Slutsky meets Marschak*

The First-Order Identification of Multi-product Production

E. Glen Weyl†

December 2009

Abstract

Marschak (1953) suggested that applied research should begin by determining the minimal set of assumptions and data needed to make a prediction of interest. Standard identification analyses correspond poorly to this search, as they have either stringent data (local average treatment effects/IV and non-parametric) or structure (parametric) requirements. Yet, in this spirit, much empirical (Chetty, 2009) and some theoretical work has asked which effects of policy interventions may be forecast from the local observable levels and derivatives of the structure. I formalize this inquiry as the *first-order* identification problem and illustrate it with the case of multi-product production. Under perfect competition or monopoly, but not standard oligopoly, Slutsky conditions for firm optimization provide strategies for identifying, for example, the effects of price controls. They also test for consistency of conjectural variations.

This manuscript is preliminary. Please do not cite it without the author's permissions, but please do send along any suggestions for improvement.

*I acknowledge the generous support of the Milton Fund, which funded excellent research assistance by Stephanie Lo, and the helpful comments of Jon Baker, Severin Borenstein Gary Chamberlain, Raj Chetty, Jerry Hausman, Ali Hortaçsu, Guido Imbens, David Laibson and David I. Levine. I am particularly grateful to Jim Heckman, a conversation with whom largely inspired the research that led to this paper. All errors and misunderstandings are, of course, my own.

†Society of Fellows, Harvard University, 78 Mount Auburn Street, Cambridge, MA 02138: weyl@fas.harvard.edu.

In a series of recent papers, Jim Heckman, Sergio Urzúa and Ed Vytlacil (Heckman and Vytlacil, 2007; Heckman and Urzúa, Forthcoming) argue that "estimators should be selected on the basis of their ability to answer...economic problems with minimal assumptions" and data. They refer to this principle as Marschak's Maxim in honor of Jacob Marschak's 1953 argument that the nature of "useful knowledge", and thus the assumptions and data needed to obtain it, depends critically on the policy question of interest. This principle appears to contrast sharply with the two dominant paradigms in empirical identification. On one hand, a growing literature on instrumental variables (IV) and local average treatment effects (LATEs) focuses on (the limited set[1] of) interventions for which experiments can be run or quasi-experiments observed. On the other, a modern "structural" literature estimates broad swathes of deep economic structure, such as an entire multi-product demand system, in order to forecast the effects of interventions. This typically requires either very large data sets or restrictive functional form assumptions. Marschak's Maxim, which seems to bridge these approaches, thus lacks a formal language.

Yet Marschak's approach is commonly applied. In fact, the session of the ASSA meetings for which this paper is being prepared is devoted to such work. Chetty (2009)'s paper (Subsections I.A-B) surveys a wide range of research, such as Einav et al. (Forthcoming)'s work (Subsection I.B), which exploits basic economic structure with few ancillary functional form assumptions to connect plausibly observable LATEs to welfare effects. My own prior work (Weyl and Fabinger, 2009) shows that, with a minimum of structural restrictions, observations of pass-through rates identify many comparative statics in industrial organization models (Subsection I.C). Yet without a formal language for such identification results, I fear they may not receive the econometric support that will help them thrive.

This paper therefore formalizes[2] (Subsection III.D) the *first-order* identification problem which lies behind such analysis. This asks whether, after observing *local* levels and derivatives of endogenous outcomes with respect to some exogenous variable one can recover the derivatives of (observable or unobservable) outcomes of interest with respect to that exogenous variable or others for which exogenous variation may be unavailable. Thus, unlike LATEs (Subsection III.A), first-order identification requires neither that the outcome of interest be observable nor that the intervention of interest vary in the data, making it applicable to a wider range of economic problems. Unlike parametric identification (Subsection III.B), first-order identification requires no parametric structure on the relationship between exogenous and endogenous variables, allowing it to treat general economic theories which rarely

[1]See, for example, Heckman and Vytlacil (2005) for a discussion of the limits on questions that can be directly answered by IV.

[2]This may be useful for several reasons summarized in the first paragraph of my conclusion (Section VI).

supply specific functional forms. Unlike nonparametric identification (Subsection III.C), the first-order approach can formalize limitations on data by reducing the number and degree of derivatives observed by the econometrician: small data sets are sufficient to identify first-order derivatives, larger ones second-order derivatives, etc. While each of the standard approaches might be extended (Subsection III.E) to include first-order identification, the necessary extension is awkward and thus it is useful to think of it as distinct.

I illustrate this general problem with an application to a classic model in my field of industrial organization - that of multi-product producers. I draw on Slutsky (1915)'s[3] argument that the matrix of cross partial derivatives of compensated demand must be symmetric. This result that has gone somewhat out of favor in recent years, as compensated demand for consumers is often difficult to observe, but which applies well to firms which are typically assumed to maximize profits free of budget constraints.

Imagine trying to predict the effect of a price control on one of two goods produced by a single firm on production and welfare (Subsection IV.A). Direct IV is likely infeasible, as price controls are rarely randomly imposed. A structural model could easily inadvertently sign the effect purely through ex-ante assumptions (Subsection IV.B), while a non-parametric model would require more data than is likely to be available. First-order identification offers a simple solution (Subsection IV.C). If the first-order effect of exogenous shocks to the cost of producing the two goods on price can be observed, Slutsky symmetry and consumer envelope conditions tie down all effects of interest under perfect competition, monopoly and consistent conjectures oligopoly (Bresnahan, 1981), but *not* under the standard static oligopoly solution concepts. These results apply generally (Subsection IV.D) to firms producing any number of products, across time and states of the world. Beyond symmetry, Slutsky's concavity conditions restrict the magnitude of cross-product production pass-through (Subsection IV.E). The results can be applied to a variety of settings of recent interest (Subsection IV.F) that may not immediate appear to be classic multi-product production, such as multi-sided platforms and demand accumulation/learning-by-doing models.

This application illustrates another benefit of first-order identification (Subsection V.A). The techniques (classical price theory) needed to solve the problem in most applications are simpler and more broadly known within economics than those typically used for parametric and non-parametric identification. In addition to making the analysts' job easier and more

[3]The use of Slutsky's ideas to illustrate my Marschakian approach repays an old intellectual debt. Slutsky's seminal work was published in an obscure Italian journal; however, Jacob Marschak, before immigrating to Britain and then United States, had attended Slutsky's lectures at the Institute of Economics in Kiev. Chipman and Lenfant (2002) suggest Marschak may have been responsible for the eventual reception of Slutsky as a father of modern demand theory. If true, it would be particularly fitting for an example from Slutsky assist in the revival of Marschak's approach to identification.

broadly accessible, this may help to make econometric identification transparent to a broader audience, aiding the credibility of empirical results based on this identification approach. Of course these advantages have corresponding drawbacks (Subsection V.B). First-order identification requires theoretical assumptions that IV strips away. It narrows the usefulness of any given analysis relative to either parametric or non-parametric identification. Where the structure imposed by parametric identification helps test underlying, broadly applied economic assumptions, first-order identification can lead to false generalization that shields theory from falsification. Thus first-order identification is most useful as a complement, not a substitute, for the alternative approaches. I conclude in Section VI by discussing directions, both empirical and theoretical, that future research along these lines might take. Proofs, and other supplements, are collected into appendices available at http://www.glenweyl.com.

I. Examples from Prior Work

Despite the lack of formalism, first-order identification is a strategy commonly applied in empirical research. On occasion what is essentially a first-order identification result has been the focus of theoretical work. In this section, to motivate the formalism below, I briefly considering a few examples. Chetty (2009) provides a far more complete survey of my the first two application areas, and I therefore keep these especially short.

A. Taxation

Perhaps the most famous first-order identification argument is Harberger (1964)'s formula[4] for the deadweight loss of taxation. Under a number of strong, but standard, assumptions[5] the marginal deadweight loss from taxing an activity is $t\frac{dA}{dt}$ where A is the quantity of the activity and t is the tax rate. This influential formula connects potentially observable behavioral effects of taxation to their (directly) unobservable welfare effect. The classic applications were by Harberger (1954, 1962) himself and a prominent recent revival is Feldstein (1999)'s calculation of the marginal deadweight loss of income taxes.

To calculate the variation (equivalent of a derivative) of a non-linear tax schedule requires more information. Saez (2001) built on Diamond (1998)'s analysis of the Mirrlees (1971) model to show that, jointly with a set of marginal welfare weights across the income distribution, the distribution of income and its elasticity with respect to marginal tax rates determines the variation. Saez (2001) and Gruber and Saez (2002) use assumptions about

[4]This is due originally to Dupuit (1844) and Jenkin (1870, 1871–1872).

[5]Including complete markets and optimization, no externalities, perfect competition and no income effects.

welfare weights and empirical measurements of the distribution of earnings and elasticities[6] to calculate, with increasing numbers of ancillary assumptions, number unobservable quantitites: 1) the marginal welfare gains from a local change in the tax schedule, 2) the optimal tax rate on the highest income taxpayers and 3) the full optimal tax schedule.

B. Insurance

Baily (1978) showed that, in a static model, the difference between the marginal utility of consumption of the employed and unemployed (the marginal rate of substitution or MRS) and the marginal rate of transformation of income between the two states identify the first-order welfare gains from unemployment insurance. Chetty (2006) showed this applies to a broad class of dynamic models of insurance and Gruber (1997), Chetty (2008) and Shimer and Werning (2007) provide different empirical strategies for calculating the necessary MRS.

Einav et al. (Forthcoming) study distortions in private insurance markets, observing that the net marginal social benefit of expanding coverage is the divergence between inverse demand and the cost of serving a marginal consumer. Competitive markets are inefficient under adverse selection because they instead equate inverse demand to the cost of serving the *average* consumer. Local variation in price identifies the slope of average costs, revealing marginal costs and therefore the marginal social benefits of expanding coverage.

C. Industrial organization

Much of the first wave of the new empirical industrial organization (IO) relied on a first-order approach. Perhaps the foundational concept of that literature (Rosse, 1970) was Lerner (1934)'s argument that demand elasticity identifies, through firm first-order conditions, the marginal cost of production. Following this lead, Baker and Bresnahan (1985) famously showed that a first-order approximation to the gains from collusion can be estimated from the elasticities of the residual two-product demand system for the colluding firms without estimating the entire industry demand system. Panzar and Rosse (1987) showed that monopoly behavior could generally be tested based on input demand elasticities obeying certain inequalities. Bresnahan and Reiss (1991) argued that the curvature of the number of firms with respect to the size of markets identifies the effect of entry on margins.

More recently this approach has been largely confined to theoretical work. Unlike in public finance, the goal is typically not to tie the unobservable welfare effect of policy changes

[6]Nathanson and Weyl (2010) extend this to allow workers to choose between well-compensated rent-seeking and enjoyable virtuous careers, requiring an additional measure of the elasticity of career substitution.

to their observable consequences, but rather to tie the potentially observable consequence of uncommon policy changes to the effects of shocks that are more likely observed in the data.

A leading example is Froeb et al. (2005) and Farrell and Shapiro (2008) argument that a first-order approximation to the price effects of mergers are determined by pass-through rates and own- and cross-elasticities *for the merging products only.* Similar results using pass-through have been derived by Aguirre et al. (Forthcoming) on third-degree monopoly price discrimination, by Bulow and Klemperer (2009) on the effects of price controls on consumer welfare in a competitive market and by Weyl (2008, 2009) on mergers in two-sided markets. Gopinath and Itskhoki (Forthcoming) show that firm exchange-rate pass-through rates identifies the frequency of price changes and confirm this empirically.

Weyl and Fabinger (2009) survey and extend these results, showing that in many IO models many qualitative effect can be signed by the pass-through rate or its slope. For example in multi-product industries strategic complements versus substitutes is identified by how pass-through compares to unity and the ratio of either consumer surplus or deadweight loss to monopoly profits are identified by pass-through.

D. Other examples

Examples of first-order identification-like exercises show up occasionally in other areas of economics. Examples include Acemoglu (2009)'s argument that complementarity between technology and labor determines the effect of labor shortages on technical progress and Lee and Saez (2008) identification optimal minimum wages from their marginal effects on employment. More broadly the tradition in applied microeconomics[7] that tests the predictions of theories, typically comparative statics, rather than the theories themselves, can be seen as (very simple) manifestations of the notion of first-order identification I formalize below.

Given the prevalence of first-order identification type reasoning in both theoretical and empirical work, it is surprising that it has no natural analog in the literature on formal identification. This is analogous to the position of IV in the early 1990s, prior to the path-breaking work of Imbens and Angrist (1994). In what follows I therefore show how providing a formal framework helps systematize the application of first-order identification.

[7]Macroeconomics has similar results. Aguiar and Hurst (2005) and Card et al. (2007) demonstrate that models of dynamic consumer optimization can be distinguished using only a few moments of behavior, such as the drop in consumption upon retirement or the sensitivity of consumption to unexpected wealth shocks.

