes others to follow

[37] If θ is not constant, then the result must be replaced with averages of $\theta\rho$.

or do the opposite) in a variety of problems in oligopoly theory, particularly strategies to deter entry or affect postentry competition as studied by Fudenberg and Tirole (1984) and Bulow, Geanakoplos, and Klemperer (1985a). In this subsection we discuss the close relationship between strategic effects and their global analogue, the strategic impact of entry, on the one hand and incidence and pass-through on the other.

First, consider quantities. Suppose that one firm increases the quantity it produces of a good that is homogeneous with other goods. This is precisely the same, from the perspective of the other firms that take quantities as given as in Cournot, as an increase in the exogenous competition discussed in Sections III and IV. Thus if the remaining firms are symmetric imperfect competitors (or a residual monopolist), whether they respond by increasing or decreasing their own output is determined by the sign of $\theta\rho - 1$ with constant marginal cost and, by the logic of Appendix A, by $[\theta + (\epsilon_D - \theta)/\epsilon_S)]\rho - 1$ more generally. Because Bulow et al. consider only duopoly settings, this strictly generalizes their results. For example, it gives a simple demonstration of why strategic substitutes are more likely under Cournot when the number of firms is large, as $\theta = 1/n$ in this case. Appendix A gives further results on the sign of the strategic effect in quantities.

Second, consider prices. When one firm raises its prices, this raises the willingness of consumers to pay for other products. In many models, the result is a parallel shift upward in willingness to pay. For example, in symmetric oligopoly models with no outside good such as the "spokes" model of Chen and Riordan (2007) or the random utility model of Perloff and Salop (1985), any individual firm raising its price increases the price of the only "outside option" for the rest of the symmetrically imperfectly competitive industry. This is equivalent to increasing willingness to pay for the remaining industry by an identical amount for all individuals. By principle of incidence 1 under symmetric imperfect competition, a unit increase in consumer willingness to pay will raise prices if and only if $1 - \rho > 0$ or, equivalently, $\rho < 1$. This explains why Gabaix et al. (2013) find that in the Perloff and Salop model, entry (a new firm entering the market and thus effectively reducing its price from infinity) raises (lowers) prices under constant marginal cost if the distribution of valuations is log convex (log concave) given the relationship between pass-through and log curvature under constant marginal cost discussed extensively in Section III. However, it holds more broadly and would also apply if costs were allowed to be nonlinear. Chen and Riordan (2008) and Quint (2013) obtain similar relationships, even while relaxing the assumption of no outside substitution.

Bulow et al.'s examples of the importance of such strategic effects are too numerous to consider here. So we just give one example of how the logic of incidence relates to the normative consequences of strategic complementarity and substitutability. Consider, as in Mankiw and Whinston (1986), a firm that enters production of goods already existing in an industry. We specialize compared to Mankiw and Whinston by assuming a symmetric industry prior to the entry and constant marginal cost (though both of these assumptions can be relaxed at some expositional costs) but generalize by allowing any symmetric oligopoly setting (including monopolistic competition or differentiated products). The new firm's entry leads to some equilibrium production by the new firm, which effectively acts as an exogenous quantity entering the market. By the logic of Section IV, deadweight loss falls by $m\theta\rho$ (for each unit sold) while the entering firm gains profits m taken from existing firms. Thus the new firm's profits on a small unit of increased production are greater (less) than its social contribution if and only if, as Mankiw and Whinston find, $\theta\rho <$ ($>$) 1; that is, there are strategic substitutes (complements) in quantities. By markup-weighted averaging and using the logic of principle of incidence 5 under symmetric imperfect competition, these results can be extended to discrete entry in a more quantitative manner than in Mankiw and Whinston's study. The intuition behind the result may also be clarified by the incidence logic: entry is excessive (insufficient) to the extent that negative (positive) real nonpecuniary externalities are created by entry by reducing (increasing) the profitable quantities of existing firms.

E. Other Applications

The working paper version of this article discussed a number of other applications in detail similar to the examples above. Here we mention them only briefly.

1. *Platforms.*—In platform industries, where individuals' utility from consumption depends on how many other individuals participate on the platform, an important trade-off is often between reducing prices and increasing network benefits. For example, in the Rochet and Tirole (2003) model, policy analysis has focused on whether individuals on one side of the market (say credit card–accepting merchants) might benefit from paying higher prices that subsidize participation on the other side of the market. Both the amount of subsidy that a price rise finances and the relative size of the benefits to inframarginal individuals from more network benefits compared to the losses from higher prices are closely related to incidence. A number of results in Rochet and Tirole's model and various extensions thereof can be characterized parsimoniously in terms of incidence; see the now-defunct working paper Weyl (2009b), as well as Weyl (2009a) and Goos, Van Cayseele, and Willekens (2012) for more details. The welfare economics of other, quite different platform models, such as Becker (1991), also turn on incidence properties; details are available on request.

2. *Mergers.*—A long line of work (Shapiro 1996; Werden 1996; Farrell and Shapiro 2010b) has established a close connection between the impact of mergers in differentiated products industries and the effects of changes in cost. These ideas have been incorporated into policy through the new US and UK horizontal merger guidelines. They suggest that agency investigators consider the equivalent of a merger in marginal cost changes to determine its competitive effects. Jaffe and Weyl (forthcoming) show that a (matrix) product of pass-through rates and these equivalent cost changes is a first-order approximation to the effect of a merger on prices, where the approximation's error is proportional to the third derivative of the demand system and the square of the size of the equivalent cost changes. In the working paper version of this article, we showed a stronger result: that because of the relationship between local pass-through and global incidence, under constant marginal cost and constant pass-through, a merger to monopoly between two firms with small diversion eliminates a fraction of consumer surplus equal to the diversion ratio between the products. Similar approximations may be used to derive the effect of mergers on profits or deadweight loss, further exploiting the logic of incidence.

3. *Product and market design.*—Spence (1975) and Johnson and Myatt (2006) ask how firms choose a nonprice characteristic of their product to affect the demand according to a specified, parametric relationship between demand and the characteristic. An alternative approach to the problem of product (or market) design is to assume that a firm or producer surplus-maximizing entity can choose any arrangement of demand and supply subject to some constraint. A simple constraint would be that total potential gains from trade are constant. The principles of incidence give simple ways to compare different arrangements subject to this constraint: firms want arrangements in which $\bar{I}$ and $\widetilde{SI}$ are both low. This, for example, means that monopolists with constant marginal costs prefer demand to be as concave as possible, that perfect competitors prefer inelastic supply and elastic demand, that symmetric competitors prefer θ to be as close to one as possible and to increase at higher prices, and that asymmetric competitors put a premium on high conduct parameters at large, high-pass-through firms.