	Endongenous	**Exogeneous**
Varying	Observable: LHS Variable **Y** / Unobservable: Welfare **W**	Observable: Covariates **X** / Unobservable: Latent Variables **E**
Unvarying		Observable: Policy **U**

II. A General Model

Suppose that, as in the discussion above, we are interested in the (average) effects[8] of change in some policy instrument p on some outcome o. In Marschak's classic example of a monopolist, p is either a production level of a monopolist or a tax on that monopolist, while o is either the firm's profits or the revenue raised by the tax. The economic structure is made up of a number of variables (potentially infinite) that can be broken into five categories:

1. Endogenous observables **Y**: these are observable outcomes (like quantity demanded). There could be infinitely many of these, as, for example, in the Saez (2001) model, the number of taxpayers at each continuous income level is the key set of observables.

2. Endogenous unobservables **W**: these are outcomes which cannot be observed directly (like consumer welfare in Marschak's industry). These will often be the welfare of some individuals which can only be measured through revealed preferences.

3. Exogenous, varying observables **X** : these are exogenous inputs to the economic structure, often called "covariates", which help determine outcomes. In Marschak's setting this might be the firm's cost, past production or past tax rates. They are exogenous in the sense that some independent variation in them can be isolated (from unobservable exogenous variables) either through direct randomization or with an instrument and thus the effect of varying them can be isolated, at least locally.

4. Exogenous, varying unobservables **E**: these exogenous inputs to the economic structure cannot be directly observed and thus are a source of noise for the econometrician. They are often referred to as "latent variables". In Marschak's example they might include unobservable demand shifters. As discussed above, the sense in which **X** is exogenous is that some variation in **X** orthogonal to **E** can be isolated. Because one can always

[8] In labor econoimcs, this known (Heckman and Vytlacil, 2002) as the "policy relevant treatment effect".

relabel any part of $\mathbf{X}$ that is not orthogonal $\mathbf{E}$ as an endogenous, observable outcome, I assume that $\mathbf{X} \perp \mathbf{E}$ for expositional simplicity[9].

5. Exogenous, unvarying observables $\mathbf{U}$: these exogenous inputs can be directly observed, but do not (exogenously) vary in the data we observe in the past. A natural example would be whether two firms have merged or not: it is clearly observed, but prior to a merger taking place, will never have varied (for those firms) in the past.

These variables are related by the *structural equations*

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}, \mathbf{U}, \mathbf{E}) \tag{1}$$

$$\mathbf{W} = \mathbf{g}(\mathbf{X}, \mathbf{U}, \mathbf{E}) \tag{2}$$

where $\mathbf{f}$ and $\mathbf{g}$ are some vector-valued functions. The outcome $o \in \mathbf{W} \cup \mathbf{Y}$ and the policy instrument $p \in \mathbf{U} \cup \mathbf{X}$. Given that $\mathbf{f}$ and $\mathbf{g}$ are arbitrary functions we can assume, without loss of generality (Matzkin, 2003) any fixed distribution of $\mathbf{E}$; all the information about the "true distribution of errors" is encoded in the structural functions. Thus if we knew the structural functions $\mathbf{f}$ and $\mathbf{g}$ we could determine the effect of a change in p on the distribution of o. However we typically will not know these functions precisely, instead having to infer at least some properties of them from observations of $\mathbf{X}$ and $\mathbf{Y}$.

A common feature of such data is that $\mathbf{X}$ varies only in a limited range about some "equilibrium" quantity $\mathbf{X}^\star$ or perhaps all data varies around a very few equilibrium values of $(\mathbf{X}^\star_1, \ldots, \mathbf{X}^\star_M)$ where M is small and particularly $M \ll N$. The notion of "$\mathbf{X}$ varying in a limited range about these values" will be made a bit more precise in the following section.

III. Identification Problems

Identification analysis (Koopmans, 1949) asks what variation in data is necessary for estimation to be possible in principle. In this section I use the model above to provide a common language for comparing the classic approaches to identification (LATEs, parametric and non-parametric). I then formally define and contrast first-order identification.

[9]A more general assumption that would suffice, as discussed below, is that each entry of $\mathbf{X}$ is orthogonal to $\mathbf{E}$ conditional on the other entries of $\mathbf{X}$. This level of detail is distracting in what follow as I have little or no discussion of actual statistical techniques or conditions. If this were to be made more precise, exogenous variables could be altered (view them as residual variation) so that they are satisfy the simple condition.

A. Local average treatment effects (LATEs)

The literature on LATEs considers condition under which the effects of (small) changes in an observable exogenous variables $p \in \mathbf{X}$ on the average value of an observable endogenous variable $o \in \mathbf{Y}$ can be observed. Formally letting

$$\mathbf{F}(\mathbf{X}, \mathbf{U}) \equiv E_{\mathbf{E}} \left[\mathbf{f}(\mathbf{X}, \mathbf{U}, \mathbf{E}) \right]$$

$$\mathbf{G}(\mathbf{X}, \mathbf{U}) \equiv E_{\mathbf{E}} \left[\mathbf{g}(\mathbf{X}, \mathbf{U}, \mathbf{E}) \right]$$

In the case when p is a continuous treatment[10], the LATE of $p \in \mathbf{X}$ on $o \in \mathbf{Y}$ at $(\mathbf{X}, \mathbf{U}) = (\mathbf{X}^\star, \mathbf{U}^\star)$ is $\left. \frac{\partial F_o}{\partial X_p} \right|_{\mathbf{X}=\mathbf{X}^\star, \mathbf{U}=\mathbf{U}^\star}$ where F_o is the entry of $\mathbf{F}$ corresponding to o and X_p is the entry of $\mathbf{X}$ corresponding to p. If p is instead discrete and, without loss of generality, we assume p comes in integer increments, the LATE of p on o at $(\mathbf{X}, \mathbf{U}) = (\mathbf{X}^\star, \mathbf{U}^\star)$ is $F_o \left(X_p^\star + 1, \mathbf{X}_{-p}^\star, \mathbf{U}^\star \right) - F_o \left(\mathbf{X}^\star, \mathbf{U}^\star \right)$ where $\mathbf{X}_{-p}^\star$ is the entries of $\mathbf{X}^\star$ other than p.

Orthogonality between $\mathbf{X}$ and $\mathbf{E}$ obviates the identification problem for LATEs (Imbens and Angrist, 1994): so long as $p \in \mathbf{X}$ and $o \in \mathbf{Y}$, the LATE of p on o should be easy to estimate at the equilibrium value $\mathbf{X}^\star$ and the invariant level of $\mathbf{U}^\star$. A large literature has developed[11] finds weaker statistical conditions for identifying LATEs (Imbens, 2004) and practical approaches to find the necessary quasi-experimental variation[12] (Angrist and Krueger, 2001) or, recently, to generate experimental variation[13] (Harrison and List, 2004).

However, in many cases quasi-experimental variation in past data does not exist and experiments are infeasible for ethical or practical reasons. This leads to p being[14] in $\mathbf{U}$. Even if quasi-experimental variation is available, the outcome of interest o will often be in $\mathbf{W}$. In

[10] Heckman and Vytlacil (1999) call this "local instrumental variables" and extensively discuss its connection to the standard discrete LATE estimator.

[11] This literature strongly complements the first-order approaches, as LATEs are its building blocks. Another focus has been the extent to which not just average effects but the distribution of over covariates (only one value of $\mathbf{X}$ is held fixed and other are allowed to vary) or latent variables can be observed. See Manski (1990, 1996, 2003) for a conservative approach based on bounding these quantities. However concerns about identifying LATEs are largely orthogonal to those about how to use models to map these to unobservable outcomes or the effects of unvarying policy instruments, so I do not discuss these literatures further.

[12] This tradition dates back at least to Schultz (1964)'s study of the agricultural production function through natural disasters. Prominent recent examples include Angrist (1990)'s study of the effects of the Vietnam draft, Levitt (1997)'s study of the effect of policing on crime and Hoxby (2000) analysis of the effect of school competition on performance. In all of these cases, variation (argued to be exogenous) in the policy instrument of interest exists in past data and the outcome of interest is observable.

[13] Actual experiments are a more recent development in this literature. Important contributions include List (2003)'s study of the endowment effect, Resnick et al. (2006)'s analysis of reputation in eBay auctions, Banerjee et al. (2007)'s study of educational programs in India and on-going large scale studies by Roland Fryer and co-authors of various educational interventions in inner-city schools in the United States.

[14] Note that for the LATE to be valid the instrument observed to vary must be exactly the policy instrument of interest. If it is just something "close" (the same treatment, but on a different subpopulation, for example), further assumptions are needed to transform it into something policy-relevant (Heckman and Vytlacil, 2002).

either of these situations, without further assumptions, the LATEs cannot be estimated. Thus the LATE approach can be applied if, and only if, plausible exogenous variation in p can be found or produced and the outcome of interest o is directly observable. This limits the direct purview of empirical methods based on this approach and is the primary motivation (Reiss and Wolak, 2007) behind so-called "structural methods" of identification, described below, that impose assumptions on $\mathbf{f}$ and/or $\mathbf{g}$ to augment the informational content of data.

B. Parametric identification

As famously argued by Lucas (1976), and earlier by Marschak (1953) and Tinbergen (1956), when past data do not include variations in the policy instruments of interest, structural assumption must be imposed to infer from data the likely effects of these interventions. Similarly, if an outcome of interest, such as welfare, cannot be directly observed, economic theory is needed to translate effects on observables into effects on unobservables.

One natural and common approach to this problem is to assume a theoretical structure that limits uncertainty to a small number of parameters[15]. Consider Marschak's simple example of a monopolist. Suppose one assumes a linear demand curve with noise $Q = a(1+\epsilon^Q)(b-P)$ where ϵ^Q is a mean-zero error term, and that some cost shifter x is observable. The monopolist's constant marginal cost is a function $cx + t + \epsilon^c$, parameterized by c, of some observable cost shifter x and a specific tax t which has, in the past, remained fixed at $t \equiv 0$. Observing the prices and quantities following random fluctuations in x, independent of ϵ^Q and ϵ^c, identifies the parameters a, b and, because the monopolist's price under linear demand is $\frac{b+cx+t+\epsilon^c}{2}$ also reveal c once b is observed. Furthermore Jevons (1871)'s formula[16] tells us that consumer surplus is $a(1+\epsilon^Q)\frac{(b-P)^2}{2}$ and thus surplus as a function of cost and the parameters is $a(1+\epsilon^Q)\frac{(b-cx-t-\epsilon^c)^2}{8}$. Thus, even though the tax rate never varies in the data and welfare can never be directly observed, the effect of a cost shock on welfare can be inferred once the parameters of the assumed structural model have been identified.

The above example demonstrates the power of parametric identification if the functional form is plausible; however, it also reveals its dangers. In this context, the assumption of linear demand implies, without any reference to data, that prices will increase half as much

[15]The attempt to estimate, typically parametric, models of the economy is perhaps the oldest tradition in empirical economics and it is thus difficult to precisely trace its origins. However rigorous modern approach to parametric identification and most modern empirical applications of it originate with the Cowles Comission and its leaders, including Ragnar Frisch, Grhard Tinter, Trygve Haavelmo, Lawrence Klein and, especially, Jacob Marschak and T. C. Koopmans. Koopmans (1949) was the first to formalize the identification problem as such. Koopmans et al. (1950) conducted one of the first formal identification analyses in this tradition, studying systems of linear structural equations. Probably the most famous early application of this approach was to macroeconomics, particularly Klein (1950)'s model of the aggregate economy.

[16]See James R. Hines (1999) for a detailed history of the measurement of consumer welfare.

as the tax does. Thus if a subsidy of a dollar is given to consumers, prices rise by fifty cents (by the neutrality of tax incidence) and thus consumer welfare net of the cost of the subsidy must fall. This result is not a general property of monopoly, as we will see in Subsection D, and is instead an artifact of the assumed functional form.

More generally, parametric identification assumes that $\mathbf{f}$ and/or $\mathbf{g}$ take a particular, known functional form. All uncertainty is collected into a finite group of parameters. Formally this is to assume that $\mathbf{f}(\mathbf{X}, \mathbf{U}, \mathbf{E}) \equiv \hat{\mathbf{f}}(\mathbf{X}, \mathbf{U}, \mathbf{E}; \gamma)$ and $\mathbf{g}(\mathbf{X}, \mathbf{U}, \mathbf{E}) \equiv \hat{\mathbf{g}}(\mathbf{X}, \mathbf{U}, \mathbf{E}; \gamma)$ where $\hat{\mathbf{f}}, \hat{\mathbf{g}}$ are known functions and γ is a finite vectors of unknown parameters. Parametric identification[17] asks whether, under certain observability assumptions, a parameter or some collection of parameters are restricted, to a set or a point. A typical observability condition would be the econometrician seeing the distribution of $\mathbf{f}(\mathbf{X}, \mathbf{U}^{\star}, \mathbf{E})$ over realizations of for all $\mathbf{E}$ for any given value of the exogenous observables $\mathbf{X}$.

Parametric identification has been a tremendous success[18] in a wide range of fields, no doubt due to the fact that the conditions necessary for parametric identification are much weaker than those for LATEs. As indicated by the monopoly example, it does not require either that p have varied in past data or that o be observable. Instead, all that is needed for an identification analysis to yield a "positive" result, thereby allowing empirical estimation, is that observable variation ties down the parameters of the models[19]. This predictive power comes at the cost of functional form assumptions, concerns about the plausibility of which has motivated interest in the last decade and a half on non-parametric approaches.