4. *Behavioral welfare analysis.*—There has been a recent revival of interest, surveyed by Mullainathan, Schwartzstein, and Congdon (2012), in Dixit and Norman's (1978) analysis of welfare when consumers may fail to act or be persuaded not to act in their own best interest. Incidence plays a crucial role in much of this analysis; we consider two examples. Farrell (2008) considers markets in which secret fees may be charged on goods that consumers do not understand. These fees act as consumer-financed subsidies of the products but may actually benefit consumers if $I > 1$. Dixit and Norman consider advertising that uniformly raises

the willingness of consumers to pay for a product. This acts as a similar subsidy, but Dixit and Norman are concerned with social welfare, which is always increased locally by such a subsidy if $\theta > 0$. Thus, instead of this local question, they ask when and to what extent advertising is excessive. While they derive qualitative conditions to ensure excessive advertising, their analysis can be made more quantitative and extended beyond the monopoly setting they consider using the logic of incidence as this determines how the gains from the subsidy are split between externalities (from the firm's perspective) to consumers and benefits to the firm. Even in rational models of uniform-shift advertising, such as that of Becker and Murphy (1993), incidence plays a crucial role in welfare analysis for similar reasons.

5. *Demand systems.*—Demand curvature plays a key role in determining pass-through and thus incidence when firms have market power. While most standard demand forms used when modeling imperfect competition allow flexibility in the elasticity of demand, few are flexible in the curvature properties of demand, as we show in related work (Fabinger and Weyl 2012). In some cases, imposed restrictions are not based on clear economic intuitions but instead derive from convenience; in others, imposing restrictions based on clear intuitions requires sacrificing tractability. We thus propose a new class of *adjustable pass-through* demand forms that substantially increase the flexibility of curvature while maintaining tractability and nesting the most common demand forms.

6. *Demand estimation.*—Atkin and Donaldson (2012) explicitly use the role of incidence as a "sufficient statistic" and the structure of our results above to analyze the degree of competition in markets and the division of surplus from globalization. They consider markets for various internationally traded commodities in different locations within developing countries in South Asia and sub-Saharan Africa. In a symmetric imperfect competition model, they impose three key assumptions: that demand curvature is constant (demand is in the Bulow and Pfleiderer [1983] class) and the same across markets for a given product, returns to scale are constant, and conduct is invariant to prices (θ is constant). They then use the variation in empirical pass-through in the face of global price shocks across geographic locations for a given product to back out θ. Integrating and using the fact that under their assumptions local and global incidence are identical, they determine the division of surplus arising from the market existing between the intermediaries and consumers. Similarly, Miller, Remer, and Sheu (2013) consider a differentiated products Nash-in-prices model in which θ corresponds to the aggregate diversion ratio.[38] They exploit common structural assumptions about demand and cost curvature that (as discussed in the previous point

[38] They consider a more general asymmetric model, but the logic is analogous.

on demand systems) impose values of ϵ_S and ϵ_{ms} to recover diversion ratios from observed pass-through rates.

VII. Conclusion

This article argues that incidence offers just as powerful a framework for organizing the analysis of comparative statics and welfare under imperfect competition as it does under perfect competition. Analysis of imperfect competition typically eschews the language of incidence and is labeled “industrial organization” or “international trade” while analysis based on incidence usually avoids imperfect competition and is labeled “public finance” or “development.” We believe that this dichotomy is false.

In fact, we have argued that, to paraphrase the conclusion of Bulow et al. (1985b), the crucial question for welfare in imperfectly competitive markets is typically not “Do these markets exhibit price competition or quantity competition or competition using some other strategic variable?”; “Are products differentiated, how many firms are there, do firms act strategically, or are they monopolistic competitors?”; or even Bulow et al.’s “Do competitors think of the products as strategic substitutes or as strategic complements?”; but rather, “What is the pass-through and incidence of a tax in this market?” Unlike the first group of questions, this last is not “new” to the theory of imperfect competition. Rather it is what, at least since the time of Marshall, economists have been asking about competitive markets to analyze a wide range of outcomes and policies. And as discussed in Section VI, once incidence and pass-through have been determined, there is, for many questions, little difference between the relevant analysis in perfectly competitive markets and imperfectly competitive markets. Thus the analysis of “strategic” industries with market power may not be as distinct as it may at first seem from the analysis of perfectly competitive markets.

One of the most important weaknesses of our analysis was that we followed nearly all partial equilibrium literature in assuming that all goods outside the industry under consideration were perfectly competitively supplied (with no externalities). Section V made clear how problematic this assumption is: if goods outside the considered industry that are either complementary to or substitutable with goods in the industry have positive markups, then firms in the industry having zero markups is not typically optimal. Considering detailed behavior outside the industry of interest would defeat much of the simplifying value of a partial equilibrium analysis. However, it would be useful to develop a version of the analysis here in which it was assumed that all goods outside the market to which consumers might substitute have some fixed, average markup from the economy rather than zero markup.

Another limitation was that all of our oligopoly models assumed complete information. A natural direction to extend our analysis would be to allow firms to be uncertain about their competitors' cost or demand and consider the impact of this informational environment on the industry's conduct. Three additional applications and extensions of the framework also seem natural:

- Shifting bargaining power from one side to the other side of the market in the Riley and Samuelson (1981) model of bilateral trade can be shown to have effects similar to changing the amount of exogenous competition facing each side of the market. Thus if the principles of incidence could be extended to such bilateral imperfect competition settings, they might be used to provide an elegant characterization of the incidence of bargaining power in bilateral trade.
- Almost all international trade models use explicit, often constant pass-through, demand forms to obtain results, which are known to vary on the basis of, for example, whether linear or constant elasticity demand is employed. It thus seems likely that incidence plays an important role in the comparative statics of such models.
- Finally, we assumed that firms' only instruments were uniform prices and all consumers were homogeneous in their value to firms. Extending the logic of incidence to cases with heterogeneously valuable consumers and nonprice, or discriminatory price, instruments as summarized by Veiga and Weyl (2012) is a promising direction for future research.