C. Non-parametric identification

Non-parametric identification[20] has the same basic goals as parametric identification, but seeks to avoid functional forms, instead relying only on the sort of "economic" assumptions

[17]Following Koopmans and other Cowles economists, formal analysis of parametric identification has flourished. I will mention just one example, in the spirit of the monopoly problem above. Bresnahan (1982) famously argued that if both demand "twisters" (elasticity shifters) and "shifters" are observable, then not only can demand and cost be identified, but so too can the conjectures of firms about how changes in their quantities will be met by changes in their competitors' quantities. Nevo (1998) extends this reasoning to differentiated product industries and Villas-Boas and Hellerstein (2006) does so for vertical relations.

[18]Famous recent applications include Heckman (1979)'s model of selection into labor market treatments, Deaton and Muellbauer (1980)'s work on consumption patterns, the dynamic stochastic general equilibrium framework that is at the core of modern macroeconomics beginning with Kydland and Prescott (1982), Wolpin (1987)'s research on employment search, Rust (1987)'s study of real options and, perhaps most influential at the moment, the demand-based analysis of industrial equilibrium by Berry et al. (1995).

[19]Partial identification, pioneered by Manski (1995), focuses on cases where not all parameters can be identified. Nonetheless most formal identification analysis, even recently, focuses on full identification.

[20]The exact origins of the non-parametric approach to identifying econometric models is less clear than those for LATEs and parametric identification, but the approach appears to have gained significant credence over the course of the 1980's (Matzkin, 2007) led by the pioneering work of Brown (1983) and Roehrig (1988).

typically used in abstract theoretical analysis, such as optimization, concavity, etc.

Returning to the monopoly example, a non-parametric approach would simply write $Q = Q(P, \epsilon^Q)$ where Q is an unknown smooth, decreasing-in-P_i function, ϵ^Q is known to the firm but not the econometrician and everything else is as before. Price is still a function of constant marginal[21] cost if demand is known, but the expression is a bit more abstract: P is the (generically unique) maximizer of $(P - c[x, \epsilon^c] - t)\, Q(P, \epsilon^Q)$ over all values of P. t is again assumed fixed at 0 and c is an unknown, assumed-increasing function of x, again assumed orthogonal to $\epsilon^Q \perp \epsilon^c$ and I assume[22] that $Q_2 > 0$ at all P[23].

The variation in x, inducing variation in P, which traces out the distribution of Q at each P also traces out the distribution of P as a function of $c(x, \epsilon^c) + t$ through firm optimization and thus allows the distribution of c as a function of x to be recovered as well. Thus it allows the (distribution of) price effects of a change from 0 in t to be observed. Similarly consumer surplus is, for any P and ϵ^Q, by Jevons's rule[24], $\int_{x=P}^{\infty} Q(x, \epsilon^Q) dx$; thus the price variation also traces out the distribution of consumer welfare, as a function of the price[25]. Therefore observation of the demand curve, through price variation, implies observation of (the distribution of) welfare as a function of the tax, even though cost is never seen to vary nor is welfare observed at all in past data. Thus at least in this simple example, none of the powerful identification imposed on this problem by economic theory is lost by removing the parametric structure. This approach allows, for example, the impact of taxes on the tax-augmented consumer welfare discussed above to be estimated rather than assumed.

More generally, non-parametric identification asks whether, after having observed the distribution of $\mathbf{f}(\mathbf{X}, \mathbf{U}^\star, \mathbf{E})$ over realizations of $\mathbf{E}$ a function of $\mathbf{X}$ (perhaps over some set) one also observes[26], say, $\mathbf{f}$ as a function of $\mathbf{U}$ or $\mathbf{g}$ as a function of $\mathbf{X}$ or $\mathbf{U}$. Matzkin (2007) provides a some what more detailed and rigorous definition along these lines.

While nonparametric identification[27] has gained substantial traction as a theoretical econometric problem, its direct application to empirical practice has been somewhat lim-

[21] This itself is a sort of parametric restriction that could be relaxed, but here this would be confusing.

[22] Some assumption of this form is needed for the "quantiles of the distribution of Q" as a function of P to reveal the quantiles of "Q as a function of P". Some weaker assumption would almost certainly suffice.

[23] This implies that observing the distribution of prices and the distribution of welfare as a function of prices reveals a distribution of welfare. Again this can likely be weakened.

[24] Hotelling (1932)'s lemma is a common generalized version of Jevons's formula. See Milgrom and Segal (2002) for a recent further generalization of this result.

[25] See Hausman and Newey (1995) for an application of this non-parametric approach when utility is not quasi-linear. Willig (1976) argues such income corrections exceed statistical uncertainty.

[26] If $\mathbf{f}$ and/or $\mathbf{g}$ are not point identified they may still be restricted by these observations to some class of functions, the natural analog here of set identification.

[27] For example, Matzkin (1993)'s study of choice models, Athey and Haile (2002)'s analysis of the identification of value distributions in auctions and, most relevant here, Berry and Haile (2009a,b)'s results on the nonparametric identification of demand systems through price and choice set variation.

ited[28]. An important reason is that, in any finite samples, many non-parametric methods do not substantially differ from a particular parametric estimate that (at that data size) they happen to correspond to[29]. Thus, despite the evidently greater flexibility of nonparametric identification results, when data is limited their direct benefits relative to parametric methods for particular empirical applications are often much less obvious.

D. First-order identification

A natural goal, then, is to maintain the flexibility of the non-parametric approach while allowing for small data sets to be formalized. This motivates the first-order approach to identification. While I believe that at least the precise formalization here is novel, the basic approach clearly has a wide range of both theoretical and empirical precedents (Section I).

First-order identification begins with a focus, as with LATEs, on the effects of a small change[30] in p from its "current" or "equilibrium" level on o. In the monopoly example, the government may be considering placing a small tax or subsidy on the monopolist and wish to know the effect this has on welfare. The goal is to non-parametrically determine this effect, with limited data. In particular, prices in the past have only been induced by cost shocks to exogenously vary a small amount and all of this variation has been "near" the current equilibrium price $P^\star$ induced by some equilibrium cost shifter $x^\star$.

Perhaps the simplest way to formalize this data limitation non-parametrically is to abandon stochastics and simply to assume that the level and some derivatives of the (deterministic) structure can be observed. In the monopoly example, the local value of P, Q, P', Q', P'' and Q'' might be assumed to be observable. Note that I have suppressed error term[31]. This has a couple logically consistent interpretations, discussed in more detail in Appendix I.A:

1. Most simply, we could suppose that like, x, the error term varies only over a small range, or appropriately separable in the structural functions, and thus value of the derivatives of the structure are constant over values of the error term. Then the non-stochastic model represents uniform local properties of $\mathbf{f}, \mathbf{g}$.

[28] Prominent exceptions include consumer and producer theory (Varian, 1982, 1983, 1984, 1985), macroeconomics (Campbell and Mankiw, 1987), auction behavior (Guerre et al., 2000) and finance (Ait-Sahalia and Lo, 1996). Yatchew (1998) provides a survey of this impressive, if relatively small, literature.

[29] The sieve non-parametric estimator (Grenander, 1981; Geman and Hwang, 1982) makes this clearest.

[30] This is essentially what Carneiro et al. (Forthcoming) call the "marginal policy-relevant treatment effect".

[31] The primary goal of suppressing the stochastic element, as discussed in the following section, is to simplify the problem in a manner that leaves out much of the statistical detail, while retaining the essence of the economic argument linking certain observations to others. The statistical detail can then be filled in at a later, separate stage of the analysis; see Chesher (2003) and Hastie and Loader (1993) for approaches to the nonparametric identification and estimation of derivatives.

2. Another, more broadly applicable assumption is that non-stochastic model represents the average values of the stochastic model. Observing the average value of derivatives of the stochastic model is equivalent to observing the derivatives of this average. Thus if it is average values of derivatives that are of interest, linking such values in the in a non-stochastic model provides a reduced-form[32] strategy for linking average values in a stochastic model. In this case the non-stochastic model represents **F** and **G**.

In Marschak's monopoly example, classical price theory tells us that, based on the observables posited above, quite a bit can be said with only weak additional assumptions. First consider the behavior of the monopolist. Suppose that her constant marginal cost cannot be directly observed. Her first-order condition for maximization is

$$P - c(x) - t = -\frac{Q(P)}{Q'(P)} \tag{3}$$

Thus observing $P^\star, Q^\star$ and Q' identifies the marginal cost (Lerner, 1934; Rosse, 1970). Price theory also links how the monopolists responds to such an increase to the local shape of demand. Assuming marginal revenue is a declining function of quantity, implicit differentiation of equation (7) gives an expression for the monopolist's local *pass-through* of an increase in cost that was first derived by Cournot (1838):

$$\frac{dP}{dc} = \frac{1}{2 - \frac{QQ''}{(Q')^2}} \tag{4}$$

Hence, identification of cost pass-through requires observing the second derivative (log-curvature) of demand at $P^\star$, which typically requires more data than observing its first derivative (elasticity). However, when this can be observed, tax pass-through, and $c'(x^\star)$, can be predicted without any functional form restrictions on demand.

Finally consider the measure of welfare discussed above, aggregating consumers and the government. By Jevons's rule, perhaps the most classic of first-order identification results, the effect on consumer welfare is just the change in price multiplied by demand $\frac{Q}{2-\frac{QQ''}{(Q')^2}}$. Because the government also earns revenue the total harm is $Q\left(\frac{1}{2-\frac{QQ''}{(Q')^2}} - 1\right)$. Thus, from this anti-firm perspective, a tax (subsidy) will be first-order beneficial if and only if $QQ'' <$ $(>)\,(Q')^2$. Given that the right hand side is always positive, linear demand will impose that a tax is attractive; it can easily be shown that constant elasticity demand favors a subsidy.

Thus, just as with both structural approaches, the effect of a change in a variable (the

[32] See Appendix I.A for caveats.

tax rate) that has never been observed to vary on a variable that is not observed at all (welfare) can be determined based on observable variation by invoking economic theory. As with nonparametric identification, this argument does not rely on any specific functional form assumptions and, in fact, the analysis can be carried out with little if any reference to the statistical details that would be needed for actual estimation, instead relying on abstract economic arguments about the nature of the structural functions.

More generally first-order identification focuses considers $\mathbf{F}$ and $\mathbf{G}$ and takes as observable a collection of levels and derivatives (essentially LATEs) of $\mathbf{F}$ at various points $(\mathbf{X}_1^\star, \dots, \mathbf{X}_M^\star)$ given a fixed level of $\mathbf{U}^\star$. For example the econometrician might observe all of the first and second-own-partial derivatives of $\mathbf{F}$ with respect to $\mathbf{X}$ at a single point $\mathbf{X}^\star$, but none of the cross partial derivatives. I now provide formal definitions of first-order identification, for this case of a single equilibrium point $\mathbf{X}^\star$ and finite dimensional variables. Appendix I.B extends this definition to the case of many equilibrium points and infinite dimensional variables. Let $|\mathbf{X}|$ represent the dimensionality of $\mathbf{X}$ and so forth for other variables.

Definition 1. *The* equilibrium levels *of the structure* $\mathbf{L}$ *is a* $|\mathbf{X}| + |\mathbf{U}| + |\mathbf{Y}|$*-dimensional real vector* $(\mathbf{X}^\star, \mathbf{U}^\star, \mathbf{F}[\mathbf{X}^\star, \mathbf{U}^\star])^\top$.

In the simple monopoly example, the equilibrium levels are just $(x^\star, 0, P^\star, q^\star)$.

Definition 2. *Let* i *be a natural number. The* i-th order varying derivatives*, denoted by* $\mathbf{D}^i$*, of the structure is a* $|\mathbf{Y}| \begin{pmatrix} |\mathbf{X}| + i - 1 \\ i \end{pmatrix}$*-dimensional real column vector.* $\mathbf{D}^i$ *represents the* i*th order partial derivatives*[33] *of entries of* $\mathbf{Y}$ *with respect to entries of* $\mathbf{X}$.

These are just the (potentially observable) derivatives or a given order i of observable outcomes with respect to all of the varying exogenous variables. The dimensionality of this vector is defined by the number of cross-partial derivatives of order i possible with $|\mathbf{X}|$ input variables and $|\mathbf{Y}|$ output variables; this is the same as $|\mathbf{Y}|$ times the number of ith order monomials in $|\mathbf{X}|$ variables. For the simple monopoly case, $\mathbf{D}^i = \left(\frac{d^iP}{dx^i}\Big|_{x=x^\star}, \frac{d^iQ}{dx^i}\Big|_{x=x^\star} \right)$, the ith derivative of demand evaluated at the equilibrium price.

Definition 3. *Let* N *be a natural number. The* up-to N-th order varying derivatives *of the structure* $\mathbf{D}^{\Sigma_i^N} = \left(\mathbf{L}^\top, \mathbf{D}^{1^\top}, \mathbf{D}^{2^\top}, \dots, \mathbf{D}^{N^\top} \right)^\top$.

In the simple monopoly examples this is

$$\left(x^\star, t^\star, P^\star, q^\star, \frac{dP}{dx}\bigg|_{x=x^\star}, \frac{dQ}{dx}\bigg|_{x=x^\star}, \frac{d^2P}{dx^2}\bigg|_{x=x^\star}, \frac{d^2Q}{dx^2}\bigg|_{x=x^\star}, \dots, \frac{d^NP}{dx^N}\bigg|_{x=x^\star}, \frac{d^NQ}{dx^N}\bigg|_{x=x^\star} \right) \tag{5}$$

[33] For concreteness and without loss of generality, assume these are listed in lexicographic order first in the ordering of $\mathbf{Y}$ and then in the monomial lexicographic ordering with respect to the entries of $\mathbf{X}$.

Definition 4. *An* N-th order point observability condition*,* $\Theta_{\cdot}^{N}$*, is a vector of functions* $\theta_{\cdot}^{N}$ *of* $\mathbf{D}^{\Sigma_i^N}$ *to the reals with the property that for any real number* x $\theta_{\cdot}^{N}=x$ *is satisfied for some, but not all, values of* $\mathbf{D}^{\Sigma_i^N}$*. This has the interpretation that the econometrician knows the exact value of all fucntions* $\theta_{\cdot}^{N}$ *(the relevant functions are point observable). An* N-th order set observability condition*,* $\Theta_{>}^{N}$*, is a vector of functions* $\theta_{>}^{N}$ *of* $\mathbf{D}^{\Sigma_i^N}$ *to the reals with the property that* $\theta_{>}^{N}>0$ *is satisfied for some, but not all, values of* $\mathbf{D}^{\Sigma_i^N}$*. This has the interpretation that the econometrician knows that the inequality* $\theta_{>}^{N}>0$ *holds (the relevant function of derivatives and levels is set observable). Finally an* N-th order observability condition *is a pair* $\Theta^N=\left(\Theta_{=}^{N},\Theta_{>}^{N}\right)$ *of* N*-th order point and set observability conditions.*

A simple second-order point observability condition in the monopoly example would be that one can all levels and first derivatives (elasticities) but not second derivatives. If we refer to the sevent entries of $\mathbf{D}^{\Sigma_i^2}$ as $(x^\star,t^\star,P^\star,q^\star,P_x,Q_x,P_{xx},Q_{xx})$ then this would be formalized by $\Theta_{=}^{2}=(x^\star,t^\star,P^\star,q^\star,P_x,Q_x)$. Knowing the pass-through rate's comparison to unity is

$$\left(Q_{xx}P_x-P_{xx}Q_x\right)Q-P_x\left(Q_x\right)^2$$

or its negative being the unique element of $\Theta_{>}^{2}$, as the above expression can easily be shown to have the same sign as pass-through less 1 (see the proof of Proposition 1 below). Together these might form Θ^2, a second-order observability condition.

Definition 5. *The* structural restriction *is a collection of functions* S *of* $(\mathbf{X},\mathbf{U})$ *to* $\mathbb{R}^{|\mathbf{Y}|+|\mathbf{W}|}$ *representing all structural functions consistent with the economic assumptions imposed.*

In the monopoly example, $\mathbf{F}(x,t)$ is $\left(P\left[c(x)+t\right],Q\left[P\left(c[x]+t\right)\right]\right)$ and if we, for the moment, assume that $\mathbf{G}$ is just the one dimensional, firm-neglecting welfare, $\mathbf{G}=W\left(P\left[c(x)+t\right]\right)+tQ\left(P\left[c(x)+t\right]\right)$ where $W(y)\equiv\int_y^\infty Q(y)dy$. S is the set of all smooth functions from $\mathbb{R}^2$ to $\mathbb{R}^3$ representable in the form above where Q is smooth and strictly decreasing wherever positive and has $\frac{Q''Q}{(Q')^2}<1$ everywhere (declining marginal revenue, c is strictly increasing and smooth and $P\left(c[x]+t\right)$ maximizes $\left(P-c[x]-t\right)Q(P)$ given x and t.

Definition 6. *A property of* $(\mathbf{F},\mathbf{G})$ *at some point* $\left(\tilde{\mathbf{X}},\tilde{\mathbf{U}}\right)$ *is said to be* point identified *given an* N*-th order observability condition* Θ^N *at point* $(\mathbf{X}^\star,\mathbf{U}^\star)$ *and a structural restriction* S *if* $\forall\mathbf{x}\in\mathbb{R}^{\left|\Theta_{\cdot}^{N}\right|}$ *every function* $(\mathbf{F},\mathbf{G})\in S$ *satisfying* $\Theta_{\cdot}^{N}=\mathbf{x}$ *and* $\Theta_{>}^{N}>0$ *at that point assigns the same value to the property of interest (as a function of* $\mathbf{x}$*). A property is said to be* sign identified *if, over the same range of settings, the property is always of the same sign (positive, negative or zero). The property will be said to be* point unidentified *or* sign unidentified *whenever they are not point identified or sign identified respectively.*

While the properties of interest are typically local levels or derivatives, in general they could be any property of the function. For example, first-order identification may also attempt to determine the effects of discrete changes[34] in p on o. Returning again to the monopoly problem, it may be instructive to state a formal result, taking for the moment $\Theta^2_>$ to be of the positive sign version proposed above, $p = t$ and $o = W + tQ$.

Proposition 1. *Given* Θ^2 *and* S, $\left.\frac{do}{dp}\right|_{x=x^\star,t=0}$ *is sign identified but point unidentified.*

Proof. The simple and illustrative proof is, for brevity, included in Appendix I.C.

Finally first-order identification mirrors other approach in its notion of model testing/over-identification. The simplest version of that exercise, deriving a signed comparative static based on the predictions of a set of economic assumptions, is a result in applied theory.

Definition 7. *An N-th order observability condition* Θ^N *at point* $(\mathbf{X}^\star, \mathbf{U}^\star)$ over-identifies *a structural restriction* S *if* $\exists \mathbf{x} \in \mathbb{R}^{|\Theta^N_.|} : \exists(\mathbf{F}, \mathbf{G}) \in S$ *satisfying* $\Theta^N_. = \mathbf{x}$ *and* $\Theta^N_> > 0$ *at that point* and $\exists \mathbf{x}' \neq \mathbf{x} : \nexists(\mathbf{F}, \mathbf{G}) \in S$ *satisfying* $\Theta^N_. = \mathbf{x}$ *and* $\Theta^N_> > 0$ *at that point.*

In the simple monopoly model if the derivative of price with respect to taxes were to become observable, this would over-identify the economic assumptions of the model.

E. Connections

Perhaps the clearest connection between first-order identification and the other approaches is its potential to be seen as the non-parametric identification of certain properties of a function. In fact, some of the most interesting recent work (Matzkin, 2003; Chesher, 2003) in nonparametric identification has focused on conditions for the identification of derivatives of the economic structure. It would then seem natural to view first-order identification as a simple subset of nonparametric identification. A model could be assumed where the Chesher-Matzkin conditions for nonparametric identification of some, assumed observable, derivatives are satisfied; economic structure could then be imposed; finally their conditions could be checked for the identification of the derivative of interest.

Such an approach is likely isomorphic to a first-order identification analysis, but nonetheless, there are several reasons why it may be useful for first-order identification to be semantically distinct. First, such an exercise seems a cumbersome way to analyze first-order identification: the requisite statistical baggage is likely to confuse rather than clarify the

[34] Werden (1996), Froeb et al. (2005) and Farrell and Shapiro (2008) tie the effect of a merger to so-called diversion ratios, mark-ups and pass-through rates. However these exercises are typically either based on first-order approximations to this discrete effect based on some natural "smooth path" from the pre-intervention to post-intervention states or only attempt to sign, rather than quantify, the effects in question.

analysis. Second, the literature on nonparametric identification of local derivatives has typically focused on derivatives of low order; however, there is little reason why it *need* do so. In principle, the same infinite local variation that identifies local first derivatives identifies local derivatives of arbitrarily high orders. If the structural functions are assumed analytic, observing all local derivatives eventually reveals the whole function. It is precisely this counter-intuitive conclusion that first-order identification tries to avoid by assuming that only derivatives of certain orders are observable. Thus first-order identification focuses on imposing *quantitative* limits on the data available through observability assumptions.

Viewing first-order identification as a parametric identification problem might, at first, seem similarly natural. One can almost always parameterize a model such that one parameter corresponds to each of the local derivatives or levels with which first-order identification seeks to observe. A classic example of this approach was Deaton and Muellbauer (1980)'s demand system, which is explicitly flexible on all elasticities and cross-elasticities of demand. One could then ask whether these parameters are identified in that particular model, under observability conditions that only reveal certain derivatives.

In practice, however, this approach is essentially never used for several reasons. First, formulating observability conditions that identify parameters corresponding to certain derivatives is formally cumbersome in the standard framework. Second, formulating logically consistent functional forms that are explicitly flexible along the desired dimensions, but not others, is a substantial task and would need to performed anew for each set of observability and economic assumptions, as well as each outcome of interest. Perhaps worse, the analytic techniques needed to establish identification could vary widely across these different functional forms, making analysis challenging. This, in turn, would make a difficult the main purpose of first-order identification: the examination of how identification varies across sets of economic and observability assumptions. Thus, while any first-order identification problem likely *can* be reformulated as a parametric, the typical goals of such analyses are distinct and the reformulation is unlikely to be useful in practice.

LATEs are more clearly distinct from first-order identification. Still, the processing of data to even determine what exactly has varied in the past, what is observable, what is exogenous to what and so forth nearly always requires some number of assumptions. Could not the process of first-order identification analysis simply be viewed as an extension, adding assumptions to change the interpretation of data until it reveals the desired effect?

The essential difference is that LATE assumptions typically aspire to be immediately persuasive to at least the majority of the economics community. First-order identification, on the other hand, is typically based on quite strong economic assumptions at a remove from the identification they imply. While first-order identification is obviously more successful when these assumptions are more plausible, the primary goal is primarily to make the assumptions, and therefore the source of identification, transparent through the use of formal economic models. These economic assumptions are unlikely to be perfectly correct, so formal statement and analysis of them is crucial to understanding the plausibility of their application.

Between any of these categories there are many shades, but most applications fall fairly clearly into one camp or another. I hope the example of the following section continues to clarify why I consider it useful to think of first-order identification as a distinct approach

IV. Slutsky and Multi-Product Producers

In this section I consider a detailed example of a novel comprehensive first-order identification analysis of the standard model of multi-product firms, building on Slutsky (1915)'s theory of demand. This example has three purposes. First, it illustrates (Subsections A-C) in a bit more detail the practical distinctions between the identification approaches described in the previous section. Second, it shows how a first-order identification analysis can be carried out in a fairly systematic way, comparing various observability and economic assumptions and illustrating the varying identification each lends to this application (Subsections D-E). Finally, it may be of some substantive interest (Subsection F).

A. An example

I begin with a simple, classical example. A company chooses how much, $\mathbf{Q} = (Q_A, Q_B)^\top$, to produce of two consumption goods, A and B. The consumption goods fetch prices $\mathbf{P} = (P_A, P_B)^\top$, in a manner elaborated below, and have given smooth, strictly convex costs of production $C(\mathbf{Q}) + \mathbf{c}^\top\mathbf{Q}$, where $\mathbf{c} = (c_A, c_B)^\top$ are the prices of one-to-one necessary input to production that the producer takes as given. Thus she earns profits $\left(\mathbf{P}^\top - \mathbf{c}^\top\right)\mathbf{Q} - C(\mathbf{Q})$.

Consider three ways in which the good's price may be determined:

1. Perfect competition: The firm takes price of each good as given at $\mathbf{P}$.

2. Monopoly: The firm chooses her quantities $\mathbf{Q}$ therefore prices of the goods according to $\mathbf{P}(\mathbf{Q})$, a smooth inverse demand system obeying standard conditions.[35].

3. Oligopoly: The firm chooses quantities $\mathbf{Q}$ taking as given alla Cournot the production of other firms $\overline{\mathbf{Q}}(\mathbf{Q})$ of other firms which together determine prices according to $\mathbf{P}(\mathbf{Q}+\overline{\mathbf{Q}})$. $\overline{\mathbf{Q}}$ is determined by an equilibrium among firms. As in the standard Cournot model,

[35] I assume the law of demand ($\nabla\mathbf{P}$ has negative diagonal elements), derivation from the optimization of a representative consumer ($\nabla\mathbf{P}$ is symmetric) and concave revenues $\mathbf{P}^\top\mathbf{Q}$ in $\mathbf{Q}$.

$\overline{\mathbf{Q}}$ depends on the conditions facing the firm of interest only through her production: $\overline{\mathbf{Q}}(\mathbf{Q})$ is the equilibrium production of all other firms given that the firm in interest produces $\mathbf{Q}$. Thus equilibrium prices with production $\mathbf{Q}$ are $\mathbf{P}\left(\mathbf{Q}+\overline{\mathbf{Q}}\left[\mathbf{Q}\right]\right)$.

The econometrician observes prices $\mathbf{P}$, cost of the necessary inputs $\mathbf{c}$, quantities produced $\mathbf{Q}$, total cost $C(\mathbf{Q})$ and, in the oligopoly case, production of other firms $\overline{\mathbf{Q}}$ all at equilibrium. She also observes the level of a government-imposed price control on good A, $\overline{P^A}$, set initially at the equilibrium P^A so that it does not bind.