Appendix A

Social Incidence with General Cost

Our results on deadweight loss assume constant marginal cost, but many of the general intuitions derived there extend to, or are even strengthened with, nonlinear costs. First, consider the relationship between pass-through and $dq/d\tilde{q}$. Skipping directly to the symmetric imperfect competition model, and not considering the notation-intensive generalization to general imperfect competition, we get

$$\theta = -\frac{p - mc(q-\tilde{q}) - t}{p}\frac{p}{(q-\tilde{q})p'} = -\frac{p - mc(q-\tilde{q}) - t}{(q-\tilde{q})p'}.$$

However, now the impact of an increase in $\tilde{q}$ directly (the partial derivative) on the right-hand side is

$$-\frac{(q-\tilde{q})mc'(q-\tilde{q}) + p - mc(q-\tilde{q}) - t}{(q-\tilde{q})^2 p'} = \frac{\theta + [(\epsilon_D - \theta)/\epsilon_S]}{q-\tilde{q}}.$$

On the other hand, the impact of a change in t is $1/(q-\tilde{q})p'$. Therefore,

$$\frac{dq}{d\tilde{q}} = \left(\theta + \frac{\epsilon_D - \theta}{\epsilon_S}\right)\rho.$$

Thus when there are declining returns to scale, $\theta\rho$ is smaller, and when there are increasing returns, $\theta\rho$ is larger than the effect of competition on quantities. Declining returns to scale reduce pass-through and increasing returns increase it, so we can say that returns to scale have a larger impact on $dq/d\tilde{q}$, driving a wedge between them even in the monopoly case of $\theta = 1$.

This competition pass-through, which we now call ρ_c, is always below unity if purely demand-driven quantity pass-through (that which would prevail with constant returns)

$$\hat{\rho} \equiv \frac{\theta}{1 + \frac{\theta}{\epsilon_\theta} + \frac{\theta}{\epsilon_{ms}}} < 1.$$

To see this note that

$$\rho_c - 1 = \frac{\theta + \frac{\epsilon_D - \theta}{\epsilon_S}}{1 + \frac{\epsilon_D - \theta}{\epsilon_S} + \frac{\theta}{\epsilon_{ms}} + \frac{\theta}{\epsilon_\theta}} - 1 = \frac{\left(1 - \frac{1}{\epsilon_{ms}} - \frac{1}{\epsilon_\theta}\right)\theta - 1}{1 + \frac{\epsilon_D - \theta}{\epsilon_S} + \frac{\theta}{\epsilon_{ms}} + \frac{\theta}{\epsilon_\theta}},$$

while

$$\hat{\rho} - 1 = \frac{\theta - \left(1 + \frac{\theta}{\epsilon_{ms}} + \frac{\theta}{\epsilon_\theta}\right)}{1 + \frac{\theta}{\epsilon_{ms}} + \frac{\theta}{\epsilon_\theta}} = \frac{\left(1 - \frac{1}{\epsilon_{ms}} - \frac{1}{\epsilon_\theta}\right)\theta - 1}{1 + \frac{\theta}{\epsilon_{ms}} + \frac{\theta}{\epsilon_\theta}}.$$

The numerators of these two expressions are the same, and both have positive denominators as long as the equilibrium is stable under constant marginal cost. Thus, assuming such stability, the sign of $\rho_c - 1$ (the strategic effect, substitutes vs. complements discussed in Sec. VI.D) is determined by that of $\hat{\rho} - 1$. Notice that decreasing returns to scale move ρ_c toward unity (compared to the constant returns case) by increasing both the numerator and the denominator while increasing returns have the opposite effect.

Finally, consider the incidence calculations. For these, it is crucial to specify the costs at which the exogenous units are produced. For this purpose, consider an alternative experiment. Rather than introducing exogenous competition, imagine that the state confiscates any profits earned on the first $\underline{q}$ units. Then equilibrium is now

$$\theta = -\frac{p - mc(q) - t}{(q - \underline{q})p'},$$

and thus $dq/d\underline{q} = \theta\rho$.

Profits per firm are now $p(q)(q - \underline{q}) - c(q) + c(\underline{q})$, so the fall in profits from an increase in $\underline{q}$ is

$$\begin{aligned} &p'\theta\rho(q - \underline{q}) + p(\theta\rho - 1) - mc(q)\theta\rho + mc(\underline{q}) \\ &= m(-\rho + \theta\rho - 1 + \alpha) = -[1 + \rho(1 - \theta) - \alpha]m, \end{aligned}$$

where $\alpha \equiv [mc(q) - mc(\underline{q})]/m$. The argument for calculating deadweight loss incidence proceeds exactly as before, so we obtain relative efficiency gains compared to profit losses of

$$\frac{\theta\rho}{1 + \rho(1 - \theta) - \alpha}.$$

These can be converted, just as in the text, to global incidence formulas. Notice that with decreasing returns, $\alpha > 0$, and thus deadweight loss is larger relative to profit than given by the formula in the text. This is intuitive because with decreasing returns, the existence of a competitive rent makes profits positive even in the absence of a monopoly distortion. When returns are increasing, $\alpha < 0$, and thus deadweight loss is larger relative to profits than the formula given in the text indicates. Again this is intuitive as with increasing returns the competitive rent is negative, reducing profits relative to deadweight loss. Thus the basic source of the divergence is that while in the main text changes in cost affected only profits, here increasing costs also affect the size of the deadweight loss triangle directly. Thus increasing or decreasing marginal costs have an additional impact on the DWL/*PS* that they do not have on the *CS*/*PS* ratio.

Appendix B

Conduct Parameter under Discrete Choice

To complement our discussion in Section IV, in this appendix we derive properties of the conduct parameter θ in the symmetric discrete-choice model, which is related by $\theta = 1 - A$ to the aggregate diversion ratio A defined after equation (1). We provide two different sufficient conditions under which θ is increasing in price. For values of different goods distributed independently and drawn from the same distribution, one of the conditions is equivalent to a condition in an associate monopoly problem, namely, $1/\epsilon_{ms}$ decreasing in price for a certain range of prices.