Let me translate this into the model of above. Exogenous variables are $\mathbf{c}$, $\overline{P^A}$ and, in the case of perfect competition, P_A and P_B. How these are partitioned between $\mathbf{X}$ and $\mathbf{U}$ will vary depending on the observability assumptions I make. $\mathbf{Y}$, observable endogenous variables, are quantities $\mathbf{Q}$, prices $\mathbf{P}$ in the imperfectly competitive models and quantities of other firms $\overline{\mathbf{Q}}$ in the oligopoly model. Unobservable consumer welfare W is, given my assumption of quasi-linearity and therefore integrability (Hotelling, 1938), derived from the envelope theorem: 0 under perfect competition, $\int_{\mathbf{q}=(0,0)}^{\mathbf{Q}} \mathbf{P}(\mathbf{q})^\top d\mathbf{q} - \mathbf{P}^\top\mathbf{Q}$ under monopoly and under oligopoly $\int_{\mathbf{q}=(0,0)}^{\mathbf{Q}} \mathbf{P}\left(\mathbf{q}+\overline{\mathbf{Q}}[\mathbf{q}]\right)^\top \left(\mathbf{I}_{2\times 2}+\nabla\overline{\mathbf{Q}}[\mathbf{q}]\right) d\mathbf{q} - \mathbf{P}^\top\mathbf{Q}$ where $\mathbf{I}$ is 2 by 2 identity matrix. The path along which the integrals are taken is irrelevant, given integrability.

The structural functions derive from firm maximization (and therefore from its unknown cost function) and the (unknown) inverse demand system. Under perfect competition the firm's first-order conditions are to equate price to marginal cost:

$$\mathbf{P} = (\nabla C)^\top(\mathbf{Q}) + \mathbf{c} \tag{6}$$

Because of convex costs, these equations have at most a unique solution[36]. The structural functions are then, in the case of perfect competition, that the quantities $\mathbf{Q} = (\nabla C)^{-1}(\mathbf{P}-\mathbf{c})$. Under monopoly, the firm equates marginal revenue and cost:

$$\mathbf{P}(\mathbf{Q}) + \mathbf{Q}^\top\nabla\mathbf{P}(\mathbf{Q}) = \nabla C(\mathbf{Q}) + \mathbf{c} \tag{7}$$

Again concave revenues, combined with convex costs, ensure this has at most one solution and again I assume it does have a solution. Let $\mathbf{M}(\mathbf{Q}) \equiv \mathbf{P}(\mathbf{Q}) + \mathbf{Q}^\top\nabla\mathbf{P}(\mathbf{Q}) - \nabla C(\mathbf{Q})$. Then here the structural functions are that $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{c})$ and that $\mathbf{P} = \mathbf{P}\left(\mathbf{M}^{-1}[\mathbf{c}]\right)$.

Finally, in the oligopoly case, the firm again acts like a monopolist, but her marginal revenue is affected by the production of other firms.

$$\mathbf{P}\left(\mathbf{Q}+\overline{\mathbf{Q}}\left[\mathbf{Q}\right]\right) + \mathbf{Q}^\top\nabla\mathbf{P}\left(\mathbf{Q}+\overline{\mathbf{Q}}\left[\mathbf{Q}\right]\right) = \nabla C(\mathbf{Q}) + \mathbf{c} \tag{8}$$

[36]For simplicity assume it does have a solution for any prices and costs. In other words, C satisfies the Inada (1963) conditions that $\frac{\partial C}{\partial Q_I} \to \infty$ as $Q_I \to \infty$ and $\frac{\partial C}{\partial Q_I} \to -\infty$ as $Q_I \to 0$.

The conditions required for this to have a unique solution are discussed in Appendix II.A, and I assume them here. Under these letting $\mathbf{O}(\mathbf{Q}) \equiv \mathbf{P}\left(\mathbf{Q}+\overline{\mathbf{Q}}\left[\mathbf{Q}\right]\right) + \mathbf{Q}^{\top}\nabla\mathbf{P}\left(\mathbf{Q}+\overline{\mathbf{Q}}\left[\mathbf{Q}\right]\right) - \nabla C(\mathbf{Q})$, the structural functions are that $\mathbf{Q} = \mathbf{O}^{-1}(\mathbf{c})$ and that $\mathbf{P} = \mathbf{P}\left(\mathbf{O}^{-1}[\mathbf{c}]\right)$.

B. LATE, parametric and non-parametric approaches

Suppose that a policy maker is considering imposing a "small" price control on the firm of interest's sale of good A produced. If the current equilibrium price of good A is $P_A^{\star}$ the policy maker is considering requiring that the good be priced at $\overline{P}_A \equiv P_A - \epsilon$ where ϵ is a small positive number. She would like an economist to predict how this will affect the production of both goods A and B, the price of good B and, perhaps, consumer welfare.

LATEs would almost certainly look for incidents in the past, in that industry or similar ones, where a price control had been imposed. If as is likely in this example, no data persuasively fitting this bill could be found the LATE approach would have little to offer. Furthermore, even in the best case it would not speak to W.

Parametric identification would place some structure on the problem, though what exactly is hard to predict. A simple example might be to assume that the cost system is Cobb-Douglas $C(Q_A, Q_B) = \eta Q_A^{\theta} Q_B^{\phi}$ with $\theta, \phi > 1$, in the monopoly case the price system is linear, $P_I(Q_I, Q_{-I}) = \alpha_I - \beta_I Q_I + \nu_I Q_{-I}$, and the oligopoly reaction functions could be derived from the demand system by assuming a number of competing firms N, using theory of Cournot oligopoly. It would then look for exogenous variation in the costs of the two goods. It is simple to show (see Appendix II.B) that variation in the cost of either of the goods is sufficient to identify all parameters. The optimization problem of the firm (or market equilibrium) could then be re-solved with the price control imposed and, assuming efficient rationing in the competitive case[37], all quantities of interest could be calculated.

Unfortunately this more practicable tack may inadvertently impose[38], rather than measure, many of the quantities of interest. The Cobb-Douglass form implies that the goods are substitutes in production. The linear demand imposes a connection between demand elasticity and pass-through of cost changes, cross-product marginal revenue effects and cross-price elasticities and immediately imposes a sign on cross-firm strategic effects (Bulow et al., 1985).

Non-parametric identification would, then, seem more attractive. Cost and prices could be left non-parametrically specified and it might shown that broad variation in both costs (see Appendix II.B) identifies the cost and price systems non-parametrically. It could also be shown that knowledge of these are sufficient to predict the effect of a price control and thus

[37] Additional distortions induced by inefficient rationing are analyzed by Bulow and Klemperer (2009).

[38] I have been miserly with the parametric approach; functional forms could almost certainly be imposed that avoided these specific restrictions. Subsection V.A discusses why this is not as unfair as it may seem.

the effect of the price control on the outcomes of interest are non-parametrically identified.

Yet in practice what procedure would the non-parametric approach advocate? Suppose we used a sieve. For any finite data set, then, a collection of parametric functions would be estimated and these might impose exactly the same unattractive restrictions one was concerned about in the parametric case. Any claim that the analysis avoided unattractive functional form restrictions would require an explicit analysis of the particular functional form used for that data size, just as in the parametric case. While the policy maker might be reassured that, if further data arrives prior to her decision the analysis might improve in ways the parametric one would not, at the actual point of decision she would be just as "stuck" with the imposed restrictions as in the parametric case.

Thus, at least superficially, it appears that in practice, the only approach to prediction available to the policy maker is some form of admittedly troubling parametric identification.

C. First-order approach

First-order identification offers a simple alternative. This is simplest to see for perfect competition. There a price control is just a reduction in the price which, in turn, is equivalent from the firm's perspective to a rise in the cost of producing A. Thus observing $\frac{\partial Q_A}{\partial c_A}$ and $\frac{\partial Q_B}{\partial c_A}$, $\frac{\partial Q_A}{\partial P_A}$ and $\frac{\partial Q_B}{\partial P_A}$ are identified. Social welfare is unaffected by the individual firm's shifts in production and by the envelope theorem the firm's profits fall, and consumer welfare rises, proportional to Q_A multiplied by the amount of the price control, assuming efficient rationing of the price controlled good. Thus by observing the first-order derivatives of observables with respect to costs, we can predict all the effects of (a small) price control.

Perhaps more surprising is that first-order identification is nearly as potent in the case of monopoly, where identification arises from the nature of the firm's optimization. First, consider welfare in market A. The price control, by the envelope theorem, increases consumer welfare by Q_A multiplied by its size and *has no first-order effect on profits* by the envelope theorem. The price control also affects the firm's optimal price on good B and therefore consumer welfare in that market. How it does so can be determined if we know the effect of the price control on production in both markets and the effects of these on prices.

We can view the monopolist as choosing any two of P_A, P_B, Q_A and Q_B, with the other two being determined by the demand system[39]. Suppose that we think of the monopolist as choosing P_A and Q_A and demand tying down Q_B and P_B. By construction, c_A is the price of Q_A. Because the firm has no budget constraint, it must *exactly* obey Slutsky (1915)'s symmetry condition, namely that $\frac{\partial P_A}{\partial c_A} = \frac{\partial Q_A}{\partial \lambda_A}$ where λ_A is the opportunity cost[40] of P_A. A

[39]Except when the goods' demand are independent.

[40]Some economists refer to this as the *shadow cost*.

price control must raise the opportunity cost of P_A in order to persuade the firm to reduce its price, as in the standard Lagrangian analysis. Thus the sign of $\frac{\partial Q_A}{\partial \overline{P_A}}$ is the opposite of that of $\frac{\partial P_A}{\partial c_A}$. Thus, a price control will increase the production of good A if and only if an increase in the cost of producing A increases the (optimally chosen) price of A. While it may seem obvious that this effect should be positive, Edgeworth in 1897 argued[41], in work published as Edgeworth (1925), that it is not immediate.

Because this argument could be applied to any choice of a two item subset of $\{P_A, P_B, Q_A, Q_B\}$, the effect of the price control on the production of good B can be determined by observing a shock to the cost of producing good B. Furthermore if we observe a cost shock to the production *both goods* we can even determine quantitatively the effects of a price control on good A on the production of either good, as these reveal the effect the price control must have opportunity cost to be effective. Thus local observations of cost shocks under monopoly determine the full effects of a local price control.

Matters differ under oligopoly. In standard static oligopoly models (e.g. Cournot), firms' conjectures about the effect of changing their production are not consistent with the actual equilibrium effect of those actions. In the simple linear demand duopoly Cournot case, firms act as if they believe a reduction in their production will leave the other firms' production unchanged, when in fact, in equilibrium, the competitor replaces half of the lost production. Thus an oligopoly producer with such *inconsistent conjectures* (Bowley, 1924; Bresnahan, 1981) need not obey Slutsky symmetry because an increase in the production of good A in fact affects the (marginal) revenue from producing good B in a way not anticipated[42] by the firm in determining its (marginal) revenue from good A. If the firms were playing consistent conjectures equilibrium[43] the monopoly reasoning would proceed with only minor changes.

Note that all of the above reasoning could be applied to cross-production effects directly. In fact, shocks to the cost of production of one good on the production of the other must be *quantitatively* the same in both directions under perfect competition, monopoly and consistent conjectures oligopoly *but not with inconsistent conjectures*. Thus observing $\frac{\partial Q_B}{\partial c_A}$ reveals $\frac{\partial Q_A}{\partial c_B}$ under these models[44]. If both can be observed, their (in)equality can be used to test whether inconsistent conjectures strategic interactions are significant. Also note that one of the commodities discussed above could be an investment rather than a commodity; in this case, the commonly queried effect of a price control on investment can be determined by

[41]This argument has been revived recently by Salinger (1991) and Kind et al. (2008) in the contexts respectively of vertical mergers and two-sided markets

[42]Of course there do exist demand systems for which the conjecture of no impact is actually consistent. For example for negative exponential inverse demand this is true globally.

[43]Dockner (1992) provides a game theoretic formal justification of such an equilibrium.

[44]The identifying power of consistent conjectures was first studied by Baker and Bresnahan (1985).

observing the effect of a shock to the cost of investing on the to-be-controlled price.

D. General results: symmetry

I now derive general formal version of the special, informal discussion in the previous subsection. Consider a firm producing N products $\mathbf{Q} = (Q_1, \dots Q_N)^\top$ which fetch prices $\mathbf{P} = (P_1, \dots, P_N)^\top$ and cost the firm $C(\mathbf{Q})+\sum_{i=1}^N c_i(Q_i, x_i)$ where C and c_i are smooth for all i and $\frac{\partial^2 c_i}{\partial Q_i \partial x_i} > 0$. I do not impose further restrictions, as modern techniques (Milgrom and Segal, 2002) allows these to be dispensed with. The firm's profits are $\mathbf{P}^\top \mathbf{Q} - C(\mathbf{Q}) - \sum_{i=1}^N c_i(Q_i, x_i)$.

The different solution concepts discussed above are:

1. Perfect competition: the firm takes as given $\mathbf{P}$.
2. Monopoly: the firm takes as given a smooth inverse demand system $\mathbf{P}(\mathbf{Q})$.
3. Oligopoly: the firm takes as given a residual inverse demand system which depends on the strategies of other firms $\boldsymbol{\sigma} \in \mathbb{R}^K$ for some natural K and is smooth in both: $\mathbf{P}(\mathbf{Q}, \boldsymbol{\sigma})$. For expositional simplicity[45] suppose that $\boldsymbol{\sigma}$ can be written as a function (in equilibrium) of $\mathbf{Q}$, as in the Cournot case. I will say the firm *plays static Nash* if she maximizes profits taking $\boldsymbol{\sigma}$ as given, *has consistent conjectures* if she maximizes taking into account that $\boldsymbol{\sigma}$ is the (correct equilibrium) function of $\mathbf{Q}$ and I will say *she has inconsistent conjectures* if she maximizes assuming $\boldsymbol{\sigma}$ is actually another function $\boldsymbol{\sigma}' \neq \boldsymbol{\sigma}$ of $\mathbf{Q}$. Note that static Nash is typically inconsistent.

In all cases the outcomes derive from the firm optimizing production[46] over all weakly positive production bundles $\mathbf{Q}$ under her beliefs about the effect of her production on profits.

In the language of the general model of Subsection III.D, exogenous variables are $\mathbf{x}$ and $\mathbf{P}$ for perfect competition or $\overline{\mathbf{P}}$, price control levels, for the imperfectly competitive models. I assume that $\mathbf{X} = \mathbf{x}$ while $\mathbf{U} = \mathbf{P}$ or $\overline{\mathbf{P}}$ as appropriate. Not all $\mathbf{x}$ vary, but I wish to allow this possibility and disallow it for price (control) shifts. Endogenous variables are $\mathbf{Q}$, $\mathbf{c}$, $\mathbf{P}$ under imperfect competition and $\boldsymbol{\sigma}$ under oligopoly. I assume all of these jointly make up $\mathbf{Y}$; that is their levels, and potentially derivatives, are observable.

[45] This assumption is without loss of generality, as if the other firms' production were a function of the firms' price, I could reformulate the problem in terms of the firm choosing prices, given a residual direct demand. This could be done for any choice variable determining the other firms' behavior which is, given a particular behavior, in one-to-one correspondence with quantity (Klemperer and Meyer, 1989).

[46] In the (nongeneric) case when there are many optimal production schedules, I assume, to tie down a unique outcome, that the firm chooses the largest one, lexicographically ordered

The structural functions linking these arise from firm maximization. Under competition

$$\mathbf{F}^{\text{comp}}(\mathbf{X}, \mathbf{P}) = (\mathbf{c}[\mathbf{Q}^\star, \mathbf{x}], \mathbf{Q}^\star)$$

where $\mathbf{Q}^\star = \text{argmax}_{\mathbf{Q}} \mathbf{P}^\top \mathbf{Q} - C(\mathbf{Q}) - \sum_{i=1}^N c_i(Q_i, x_i)$. Under monopoly

$$\mathbf{F}^{\text{monop}}(\mathbf{c}[\mathbf{X}], \mathbf{X}, \overline{\mathbf{P}}) = (\mathbf{c}[\mathbf{Q}^\star, \mathbf{x}], \mathbf{Q}^\star, \mathbf{P}[\mathbf{Q}^\star])$$

where $\mathbf{Q}^\star = \text{argmax}_{\mathbf{Q}} \left(\mathbf{P}[\mathbf{Q}] \vee \overline{\mathbf{P}}\right)^\top \mathbf{Q} - C(\mathbf{Q}) - \sum_{i=1}^N c_i(Q_i, x_i)$ and the $\vee$ operator denotes the entry-wise minimum. Finally under oligopoly with conjecture $\boldsymbol{\sigma}'$ (flat at the equilibrium under static Nash, equal to the actual $\boldsymbol{\sigma}$ under consistentcy)

$$\mathbf{F}^{\text{oligop}}(\mathbf{X}, \overline{\mathbf{P}}) = (\mathbf{c}[\mathbf{Q}^\star, \mathbf{x}], \mathbf{Q}^\star, \mathbf{P}[\mathbf{Q}^\star, \boldsymbol{\sigma}(\mathbf{Q}^\star)], \boldsymbol{\sigma}[\mathbf{Q}^\star])$$

where $\mathbf{Q}^\star = \text{argmax}_{\mathbf{Q}} \left(\mathbf{P}[\mathbf{Q}, \boldsymbol{\sigma}'(\mathbf{Q})] \vee \overline{\mathbf{P}}\right)^\top \mathbf{Q} - C(\mathbf{Q}) - \sum_{i=1}^N c_i(Q_i, x_i)$. Any standard model has conjectures that are at least locally consistent, that is $\boldsymbol{\sigma}'(\mathbf{Q}^\star) = \boldsymbol{\sigma}(\mathbf{Q}^\star)$. Thus the structural restriction S are, in each case, the set of all functions which may arise from cost functions, price systems, reactions and conjectures satisfying my assumptions.

Theorem 1. *Consider the further restriction that, at* $(\mathbf{X}^\star, \mathbf{P}^\star)$ *or* $(\mathbf{X}^\star, \overline{\mathbf{P}}^\star)$ *as appropriate, the structural function is twice differentiable. Then*

1. *Under perfect competition, monopoly and consistent conjectures oligopoly, for any* $N > 1$*th-order point or set observability condition including* $\frac{\partial Q_i}{\partial x_j}$ *(or its negative) for* $i \neq j$, $\frac{\partial Q_j}{\partial x_i}$ *is sign identified. If these are contained in the point observability condition and this also includes* $\frac{\partial^2 c_i}{\partial x_i \partial Q_i}$ *and* $\frac{\partial^2 c_j}{\partial x_j \partial Q_j}$ *then* $\frac{\partial Q_j}{\partial x_i}$ *is point identified.*
2. *Under perfect competition if the (point or set) observability condition includes* $\frac{\partial Q_i}{\partial x_j}$ *(or its negative) for any* i, j *then* $\frac{\partial Q_i}{\partial P_j}$ *and* $\frac{\partial Q_j}{\partial P_i}$ *are sign identified. If this is in the point observability condition, as is* $\frac{\partial^2 c_j}{\partial x_j \partial Q_j}$, *then* $\frac{\partial Q_i}{\partial P_j}$ *and* $\frac{\partial Q_j}{\partial P_i}$ *are point identified.*
3. *Under monopoly or consistent conjectures oligopoly if the (point or set) observability condition includes* $\frac{\partial P_i}{\partial x_j}$ *(or its negative) for any* i, j *then* $\frac{\partial Q_j}{\partial \overline{P}_i}$ *is sign identified. If all first-order price and quantity effects of all* x_i *are observable, then* $\frac{\partial Q_j}{\partial \overline{P}_i}$ *is point-identified.*
4. *Under inconsistent conjectures oligopoly, even if* $\frac{\partial \boldsymbol{\sigma}}{\partial x_i}$ *is observed for all* i*, the objects that are identified above are unidentified.*

Under any (consistently) optimizing model, responses to cost shocks must satisfy Slutsky's symmetry conditions. With perfect competition the response to a change in price is identical to change in cost. With imperfect, but consistent conjectures, competition prices are choice variables so Slutsky's symmetry conditions tie the effect of a cost shock to that of a price control, which raises the opportunity cost having a higher price. Under inconsistent conjectures anything goes; my proof of this in the appendix uses a simple Cournot example, so this is not rectified by structuring the interaction. This is one complication accompanying oligopoly which arises from the inconsistency of conjectures, rather than from strategic interactions themselves. There is true more generally of oligopoly identification: in Appendix II. C I show that Baker and Bresnahan (1988)'s extension of Rosse (1970)'s classic reasoning to recover of an oligopolist's marginal costs from exogenous variation in *only its own cost* is valid only under consistent conjectures the authors assume.

Proof. See Appendix II.C.

The key to my proof, which generalizes that of Slutsky and others to allow for demand that is not differentiable globally, is Milgrom and Segal (2002)'s envelope theorem for general choice sets and preferences. It has long been known (Hotelling, 1938; Silberberg, 1971) that the Slutsky conditions are a corollary of the envelope theorem, so it should not be surprising that an extension of the latter extends the former. Once Slutsky's conditions are established, the results follow directly from the logic of the previous section. Of course the envelope theorem depends on quasi-linearity of preferences in money, which is what makes the conditions more plausible for firms, which have no income effects, than individuals.

The models can be tested based on symmetry conditions, as my over-identification results stated formally in Appendix II.D show. This provides a test, in imperfectly competitive market of monopoly or consistency of conjectures against inconsistent conjectures oligopoly.

E. General results: concavity/stability

As Slutsky emphasized, his symmetry condition is necessary, but not sufficient, for differentiable behavior to arise from optimization. They represent only the first-order conditions for optimization. Second-order conditions also place bounds on the firm's responses to various cost shocks. In the single product case these amount to the law of demand, which carries over to the multi-product context: a rise in marginal cost at all points reduces production. However a novel element arises in the multivariate context. This roughly states that cross-partial derivatives cannot be too large in magnitude relative to own-partial derivatives. Both of these results provide natural tests[47] of the model.

[47] They could also be used for (set) identification, but these applications strike me as less natural and I therefore present the results in terms of over-identification.

Theorem 2. *Maintain the hypothesis of Theorem 1 and let* $\mathbf{H} = \left[\frac{\partial Q_i}{\partial x_j}\right]_{ij}$ *at the equilibrium point. Suppose that the observability condition contains all entries of some principal submatrix of* $\mathbf{H}$*. Then the model is over-identified under monopoly, competition or consistent conjectures oligopoly. In particular, the relevant principal minor must be of sign* $(-1)^i$*, where* i *is the order of the submatrix, to be consistent with the model. Note that any 1x1 submatrix is just some* $\frac{\partial Q_i}{\partial x_i}$ *for some* x_i *so the requirement is that this be weakly negative.*

The result puts no restriction, in either direction, on inconsistent conjectures oligopoly[48]. Specifically there is no expectation that these conditions will typically be violated with inconsistent conjectures and thus they do not form a very powerful test, as with symmetry.

Proof. See Appendix II.C.

The result follows directly from second-order conditions: the Hessian of profits with respect to costs must be negative definite. Equilibrium responses of quantities to cost shocks are the inverse of this matrix and thus must also be negative definite[49].

F. Applications

To make the preceding results more concrete and show how they might be used in practice, in this section I discuss a few potential applications of some recent policy or empirical interest.

Testing for inconsistent conjectures

Perhaps the clearest application of the above results are as a test for the consistency of the firm's conjectures. Consistency of conjectures is a useful solution concept because of its identifying power. However it is one of the most controversial economic assumptions entering into such strategies. It therefore seems useful to test[50] it, at least in some markets, by examining whether production reactions to firm-specific cost shocks are in fact symmetric.

[48] This is the current state of the theoretical literature. The only conditions known to ensure an analogous test of oligopoly equilibrium to that for monopoly and competition in Theorem 2 hold only under very special conditions such as all entries of $\mathbf{H}$ being weakly negative (Milgrom and Shannon, 1994; Echenique, 2002) or if $\boldsymbol{\sigma}$ enters into P only through some scalar aggregator (Acemoglu and Jensen, 2009). However, I believe that this is simply a hole in the literature and that, at any equilibrium which is stable in a fairly mild sense, Theorem 2 would in fact, unlike Theorem 1's positive results, apply to even inconsistent conjectures oligopoly. I am currently working to prove this (Dittmer and Weyl, 2010) in collaboration with Andrew Dittmer. If we succeed I hope a future version of this paper will include these results.

[49] The exact effects of x_i on cost need not be known, as these scale up all entries in any given column of $\mathbf{H}$ and thus have no effect on the sign of any principal minor.

[50] Such tests would be useful because while consistency of conjectures has be tested in the laboratory (Fouraker and Siegel, 1963; Dolbear et al., 1968; Holt, 1985), the limited number of tests on industry data (Liang, 1989) have relied on strong assumptions about the demand structure.

Some researchers are sufficiently skeptical (Makowski, 1987) of consistent conjectures as a solution concept that they might interpret a success of their test as indicating the absence of strategic interactions (monopoly or perfect competition).

Dynamic cost and demand effects

One of the most classical topics in economics is cross-temporal complementarities in the production (Arrow, 1962) or consumption (Klemperer, 1987; Becker and Murphy, 1988; Abel, 1990; Constantinides, 1990) of certain goods. Foster et al. (2009) measure the latter effect for the products of industrial plants using a structural model of a dynamic industry demand system. This model is parametrically identified by restrictions on the serial correlation of observable exogenous demand fluctuations and cross-sectional variation in the age of plants.

Of course it would be preferable if these assumptions could be relaxed and identification were still possible. Two features of the Foster et. al. setting suggest this may be possible using the results derived above. First, they assume a monopoly model, so the results above would seem to be applicable. Second, they have detailed cost data and thus might plausibly be able to synthesize exogenous cost shocks, though of course this would require some external source of variation which does not currently appear in the paper.