A. Discrete-Choice Duopoly

Starting from equation (3), we observe that $1/A$ may be expressed as a weighted average of the function $\hat{f}(u)$, defined as

$$\hat{f}(u) \equiv -\frac{\displaystyle\int_{-\infty}^{u} f_1(u, u')\,du'}{f(u, u)}$$

over the interval (p, ∞) with weight function $w(u) \equiv f(u, u)$:[39]

$$\begin{aligned}\frac{1}{A} &= \frac{\int_{-\infty}^{p} f(p, u)\,du + \int_{p}^{\infty} f(u, u)\,du}{\int_{p}^{\infty} f(u, u)\,du} \\ &= -\frac{\int_{p}^{\infty}\int_{-\infty}^{u} f_1(u, u')\,du'\,du}{\int_{p}^{\infty} f(u, u)\,du} \\ &= \frac{\int_{p}^{\infty} \hat{f}(u) w(u)\,du}{\int_{p}^{\infty} w(u)\,du}.\end{aligned} \tag{B1}$$

Here the second equality may be verified by noting that

$$f(p, u_2) = -\int_{p}^{\infty} f_1(u_1, u_2)\,du_1,$$

$$f(u_2, u_2) = -\int_{u_2}^{\infty} f_1(u_1, u_2)\,du_1,$$

$$1_{u_1>p>u_2} + 1_{u_1>u_2>p} = 1_{u_2<u_1} 1_{u_1>p},$$

up to a set of measure zero, and

$$\begin{aligned}\int_{-\infty}^{p}\int_{p}^{\infty} f_1(u_1, u_2)\,du_1\,du_2 + \int_{p}^{\infty}\int_{y}^{\infty} f_1(u_1, u_2)\,du_1\,du_2 \\ = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(u_1, u_2)(1_{u_1>p>u_2} + 1_{u_1>u_2>p})\,du_1\,du_2 \\ = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f_1(u_1, u_2) 1_{u_2<u_1}\,du_2\right] 1_{u_1>p}\,du.\end{aligned}$$

Equation (B1) has two immediate consequences: (*a*) If $\int_{-\infty}^{u} f_1(u_1, u_2)\,du_2 \geq 0$ in some neighborhood of p, then $\hat{f}(u) \leq 0$ in this neighborhood. Since A is manifestly positive, equation (B1) implies that in this neighborhood A decreases with p, and consequently θ increases with p. In this way we recover the result that was obtained in Section IV by more elementary methods. (*b*) If $\hat{f}(u)$ is increasing for

[39] Note that all equalities in (B1) will hold even for $f(u_1, u_2)$ that is not symmetric in its arguments (i.e., $f(u_1, u_2) \neq f(u_2, u_1)$ in general) provided that we replace A with the diversion ratio d_1^2 between good 1 and good 2 defined in Sec. V. See also n. 17.

all $u \in (\underline{p}, \infty)$, then $\theta = 1 - A$ is increasing in p for all $p \in (\underline{p}, \infty)$. We now apply this sufficient condition for θ to be increasing to an important special case.

If the value distributions are independent and identical across goods and have cumulative distribution function (CDF) $G(u)$, that is, if $f(u_1, u_2) = G'(u_1)G'(u_2)$, then $\hat{f}(u) = -G(u)G''(u)/G'^2(u)$. In this case it is helpful to think of $\hat{f}(u)$ in terms of an auxiliary single-good monopoly problem with demand $\tilde{q}(\tilde{p}) \equiv G(-\tilde{p})$. If the demand curve in this associate problem is interpreted as arising from a discrete-choice setting in which each customer can buy at most one unit of the good, the CDF $\tilde{G}(\tilde{u})$ of customers' valuations is related to $G(u)$ by a reversal: $\tilde{G}(\tilde{u}) = 1 - \tilde{G}(-\tilde{u})$. In terms of the auxiliary monopoly problem, the expression $\hat{f}(u) = -G(u)G''(u)/G'^2(u)$ may be written as $(1/\epsilon_{ms}) - 1$ at price $\tilde{p} = -u$ because

$$\epsilon_{ms} \equiv \frac{ms}{ms'\tilde{q}} = \frac{\tilde{q}'^2}{\tilde{q}'^2 - \tilde{q}\tilde{q}''}.$$

Our observation b discussed above then leads to the following conclusion. If $1/\epsilon_{ms}$ is decreasing in $\tilde{p}$ for $\tilde{p} \in (-\infty, -\underline{p})$,[40] then $\hat{f}(u)$ is increasing in u for $u \in (\underline{p}, \infty)$, and consequently, θ is increasing in p for $p \in (\underline{p}, \infty)$.

In Fabinger and Weyl (2012) we show that $1/\epsilon_{ms}$ is globally decreasing for the normal (Gaussian), logistic, and type I extreme value (Gumbel) distributions as well as for their reversals $\tilde{G}(\tilde{u}) = 1 - \tilde{G}(-\tilde{u})$. As a result, for these distributions the duopoly conduct parameter θ rises in response to an increase in the price p.

B. *Generalization to Discrete-Choice Oligopoly*

For oligopoly with n symmetric firms, the aggregate diversion ratio at $p_1 = p_2 = \cdots = p_n \equiv p$ equals

$$A \equiv -\sum_{i=2}^{n} \frac{\partial q_i}{\partial p_1} \Big/ \frac{\partial q_1}{\partial p_1} = -\sum_{i=2}^{n} \frac{\partial q_1}{\partial p_i} \Big/ \frac{\partial q_1}{\partial p_1} = -\frac{\partial Q_1}{\partial p_s} \Big/ \frac{\partial Q_1}{\partial p_1}$$

at $p_1 = p_s$, where $Q_1(p_1, p_s)$ is defined as $q_1(p_1, p_2, \ldots, p_n)$ evaluated with prices $p_2, p_3, \ldots, p_n$ set equal to a "symmetric price" p_s. Since the final expression contains only derivatives of the function Q_1 with respect to p_1 and p_s, for the purposes of evaluating A, we can think of the oligopoly problem as a duopoly problem with good 1 and a "composite good," namely, the right to consume a single unit of any single good from the set $\{2, 3, \ldots, n\}$. The price of the composite good is p_s. The joint CDF $F(u_1, u_s)$ of the duopoly problem is related to the joint CDF $F(u_1, u_2, \ldots, u_n)$ of the original oligopoly problem by $F(u_1, u_s) = F(u_1, u_s, \ldots, u_s)$.

Our duopoly result (B1) then implies for the oligopoly aggregate diversion ratio A at a given p[41]

[40] As mentioned previously, if we let the auxiliary monopoly problem have constant marginal cost, then $1/\epsilon_{ms}$ locally decreasing in price $\tilde{p}$ is equivalent to the pass-through rate locally increasing in price $\tilde{p}$, since $\rho(\tilde{p}) = \epsilon_{ms}/(\epsilon_{ms} + 1)$.

[41] Here we use the asymmetric duopoly version mentioned in n. 39; see also n. 17.

$$\frac{1}{A} = \frac{\int_{\underline{p}}^{\infty} \hat{f}(u) w(u) du}{\int_{\underline{p}}^{\infty} w(u) du},$$

$$\hat{f}(u) \equiv -\frac{1}{n-1} \frac{F_{11}(u, u, \ldots, u)}{F_{12}(u, u, \ldots, u)},$$

$$w(u) \equiv (n-1) F_{12}(u, u, \ldots, u),$$

since

$$F_{12}(u, u) = \sum_{i=2}^{n} F_{1i}(u, u, \ldots, u) = (n-1) F_{12}(u, u, \ldots, u).$$

In analogy with the duopoly case, we identify two immediate consequences: (a) The function $F_{11}(u, u, \ldots, u)$ locally nonnegative leads to locally increasing $\theta = 1 - A$. (b) If $\hat{f}(u)$ is increasing for all $u \in (\underline{p}, \infty)$, then $\theta = 1 - A$ is increasing in p for all $p \in (\underline{p}, \infty)$.

For values independent and drawn from the same distribution with CDF $G(u)$, we have[42]

$$\hat{f}(u) = -\frac{1}{n-1} \frac{G(u) G''(u)}{G'^2(u)}.$$

This means that the same sufficient condition on $G(u)$ for $\theta = 1 - A$ to be increasing that we discussed in the duopoly case applies to the case of oligopoly as well: if $1/\epsilon_{ms}$ is decreasing in $\tilde{p}$ for $\tilde{p} \in (-\infty, -\underline{p})$, then $\hat{f}(u)$ is increasing in u for $u \in (\underline{p}, \infty)$, and consequently, θ is increasing in p for $p \in (\underline{p}, \infty)$. As mentioned before, this condition is globally satisfied for normal (Gaussian), logistic, and type I extreme value (Gumbel) distributions.

Appendix C

Pass-Through in the General Model

In this appendix we discuss pass-through in the general model of Section V, which relaxes the assumption of symmetry between firms and is likely to be particularly useful for applied work. We demonstrate that even in this general case the pass-through rate is determined by the same forces, and we provide explicit formulas for the pass-through rate in terms of other important economic variables.

[42] For a given CDF $G(\cdot)$ and a given u, $\hat{f}(u) \to 0$ as $n \to \infty$. This does not imply, however, that A could exceed one for large enough n. The relevant weighted average $\hat{f}(u)$ stays above one thanks to the fact that $w(u)$ depends on n and with increasing n its region of large values shifts to higher u, where $\hat{f}(u)$ is large.

A. The Conduct Parameter Matrix and the Lerner Condition

Let us first introduce notation that is particularly well suited for manipulation of mathematical objects with more than two indices. There are two types of indices. Indices $i, j, k, \ldots$ are the standard indices taking values in $\{1, 2, \ldots, n\}$. Whenever these are summed over, there will be an explicit summation sign. The other types of indices are $a, b, c, \ldots$. These have exactly the same meaning, except that they are subject to the Einstein summation convention: In each product of elementary factors, any such index (i.e., any such letter $a, b, c, \ldots$) can appear at most twice. If it appears twice, it is being summed over even though no explicit summation sign appears. Also, derivatives with respect to the ith (or ath) argument will be denoted by subscript i (or a) after a comma. So, for example, $q_{i,j} \equiv \partial q_i / \partial p_j$ (here q_i is a function of $(p_1, \ldots, p_n)$). Whenever we use this notation for derivatives, quantities are assumed to be functions of prices, or vice versa.

The definition (4) of θ_i is

$$\theta_i \frac{d\mathbf{p}}{d\xi_i}\mathbf{q}^T + \frac{d\mathbf{q}}{d\xi_i}(\mathbf{m} - \mathbf{t})^T = 0.$$

The ith firm here can choose outcomes from its one-dimensional choice set, embedded in the n-dimensional space of all possible production vectors $\mathbf{q}$ (or alternatively the n-dimensional space of all possible price vectors $\mathbf{p}$; these are equivalent since the demand system is given). The choice set is parameterized by ξ_i. For example, $d\mathbf{q}/d\xi_i$ is the tangent vector along the choice set of firm i. Now let us write this equation in components:

$$\theta_i \frac{dp_a}{d\xi_i} q_a + \frac{dq_a}{d\xi_i}(m_a - t_a) = 0.$$

In this equation, p_a and q_a are ordinary functions of just one variable: ξ_i with a definite i. The change $d\xi_i$ always represents motion along just one choice variable, the one that firm i can choose. We can, however, also consider directions along which other firms can choose outcomes. At each point in the q-space there will be n such choice directions. We can think of these choice lines as forming an alternative coordinate system in the q-space.[43] Each point may be represented by a vector $(\xi_1, \ldots, \xi_n)$. The p_a and q_a may now be thought of as functions of $(\xi_1, \ldots, \xi_n)$. With this interpretation, it is appropriate to use partial derivative symbols. Since $q_a = (\partial q_a / \partial \xi_i) p_{b,a}$, we have

$$\theta_i \frac{\partial q_a}{\partial \xi_i} p_{b,a} q_b + \frac{\partial q_a}{\partial \xi_i}(m_a - t_a) = 0$$
$$\Rightarrow \sum_i \frac{\partial \xi_i}{\partial q_a} \theta_i \frac{\partial q_c}{\partial \xi_i} p_{b,c} q_b + m_a - t_a = 0,$$

where the first form of the equation was multiplied by the matrix inverse to $\partial q / \partial \xi$, that is, by the matrix with elements $\partial \xi_i / \partial q_a$.

[43] Of course there are many such coordinate systems, differing by (position-dependent) redefinitions of individual parameters $\xi_1, \ldots, \xi_n$. Here we simply consider one fixed coordinate system of this kind.

Denote by θ_{ij} the elements of a diagonal matrix that has $\theta_1, \ldots, \theta_n$ on its diagonal; that is, $\theta_{ij} \equiv \theta_i \delta_{ij}$, where δ_{ij} is the Kronecker delta, equal to one for $i = j$ and zero otherwise. Then

$$\frac{\partial \xi_d}{\partial q_a} \theta_{de} \frac{\partial q_c}{\partial \xi_e} p_{b,c} q_b + m_a - t_a = 0.$$

The matrix θ_{de} naturally "lives" in the coordinate system $(\xi_1, \ldots, \xi_n)$. But it can be transformed into the q-coordinate system in the q-space by a similarity transformation using the Jacobian matrix $\partial q_c / \partial \xi_e$. Let us denote this transformed matrix by

$$\tilde{\theta}_{ac} \equiv \frac{\partial \xi_d}{\partial q_a} \theta_{de} \frac{\partial q_c}{\partial \xi_e}.$$

Intuitively, the matrix $\tilde{\theta}_{ac}$ tells us how competitive firms whose altered choices could move us in a particular direction $(e_1, \ldots, e_n)$ in the output space (q-space) are:[44] if these firms are very competitive, then $e_a \tilde{\theta}_{ab} e_b$ is small, and vice versa. Taking into account that

$$\begin{aligned} \tilde{\theta}_{ac} &\equiv \frac{\partial \xi_d}{\partial q_a} \theta_{de} \frac{\partial q_c}{\partial \xi_e}, \\ p_{b,c} q_b &= -ms_c, \\ m_a &= p_a - mc_a, \end{aligned}$$

we obtain

$$\tilde{\theta}_{ac} p_{b,c} q_b + m_a - t_a = 0 \Rightarrow \tilde{\theta}_{ac} ms_c = p_a - mc_a - t_a.$$

This formula generalizes the "Lerner condition" $\theta = [(p - mc - t)/p]\epsilon_D$ used in Section IV.