Their setting involves the production of a single commodity in each of a number of periods, rather than many commodities in a single period as my preceding analysis seemed to. However, it can easily be seen that, so long as the monopolist is dynamically consistent or able to commit to her production schedule ex-ante, these two settings are equivalent, as formally shown in Appendix II.E. Production in each period[51] can be seen as different commodities in the spirit of Arrow and Debreu (1954).

Thus if Foster et. al. could isolate *either* exogenous variation in current costs of production and observe their affect on average future production *or* shocks to average future cost (perhaps through futures prices for inputs) and observe their effect on current production, they could non-parametrically test the complementarities of interest[52]. Furthermore, if both types of data were available, the model could simultaneously be tested. More broadly any

[51]Uncertainty introduces a slight complication. In this case there is one commodity not just for each time period, but for each state of the world and time period. However we can relabel these commodities (see Appendix II.E) so that one stands in for the *average* production in each period and the others represent excess production in each of the states of the world, in that period, subtracted from some omitted state of the world. A "clean" cost shock to this average production would be one that affects costs only through the average produced: that is a marginal cost shock that is constant across states of the world. The effect of a shock to current cost on average future production must then, by Slutsky's symmetry, be identical the effect of a shock to this average cost on current production.

[52]In fact this would test complementarities in firm profits and thus not distinguish between the cost and demand side, but this could easily be solved by using their direct cost estimates.

dynamic and dynamically consistent production model can be (partially, locally and with the right data) identified or tested using the results above.

Production under uncertainty

The same arguments as in the preceding section may be applied to a firm producing in many states of the world in a single period, following an initial stage of investment. Often production across states may be complementary, if they all rely on common inputs that must be purchased in advance, or substitutable, if building the facilities needed to produce in one state (warm weather) precludes producing in another (cold weather). Foretold shocks to the cost of producing in one state must have Slutsky symmetric effects in both directions.

Mergers of multi-product firms

Antitrust analysis of mergers between multi-product firms must account for the effect that the anti-competitive upward pricing pressure the merger exerts on one of the merging products may have on the prices of other products of that firm (Moresi et al., 2008). As discussed above and especially in Salinger (1991), Edgeworth's paradox of taxation implies that upward pricing pressure on one of the products of a multi-product firm need not raise prices on that or any of the firm's products. However, as long as one is willing to assume the firm has had consistent conjectures in the past, the sign and even magnitude of some of the cross-production and pricing effects can be fully or partially identified with limited data.

Multi-sided platforms

One type of multi-product firm with market power of particular interest in recent years has been those whose products have consumption externalities to consumers of different products produced by the firm. Examples include advertising-based industries, payment cards, internet service providers and operating systems. Assuming, as is common, that the goods are not direct substitutes or complements for one another, any cross-price effects of production result from externalities (Rochet and Tirole, 2006; Weyl, Forthcoming). Combined with the results above, this can be use to answer an number of questions in models of multi-sided platforms: for example, the effects of price controls on participation on both sides can be determined by observing price shocks. Market power can be identified and costs recovered through cost variation using my general formula for market power from Weyl (Forthcoming). The sign of cross production effects can be identified from a cost shock for either good. Hopefully some of these results can be applied in the growing empirical literature on multi-sided platforms, which has been pioneered by Rysman (2004, 2007).

V. Marschak's Maxim: Advantages and Drawbacks

First-order identification provides a language for answering Marschak's query: what are "minimal" collections of structural assumptions and observations needed to sign or fully identify a level or derivative of economic interest? It does so by formalizing the "amount" of structural assumptions through the size of the structural restriction and "data requirements" by the number of derivatives assumed observable. More generally than the above examples, the relative merits of that approach turn on the appropriateness of Marschak's Maxim.

A. Advantages

Perhaps the most important advantage of the Marschakian approach, relative to parametric identification, is the transparency it imposes on the use of structural economic assumptions for identification. To take a simple example, Sumner (1981) argued that high pass-through rates in the cigarette industry suggest high market power. Bulow and Pfleiderer (1983) showed that this connection between pass-through and market power is an artifact of the constant elasticity demand form Sumner assumed. Of course rather than assuming constant elasticity, Sumner might have explicitly assumed that demand was such that high pass-through rates imply high market power. This assumption would imply, via a first-order identification argument, that the first-order effect of cost shocks on prices identifies market power. This identification might not have been very plausible, however, given the weak motivation of this assumption. The basic trouble with the parametric approach is thus not that the assumptions driving its identification are *necessarily* implausible, but rather that their role in identification is often obscure and therefore it is difficult even to judge[53] their plausibility. By focusing on broad and transparent economic , rather than functional form, assumptions first-order identification helps clarify the source of identification.

The advantages of first-order identification relative to LATEs and non-parametric identification are simpler. As extensively discussed above, in many cases LATEs are simply inapplicable without some structural assumptions, which first-order identification allows. Because data sets are always finite and first-order identification has an explicit technology for formalizing this, which non-parametric identification lacks, first-order identification provides a clearer sense of whether non-parametric identification can "work" on given data.

First-order identification has another, practical advantage over other structural methods. Techniques typically used to derive first-order identification results are the essence of Neoclassical price theory which many advanced undergraduate and almost all graduate students are familiar with: the implicit function theorem, envelope conditions, basic linear algebra,

[53] In Mazur and Weyl (2010), we provide a formal model of assumptions that are "easier to judge".

etc. This is a striking contrast to methods typically required to derive nonparametric and sometimes even parametric identification results, which are typically drawn from statistical theory familiar only to a specialized subset of econometricians. This has three advantages.

First and most importantly, it augments the transparency of the the results, making clear not only the "transparent" economic assumptions needed for an identification result but exactly how these assumptions drive the result. This in turn provides a clearer picture of how robust the results may be. Second, this allows the results to be communicated and therefore put into practice more rapidly. Finally, because the techniques involved are essentially equivalent to those used for the standard comparative statics analysis, they put contributions to econometrics and empirical work within the grasp of applied theorists[54].

A final virtue of the first-order approach is its flexibility. While it may seem that studying first-order identification limits one to interest in first-order effects while parametric identification computes full equilibrium effects, this is largely a mirage. Identification of second- or higher-order effects can easily be incorporated, though of course more data will typically be required to estimate these. As higher order effects are estimated, by Taylor's Theorem, the entire function is traced out and the precise effect of discrete changes can be identified. Thus first-order identification is no more "local" than is required by the data[55].

B. Drawbacks

Of course with these advantages come several important drawbacks of first-order identification and the Marschakian approach more broadly. Relative to LATEs, first-order identification requires strong, restrictive and sometimes even implausible structural economic assumptions. Thus when it is feasible to directly measure them, LATEs are preferable and will be more persuasive[56] to a broad audience than any analysis based on first-order identification, however relaxed the assumptions it is based on are.

Relative to nonparametrics, the scope of any first-order identification analysis is very limited. While in principle high-order effects could be studied using the first-order methods, in practice this would quickly become cumbersome. Nonparametric identification analysis provides an overview of sorts, pointing out which approaches are inherently doomed to failure

[54]A common criticism of applied theory is that many outcomes are ambiguous; in this case, critics assert, what is the contribution of the theory? First-order identification provides a simple answer: to show what exactly would need to be measured to sign or measure the comparative static of interest. Thus many if not most such "ambiguous" applied theory results can instead be viewed as positive first-order identification results, easing the communication between economic theory and econometrics.

[55]An applied example is described in Appendix I.D.

[56]Imbens (2009) surveys the benefits of LATEs, holding fixed the question of economic interest an the feasibility of measurement. In practice whether a structural or LATEs approach is appropriate will depend on the time and resource cost of generating appropriate data.

and which ones may succeed with sufficient data. Thus nonparametric identification is a very strong complement for first-order identification.

First-order identification has three primary disadvantages relative to the parametric approach. First, there are many settings where data is limited and patchy over a fairly wide range, rather than rich in a limited range. While this could be thought of as identifying some derivatives about a midpoint or series of point (Hastie and Loader, 1993), the choice of points is fairly arbitrary. Second, under the parametric approach once the model is identified it can be used for a wide range of purposes (Hausman, 1997), many of which are not even intended ex-ante. Consider, for example, the enormous range of applications of the framework of Berry et al. (1995), including the authors' own study of voluntary export restraints (Berry et al., 1999), which was certainly not the initial motivation of their demand estimation. To implement first-order identification would require a separate analysis for each new data source, question of policy interest, etc. This imposes a substantial analytic burden[57] on applied researchers, slowing the pace of cumulation of economic knowledge.

Third, even if infinite time and analytical resources were available, parametric models do have a potential *cumulative scientific* (as opposed to applied) advantage over nonparametric models. As Popper (1959) famously argued[58], the validity of scientific theories depends not only on their plausibility, generality and consistency with observed data, but also on the extent to which they are falsifiable. Models which are difficult to falsify are also difficult to believe[59], unless their assumptions are inherently very compelling. Parametric models, because they are more restricted, are also often more falsifiable[60]. Such opportunities for falsification are particularly useful when the underlying economic assumptions being tested are applied in a range of contexts, allowing the results to inform economic science.

[57]When a particular question of interest is sufficiently important to merit the detailed focus required for first-order identification analysis is a judgement call applied researches will have to make. However, when it is available, such analysis will likely often be at least more transparent, if not more persuasive, than the more straight-forward parametric approach.

[58]See Blumer et al. (1987, 1989) for a formalization.

[59]In Weyl (2007) I formalize a very special case of this trade-off.

[60]This requires the functional form has independent plausibility so that the falsification attributed to underlying economic assumptions. Often in practice, the use of implausible or arbitrary functional forms undermines this process, because to the extent that tests of the model are rejected (if they are even checked) the failure is typically attributed to the functional form (if to anything) rather than to the underlying economic assumptions about optimization, information, the strategic environment, etc. This can actually have the opposite of the intended falsifying effect, shielding maintained assumptions from falsification in practice, even if models are rejected in principle. However, in principle, false generalization is a pitfall to be avoided; plausible restrictions *augment* not detract from the credibility of a model's predictions (Friedman, 1953). Generalization, to the extent it is undertaken, should always seek to remove the most implausible assumptions first. To the extent that it enables or encourages blind generalizations that shield underlying, maintained economic assumptions from falsification, first-order identification misses many of the advantages of a fully specified structural model (Deaton, 2009, Forthcoming).

Finally, any approach that suppresses the stochastic model has the significant disadvantage of requiring more steps to bring the identification result to empirical practice. This is the corresponding downside of the extreme simplicity benefits of first-order identification.

Thus it should be clear that first-order identification is properly viewed as only one complementary approach in the search for identification, one most appropriate in combination or at least dialogue with the other classic approaches. Yet despite its drawbacks, I believe its substantial virtues make it a useful tool in an econometrician's arsenal.

VI. Conclusion

This paper has argued that first-order identification is a useful way to systematize a common methodology in applied economics. The goal of this formalization is three fold. First, by posing first-order identification as a formal econometric problem I hope to encourage applied econometricians to view it as such, devoting at least some attention to exploring the first-order identification of practical models. Second, I hope to persuade applied economic theorists that they could make the results of their analysis more useful to econometricians and empiricists if ambiguities in models were instead framed as first-order identification results. Finally I aim to supply a somewhat systematic language and set of advantages and disadvantages for empiricists to keep in mind when considering whether to use a first-order, parametric, non-parametric or LATE/IV approach to studying a question of interest.

I suspect (hope) this paper raises more questions than it answers and thus conclude by discussing some of the important directions in which the inquiry here might be extended.

A popular technique of non-parametric estimation is the so-called "sieve" method of Grenander (1981) and Geman and Hwang (1982), in which increasingly complex models are estimated as the quantity of data increases. A natural extension of my reasoning would be a Marschakian sieve. First, determine which local derivatives (and/or levels) must be estimated to determine $\frac{\partial o}{\partial p}$ and ensure the first-level model is flexible on all parameters. Then do the same for $\frac{\partial^2 o}{\partial p^2}$ and ensure the second-level model is flexible on these and so forth. The analysis raises other econometric issues. What size data corresponds to various derivatives? What are the appropriate assumptions about boundedness of higher-order effects that allow variation to be thought of as "local" and identifying local derivatives?

Discontinuities in structural functions challenge first-order identification. These may occur in insufficiently concave environments. Suppose that, in the local range where we have thus far observed data, production is differentiable but we suspect it may jump. What derivatives of the underlying profit function (revealed by behavior) would we need to observe to estimate where, and to where, the jump occurs?

A step closer to applications, I hope that first-order identification analysis may prove feasible in non-parametric versions of influential, complex structural models. Four prominent areas immediately suggest themselves: dynamic stochastic general equilibrium models (Kydland and Prescott, 1982), structural models of labor market interventions (Heckman and Robb, 1985; Carneiro et al., Forthcoming), multi-product demand systems (Berry et al., 1995) and industrial dynamics (Ericson and Pakes, 1995) . Explicit analysis of these models, separating the role of theory clearly from that of data, could both help make the results more comprehensible and credible for policy makers as well as making clearer to econometricians the robustness of the results from these models. While this seems daunting, Chetty (2009)'s and, I hope, my above analysis have shown that even complex models include basic principles of optimization that allow identification without restrictive assumptions.