B. The Pass-Through Matrix

In order to derive an expression for the (inverse) pass-through matrix, let us consider infinitesimal changes in economic variables induced by a small change $d\mathbf{t} = (dt_1, \ldots, dt_n)$ in taxes. Totally differentiating $-\tilde{\theta}_{ac} ms_c + p_a - mc_a - t_a = 0$ gives

$$-\sum_j d(\tilde{\theta}_{ij} ms_j) + dp_i - dmc_i - dt_i = 0,$$

which may be rewritten as

$$-\sum_k \sum_j (\tilde{\epsilon}_{\tilde{\theta} ijk} + \tilde{\epsilon}_{msjk}) \frac{\tilde{\theta}_{ij} ms_j}{q_k} dq_k + dp_i - \sum_k \tilde{\epsilon}_{Sik} \frac{mc_i}{q_k} dq_k - dt_i = 0.$$

[44] The vector $(e_1, \ldots, e_n)$ is assumed to be normalized to one, i.e., $e_a e_a = 1$.

We define the elasticities $\tilde{\epsilon}_{\tilde{\theta}ijk}$, $\tilde{\epsilon}_{msjk}$, and $\tilde{\epsilon}_{Sik}$ appearing in this equation and, for a future reference, the demand elasticity matrix ϵ_{Dkl} and its inverse $\tilde{\epsilon}_{Dkj}$ as[45]

$$\begin{aligned}
\tilde{\epsilon}_{\tilde{\theta}ijk} &\equiv \frac{q_k}{\tilde{\theta}_{ij}} \frac{\partial \tilde{\theta}_{ij}}{\partial q_k}, \\
\tilde{\epsilon}_{msjk} &\equiv \frac{q_k}{ms_j} \frac{\partial ms_j}{\partial q_k}, \\
\tilde{\epsilon}_{Sik} &\equiv \frac{q_k}{mc_i} \frac{\partial mc_i}{\partial q_k}, \\
\epsilon_{Dkl} &\equiv -\frac{p_l}{q_k} q_{k,l}, \\
\tilde{\epsilon}_{Dkj} &\equiv -\frac{q_j}{p_k} p_{k,j}.
\end{aligned}$$

Since

$$-\frac{dq_k}{q_k} = -\sum_l \frac{q_{k,l}}{q_k} dp_l = \sum_l \frac{\epsilon_{Dkl}}{p_l}$$

and $dp_i = \sum_l \delta_{il} dp_l$, we obtain

$$\sum_l \left\{ \delta_{il} + \left[\tilde{\epsilon}_{Sia} mc_i + \sum_j (\tilde{\epsilon}_{\tilde{\theta}ija} + \tilde{\epsilon}_{msja}) \tilde{\theta}_{ij} ms_j \right] \frac{\epsilon_{Dal}}{p_l} \right\} dp_l = dt_i.$$

This means that the li element of the matrix $\boldsymbol{\rho}^{-1}$ inverse to the pass-through matrix $\boldsymbol{\rho} \equiv d\mathbf{p}/d\mathbf{t}$ is[46]

$$(\boldsymbol{\rho}^{-1})_{li} = \delta_{il} + \left[\tilde{\epsilon}_{Sia} mc_i + \sum_j (\tilde{\epsilon}_{\tilde{\theta}ija} + \tilde{\epsilon}_{msja}) \tilde{\theta}_{ij} ms_j \right] \frac{\epsilon_{Dal}}{p_l}.$$

This general formula can be transformed into two alternative forms:

$$\begin{aligned}
(\boldsymbol{\rho}^{-1})_{li} &= \delta_{il} + \epsilon_{Dal} \tilde{\epsilon}_{Sia} \frac{p_i - t_i}{p_l} + \epsilon_{Dal} \sum_j \sum_k (\tilde{\epsilon}_{\tilde{\theta}ija} + \tilde{\epsilon}_{msja} - \tilde{\epsilon}_{Sia}) \tilde{\theta}_{ij} \tilde{\epsilon}_{Dkj} \frac{p_k q_k}{p_l q_j}, \\
(\boldsymbol{\rho}^{-1})_{li} &= \delta_{il} + \tilde{\epsilon}_{Sia} \epsilon_{Dal} \frac{mc_i}{p_l} + \sum_j \tilde{\theta}_{ij} (\tilde{\epsilon}_{\tilde{\theta}ija} + \tilde{\epsilon}_{msja}) \epsilon_{Dal} \tilde{\theta}_{jb}^{-1} \frac{p_b - mc_b - t_b}{p_l}.
\end{aligned}$$

[45] Note that the elasticities $\tilde{\epsilon}_{\tilde{\theta}ijk}$, $\tilde{\epsilon}_{msjk}$, and $\tilde{\epsilon}_{Sik}$ have inverse meaning relative to those (without a tilde) used in Sec. IV. Also, $\epsilon_{Dia}\tilde{\epsilon}_{Daj} = \delta_{ij}$ and $\tilde{\epsilon}_{Dia}\epsilon_{Daj} = \delta_{ij}$.

[46] Vectors such as $\mathbf{q}$ should be thought of as row vectors. For consistency, for a given matrix $\mathbf{A}$ we denote by A_{ij} the ij element of its transpose, i.e., $A_{ij} = (\mathbf{A}^T)_{ij}$. Note also that in our convention

$$(\boldsymbol{\rho})_{ij} = \left(\frac{d\mathbf{p}}{d\mathbf{t}} \right)_{ij} = \frac{\partial p_i}{\partial t_j}.$$

The first form was obtained by substituting for marginal cost from $mc_i = -\tilde{\theta}_{ic} ms_c + p_i - t_i$ and then for marginal consumer surplus from

$$ms_j = -\sum_k p_{k,j} q_k = \sum_k \tilde{\epsilon}_{Dkj} \frac{p_k q_k}{q^j}.$$

It represents a direct generalization of (2). In deriving the second form we used instead $ms_j = \tilde{\theta}^{-1}_{jb}(p_b - mc_b - t_b)$, where the inverse conduct parameter matrix $\tilde{\theta}^{-1}$ is assumed to exist.

References

Aguirre, Iñaki, Simon George Cowan, and John Vickers. 2010. "Monopoly Price Discrimination and Demand Curvature." *A.E.R.* 100 (4): 1601–15.

Anderson, Simon P., André de Palma, and Brent Keider. 2001. "Tax Incidence in Differentiated Product Oligopoly." *J. Public Econ.* 81 (2): 173–92.

Anderson, Simon P., André de Palma, and Jacques-François Thisse. 1992. *Discrete Choice Theory of Product Differentiation.* Cambridge, MA: MIT Press.

Anderson, Simon P., and Maxim Engers. 1992. "Stackelberg versus Cournot Oligopoly Equilibrium." *Internat. J. Indus. Org.* 10 (1): 127–35.

Anderson, Simon P., and Régis Renault. 2003. "Efficiency and Surplus Bounds in Cournot Competition." *J. Econ. Theory* 113 (2): 253–64.

Armstrong, Mark, and John Vickers. 2010. "A Model of Delegated Project Choice." *Econometrica* 78 (1): 213–44.

Armstrong, Mark, John Vickers, and Jidong Zhou. 2009. "Prominence and Consumer Search." *RAND J. Econ.* 40 (2): 209–33.

Armstrong, Mark, and Jidong Zhou. 2011. "Paying for Prominence." *Econ. J.* 121 (556): F368–F395.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *A.E.R.* 98 (5): 1998–2031.

Atkin, David, and Dave Donaldson. 2012. "Who's Getting Globalized? Intranational Trade Costs and World Price Pass-Through in South Asia and Sub-Saharan Africa." Manuscript, http://www.econ.yale.edu/%7eda334/.

Atkinson, Anthony B., and Joseph E. Stiglitz. 1980. *Lectures on Public Economics.* New York: McGraw-Hill.

Bagnoli, Mark, and Ted Bergstrom. 2005. "Log-Concave Probability and Its Applications." *Econ. Theory* 26 (2): 445–69.

Becker, Gary S. 1991. "A Note on Restaurant Pricing and Other Examples of Social Influences on Price." *J.P.E.* 99 (5): 1109–16.

Becker, Gary S., and Kevin M. Murphy. 1993. "A Simple Theory of Advertising as a Good or a Bad." *Q.J.E.* 108 (4): 941–64.

Borell, C. 1975. "Convex Set Functions in *d*-Space." *Periodica Mathematica Hungarica* 6 (2): 111–36.

Borenstein, Severin. 1988. "On the Efficiency of Competitive Markets for Operating Licenses." *Q.J.E.* 103 (2): 357–85.

Bowley, Arthur L. 1924. *The Mathematical Groundwork of Economics.* Oxford: Oxford Univ. Press.

Brander, James A., and Barbara J. Spencer. 1981. "Tariffs and the Extraction of Foreign Monopoly Rents under Potential Entry." *Canadian J. Econ.* 14 (3): 371–89.

Bresnahan, Timothy F. 1989. "Empirical Studies of Industries with Market Power." In *Handbook of Industrial Organization*, vol. 2, edited by Richard Schmalensee and Robert Willig, 1011–57. Amsterdam: North-Holland.

Bulow, Jeremy I., John D. Geanakoplos, and Paul D. Klemperer. 1985a. "Holding Idle Capacity to Deter Entry." *Econ. J.* 95 (377): 178–82.

———. 1985b. "Multimarket Oligopoly: Strategic Substitutes and Complements." *J.P.E.* 93 (3): 488–511.

Bulow, Jeremy I., and Paul Pfleiderer. 1983. "A Note on the Effect of Cost Changes on Prices." *J.P.E.* 91 (1): 182–85.

Chen, Yongmin, and Michael H. Riordan. 2007. "Price and Variety in the Spokes Model." *Econ. J.* 117 (522): 897–921.

———. 2008. "Price-Increasing Competition." *RAND J. Econ.* 39 (4): 1042–58.

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-Form Methods." *Ann. Rev. Econ.* 1:451–88.

Cournot, Antoine A. 1838. *Recherches sur les principes mathematiques de la theorie des richess.* Paris: Hachette.

Cowan, Simon. 2012. "Third-Degree Price Discrimination and Consumer Surplus." *J. Indus. Econ.* 60 (2): 333–45.

Dasgupta, Partha, and Eric Maskin. 2000. "Efficient Auctions." *Q.J.E.* 115 (2): 341–88.

Delipalla, Sofia, and Michael Keen. 1992. "The Comparison between Ad Valorem and Specific Taxation under Imperfect Competition." *J. Public Econ.* 49 (3): 351–67.

Dixit, Avinash K., and Victor Norman. 1978. "Advertising and Welfare." *Bell J. Econ.* 9 (1): 1–17.

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

Dupuit, Arsène Jules Étienne Juvénal. 1844. *De la mesure de l'utilité des travaux publics.* Paris.

Edelman, Benjamin, Michael Ostrovsky, and Michael Schwarz. 2007. "Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords." *A.E.R.* 97 (1): 242–59.

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

Farrell, Joseph. 2008. "Some Welfare Analytics of Aftermarkets." Manuscript, http://www.law.northwestern.edu/searlecenter/papers/Farrell_Analytics.pdf.

Farrell, Joseph, and Carl Shapiro. 2010a. "Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition." *B. E.J. Theoretical Econ.: Policy and Perspective* 10 (1).

———. 2010b. "Recapture, Pass-Through and Market Definition." *Antitrust Law J.* 76 (3): 585–604.

Fudenberg, Drew, and Jean Tirole. 1984. "The Fat-Cat Effect, the Puppy-Dog Plow, and the Lean and Hungry Look." *A.E.R.* 74 (2): 361–66.

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

Genesove, David, and Wallace P. Mullin. 1998. "Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890–1914." *RAND J. Econ.* 29 (2): 355–77.

Gomes, Renato. 2012. "Optimal Auction Design in Two-Sided Markets." Manuscript, http://sites.google.com/site/northwesternrenatogomes/Home/research.

Goos, Maarten, Patrick Van Cayseele, and Bert Willekens. 2012. "Platform Pricing in Matching Markets." Manuscript, http://www.econ.kuleuven.be/public/n06022/M2M-20120711-WP.pdf.

Holmes, Thomas J. 1989. "The Effects of Third-Degree Price Discrimination in Oligopoly." *A.E.R.* 79 (1): 244–50.

Holmström, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *J. Law, Econ., and Org.* 7 (special issue): 24–52.

Jaffe, Sonia, and E. Glen Weyl. 2010. "Linear Demand Systems Are Inconsistent with Discrete Choice." *B. E.J. Theoretical Econ. (Advances)* 10 (1): art. 52.

———. Forthcoming. "The First-Order Approach to Merger Analysis." *American Econ. J.: Microeconomics.*

Jehiel, Philippe, and Benny Moldovanu. 2001. "Efficient Design with Interdependent Valuations." *Econometrica* 69 (5): 1237–59.

Jenkin, Henry Charles Fleeming. 1871–72. "On the Principles Which Regulate the Incidence of Taxes." *Proc. Royal Soc. Edinburgh* 7:618–31.

Johnson, Harry G. 1953–54. "Optimum Tariffs and Retaliation." *Rev. Econ. Studies* 21 (2): 142–53.

Johnson, Justin P., and David P. Myatt. 2006. "On the Simple Economics of Advertising, Marketing, and Product Design." *A.E.R.* 96 (3): 756–84.

Kimmel, Sheldon. 1992. "Effects of Cost Changes on Oligopolists' Profits." *J. Indus. Econ.* 40:441–49.

Klemperer, Paul D., and Margaret A. Meyer. 1989. "Supply Function Equilibria in Oligopoly under Uncertainty." *Econometrica* 57 (6): 1243–77.

Laffont, Jean-Jacques, and Jean Tirole. 1993. *A Theory of Incentives in Procurement and Regulation.* Cambridge, MA: MIT Press.

Lerner, A. P. 1934. "The Concept of Monopoly and the Measurement of Monopoly Power." *Rev. Econ. Studies* 1 (3): 157–75.

Mankiw, N. Gregory, and Michael D. Whinston. 1986. "Free Entry and Social Inefficiency." *RAND J. Econ.* 17 (1): 48–58.

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

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

Miller, Nathan H., Marc Remer, and Gloria Sheu. 2013. "Using Cost Pass-Through to Calibrate Demand." *Econ. Letters* 118 (3): 451–54.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon. 2012. "A Reduced-Form Approach to Behavioral Public Finance." *Ann. Rev. Econ.* 4: 511–40.

Nocke, Volker, and Michael D. Whinston. Forthcoming. "Merger Policy with Merger Choice." *A.E.R.*

Ossa, Ralph. 2011. "A 'New Trade' Theory of GATT/WTO Negotiations." *J.P.E.* 119 (1): 122–52.

Perloff, Jeffrey M., and Steven C. Salop. 1985. "Equilibrium with Product Differentiation." *Rev. Econ. Studies* 52 (1): 107–20.

Pigou, Arthur C. 1920. *The Economics of Welfare.* London: Macmillan.

Prékopa, András. 1971. "Logarithmic Concave Measures with Application to Stochastic Programming." *Acta Scientiarum Mathematicarum (Szeged)* 32:301–16.

Quint, Daniel. 2013. "Imperfect Competition with Complements and Substitutes." Manuscript, http://www.ssc.wisc.edu/~dquint/papers/quint-complements-substitutes.pdf.

Reny, Philip J., Simon J. Wilkie, and Michael A. Williams. 2012. "Tax Incidence under Imperfect Competition: Comment." *Internat. J. Indus. Org.* 30 (5): 399–402.

Riley, John G., and William F. Samuelson. 1981. "Optimal Auctions." *A.E.R.* 71 (3): 381–92.
Robinson, Joan. 1933. *The Economics of Imperfect Competition.* London: Macmillan.
Rochet, Jean-Charles, and Lars A. Stole. 2003. "The Economics of Multidimensional Screening." In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, vol. 1, edited by Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, 150–97. Cambridge: Cambridge Univ. Press.
Rochet, Jean-Charles, and Jean Tirole. 2003. "Platform Competition in Two-Sided Markets." *J. European Econ. Assoc.* 1 (4): 990–1029.
Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Rev. Econ. Studies* 68 (1): 205–29.
Schmalensee, Richard. 1981. "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination." *A.E.R.* 71 (1): 242–47.
Seade, Jesus. 1985. "Profitable Cost Increases and the Shifting of Taxation: Equilibrium Response of Markets in Oligopoly." Manuscript, http://ideas.repec.org/p/wrk/warwec/260.html.
Shaked, Avner, and John Sutton. 1982. "Relaxing Price Competition through Product Differentiation." *Rev. Econ. Studies* 49 (1): 3–13.
Shapiro, Carl. 1996. "Mergers with Differentiated Products." *Antitrust* 10:23–30.
Spence, A. Michael. 1975. "Monopoly, Quality, and Regulation." *Bell J. Econ.* 6 (2): 417–29.
Spengler, Joseph J. 1950. "Vertical Integration and Antitrust Policy." *J.P.E.* 58 (4): 347–52.
Telser, Lester G. 1972. *Competition, Collusion and Game Theory.* Chicago: Aldine-Atherton.
Varian, Hal R. 1985. "Price Discrimination and Social Welfare." *A.E.R.* 75 (4): 870–75.
Veiga, André, and E. Glen Weyl. 2012. "Multidimensional Product Design." Manuscript, http://ssrn.com/abstract=1935912.
Villas-Boas, Sofia Berto, and Rebecca Hellerstein. 2006. "Identification of Supply Models of Retailer and Manufacturer Oligopoly Pricing." *Econ. Letters* 90 (1): 132–40.
von Stackelberg, Heinrich F. 1934. *Marktform und Gleichgwicht.* Vienna: Julius Springer.
Werden, Gregory J. 1996. "A Robust Test for Consumer Welfare Enhancing Mergers among Sellers of Differentiated Products." *J. Indus. Econ.* 44 (4): 409–13.
Weyl, E. Glen. 2009a. "Monopoly, Ramsey and Lindahl in Rochet and Tirole (2003)." *Econ. Letters* 103 (2): 99–100.
———. 2009b. "The Price Theory of Two-Sided Markets." Manuscript, http://ssrn.com/abstract=1324317.
Weyl, E. Glen, and Jean Tirole. 2012. "Market Power Screens Willingness-to-Pay." *Q.J.E.* 127 (4): 1971–2003.