Finally, existing results on first-order identification have a number promising applications, including to practical policy making. The US antitrust enforcement agencies are currently revising the guidelines for horizontal mergers. The chiefs of the economics divisions of these agencies are currently Joe Farrell and Carl Shapiro, authors of a paper outlining a first-order approach to merger analysis. While the outcome is hard to predict, there seems to be a reasonable chance that antitrust policy in coming years will put the principles Jacob Marschak used to study monopoly into the service of preventing it from arising.

References

Abel, Andrew B., "Asset Prices under Habit Formation and Catching up wiht the Joneses," *American Economic Review: Papers and Proceedings*, 1990, *80* (2), 38–42.

Acemoglu, Daron, "When Does Labor Scarcity Encourage Innovation," 2009. http://econ-www.mit.edu/files/4346.

_ **and Martin Kaae Jensen**, "Aggregate Comparative Statics," 2009. http://econ-www.mit.edu/files/4870.

Aguiar, Mark and Erik Hurst, "Consumption versus Expenditure," *Journal of Political Economy*, 2005, *113* (5), 919–948.

Aguirre, Iñaki, Simon George Cowan, and John Vickers, "Monopoly Price Discrimination and Demand Curvature," *American Economic Review*, Forthcoming.

Ait-Sahalia, Yacine and Andrew Lo, "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 1996, *64* (3), 527–560.

Angrist, Joshua D., "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, 1990, *80* (3), 313–336.

_ **and Alan B. Krueger**, "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 2001, *15* (4), 69–85.

Arrow, Kenneth J., "The Economic Implications of Learning by Doing," *Review of Economic Studies*, 1962, *29* (3), 155–173.

_ **and Gerard Debreu**, "Existence of an Equilibrium for a Competitive Economy," *Econometrica*, 1954, *22* (3), 265–290.

Athey, Susan and Philip A. Haile, "Identification of Standard Auction Models," *Econometrica*, 2002, *70* (6), 2107–2140.

Baily, Martin Neil, "Some Aspects of Optimal Unemployment Insurance," *Journal of Public Economics*, 1978, *10* (3), 379–402.

Baker, Jonathan B. and Timothy F. Bresnahan, "The Gains from Merger or Collusion in Product-Differentiated Industries," *Journal of Industrial Economics*, 1985, *33* (4), 427–444.

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

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden, "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics*, 2007, *122* (3), 1235–1264.

Becker, Gary S. and Kevin M. Murphy, "A Theory of Rational Addiction," *Journal of Political Economy*, 1988, *96* (4), 675–700.

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

Berry, Steven, James Levinsohn, and Ariel Pakes, "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy," *American Economic Review*, 1999, *89* (3), 400–430.

Berry, Steven T. and Philip A. Haile, "Identification of Discrete Choice Demand from Market Level Data," 2009. http://www.econ.yale.edu/∼pah29/marketdata.pdf.

_ **and** _ , "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," 2009. http://www.econ.yale.edu/∼pah29/microdata1108.pdf.

Blumer, Anselm, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth, "Occam's Razor," *Information Processing Letters*, 1987, *24* (6), 377–380.

_ **,** _ **,** _ **, and** _ , "Learnability and the Vapnik-Chervonenkis Dimension," *Journal of the Association for Computing Machinery*, 1989, *36* (4), 929–965.

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

Bresnahan, Timothy F., "Duopoly Models with Consistent Conjectures," *American Economic Review*, 1981, *71* (5), 934–945.

_ , "The Oligopoly Solution Concept is Identified," *Economics Letters*, 1982, *10* (1–2), 87–92.

_ , "Empirical Studies of Industries with Market Power," in Richard Schmalensee and Robert Willig, eds., *Handbook of Industrial Organization*, Vol. 2, Oxford, UK: Elsevier, 1989, pp. 1011–1057.

_ **and Peter C. Reiss**, "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 1991, *99* (5), 977–1009.

Brown, Bryan W., "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 1983, *51*, 175–196.

Bulow, Jeremy I. and Paul Klemperer, "Price Controls and Consumer Surplus," 2009. http://www.gqq10.dial.pipex.com/.

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

_ **, John D. Geanakoplos, and Paul D. Klemperer**, "Multimarket Oligopoly: Strategic Substitutes and Compliments," *Journal of Political Economy*, 1985, *93* (3), 488–511.

Campbell, John Y. and N. Gregory Mankiw, "Are Output Fluctuations Transitory?," *Quarterly Journal of Economics*, 1987, *102* (4), 857–880.

Card, David, Raj Chetty, and Andrea Weber, "Cash-On-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market," *Quarterly Journal of Economics*, 2007, *122* (4), 1511–1560.

Carneiro, Pedro, James J. Heckman, and Edward Vytlacil, "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin," *Econometrica*, Forthcoming.

Chesher, Andrew, "Identification in Nonseparable Models," *Econometrica*, 2003, *71* (5), 1405–1441.

Chetty, Raj, "A General Formula for the Optimal Level of Social Insurance," *Journal of Public Economics*, 2006, *90* (10–11), 1879–1901.

_ , "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, 2008, *116* (2), 173–232.

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

Chipman, John S. and Jean-Sébastien Lenfant, "Slutsky's 1915 Article: How it Came to be Found and Interpreted," *History of Political Economy*, 2002, *34* (3), 553–597.

Constantinides, George M., "Habit Formation: A resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 1990, *98* (3), 519–543.

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

Deaton, Angus, "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development," 2009. http://www.princeton.edu/~deaton/downloads/Instruments%20ofent%20v1d_mar09_all.pdf.

_ , "Understanding the Mechanisms of Economic Development," *Journal of Economic Perspectives*, Forthcoming.

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

Diamond, Peter A., "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Tax Rates," *American Economic Review*, 1998, *88* (1), 83–95.

Dittmer, Andrew and E. Glen Weyl, "The Correspondence Principle and the Law of Demand," 2010. This paper is currently being prepared. Contact me at weyl@fas.harvard.edu for notes.

Dockner, Engelbert J., "A Dynamic Theory of Conjectural Variations," *Journal of Industrial Economics*, 1992, *40* (4), 377–395.

Dolbear, F. T., L. B. Lave, G. Bowman, A. Liberman, E. Prescott, F. Rueter, and R. Sherman, "Collusion in Oligopoly: An Experiment on the Effect of Numbers and Information," *Quarterly Journal of Economics*, 1968, *82* (2), 240–259.

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

Echenique, Federico, "Comparative Statics by Adaptive Dynamics and the Correspondence Principle," *Econometrica*, 2002, *70* (2), 833–844.

Edgeworth, Francis Ysidro, *The Pure Theory of Monopoly*, Vol. 1, New York: Burt Franklin, 1925.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen, "Estimating Welfare in Insurance Markets Using Variation in Prices," *Quarterly Journal of Economics*, Forthcoming.

Ericson, Richard and Ariel Pakes, "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 1995, *62* (1), 53–82.

Farrell, Joseph and Carl Shapiro, "Antiturst Evaluation of Hoirzontal Mergers: An Economic Alternative to Market Definition," 2008. http://faculty.haas.berkeley.edu/shapiro/alternative.pdf.

Feldstein, Martin S., "Tax Avoidance and the Deadweight Loss of the Income Tax," *Review of Economics and Statistics*, 1999, *81* (4), 674–680.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "The Slow Growth of New Plants: Learning about Demand?," 2009. http://home.uchicago.edu/∼syverson/learningaboutdemand.pdf.

Fouraker, L. E. and S. Siegel, *Bargaining Behavior*, New York: McGraw-Hill, 1963.

Friedman, Milton, *The Methodology of Positive Economics*, Chicago: University of Chicago Press, 1953.

Froeb, Luke, Steven Tschantz, and Gregory J. Werden, "Pass-Through Rates and the Price Effects of Mergers," *International Journal of Industrial Organization*, 2005, *23* (9–10), 703–715.

Geman, Stuart and Chil-Ruey Hwang, "Nonparametric Maximum Likelihood Estimation by the Method of Sieves," *Annals of Statistics*, 1982, *10* (2), 401–414.

Gopinath, Gita and Oleg Itskhoki, "Frequency of Price Adjustment and Pass-Through," *Quarterly Journal of Economics*, Forthcoming, *125* (2).

Grenander, Ulf, *Abstract Inference*, Wiley: New York, 1981.

Gruber, Jon and Emmanuel Saez, "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics*, 2002, *84* (1), 1–32.

Gruber, Jonathan, "The Consumption Smoothing Benefits of Unemployment Insurance," *American Economic Review*, 1997, *87* (1), 192–205.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong, "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 2000, *68* (3), 525–574.

Harberger, Arnold C., "Monopoly and Resource Allocation," *American Economic Review: Papers and Proceedings*, 1954, *44* (2), 77–87.

_ , "The Incidence of the Corporation Income Tax," *Journal of Political Economy*, 1962, *70* (3), 215–240.

_ , "The Measurement of Waste," *American Economic Review*, 1964, *54* (3), 58–76.

Harrison, Glenn W. and John A. List, "Field Experiments," *Journal of Economic Literature*, 2004, *42* (4), 1009–1055.

Hastie, Trevor and Clive Loader, "Local Regression: Automatic Kernel Carpentry," *Statistical Science*, 1993, *8* (2), 120–128.

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

_ **and Whitney K. Newey**, "Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss," *Econometrica*, 1995, *63* (6), 1445–1476.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

_ **and Edward J. Vytlacil**, "Local Instrumetnal Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 1999, *96* (8), 4730–4734.

_ **and** _ , "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6B, Elsevier B. V.: Amsterdam, Holland, 2007.

_ **and Edward Vytlacil**, "Policy-Relevant Treatment Effects," *American Economic Review: Papers and Proceedings*, 2002, *91* (2), 107–111.

_ **and** _ , "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 2005, *73* (3), 669–738.

_ **and Richard Robb**, "Alternatives Methods for Evalutating the Impacts of Interventions: an Overview," *Journal of Econometrics*, 1985, *30* (1–2), 239–267.

_ **and Sergio Urzúa**, "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify," *Journal of Econometrics*, Forthcoming.

Hines, Jr. James R., "Three Sides of Harberger Triangles," *Journal of Economic Perspectives*, 1999, *13* (2), 167–188.

Holt, Charles A., "An Experimental Test of the Consistent-Conjectures Hypothesis," *American Economic Review*, 1985, *75* (3), 314–325.

Hotelling, Harold, "Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions," *Journal of Political Economy*, 1932, *40* (5), 577–616.

_ , "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica*, 1938, *6* (3), 242–269.

Hoxby, Caroline M., "Does Competition among Public Scools Benefit Students and Taxpayers," *American Economic Review*, 2000, *90* (5), 1209–1238.

Imbens, Guido W., "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *Review of Economics and Statistics*, 2004, *86* (1), 4–29.

_ , "Better LATE than Nothing: Some Comments on Deaton (2009 and Heckman and Urzúa (Forthcoming)," 2009. http://www.economics.harvard.edu/faculty/imbens/files/bltn_09apr28.pdf.

_ **and Joshua D. Angrist**, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

Inada, Ken-Ichi, "On a Two-Sector Model of Economic Growth: Comments and a Generalization," *Review of Economic Studies*, 1963, *30* (2), 119–127.

Jenkin, H. C. Fleeming, "The Graphic Representation of the Laws of Supply and Demand and their Application to Labour," *Recess Studies*, 1870.

_ , "On the Principles which Regulate the Incidence of Taxes," *Proceedings of the Royal Society of Edinburgh*, 1871–1872, pp. 618–631.

Jevons, William Stanley, *The Theory of Political Economy*, London: MacMillan, 1871.

Kind, Hans Jarle, Mark Koethenbuerger, and Guttorm Schjelderup, "Efficiency Enhancing Taxation in Two-Sided Markets," *Journal of Public Economics*, 2008, *92* (5–6), 1531–1539.

Klein, Lawrence R., *Economic Flutuations in the United States, 1921-1941* number 11. In 'Cowles Commission Monograph.', New York, NY: John Wiley, 1950.

Klemperer, Paul D., "Markets with Consumer Switching Costs," *Quarterly Journal of Economics*, 1987, *102* (2), 375–394.

_ **and Margaret A. Meyer**, "Supply Function Equilibria in Oligopoly under Uncertainty," *Econometrica*, 1989, *57* (6), 1243–1277.

Koopmans, Tjalling C., "Identification Problems in Econometric Model Construction," *Econometrica*, 1949, *17* (2), 125–144.

_ **, Herman Rubin, and Roy B. Leipnik**, *Measuring the Equation System of Dynamic Economics*, Vol. 10 of *Cowles Commission Monograph*, New York, NY: John Wiley, 1950.

Kydland, Finn E. and Edward C. Prescott, "Time to Build and Aggregate Fluctuations," *Econometrica*, 1982, *50* (6), 1345–1370.

Lee, David and Emmanuel Saez, "Optimal Minimum Wage Policy in Competitive Labor Markes," 2008. http://papers.ssrn.com.ezp-prod1.hul.harvard.edu/sol3/papers.cfm?abstract_id=1267551.

Lerner, Abba P., "The Concept of Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies*, 1934, *1* (3), 157–175.

Levitt, Steven D., "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review*, 1997, *87* (3), 270–290.

Liang, J. Nellie, "Price Reaction Fucntions and Conjectural Variations: An Application to the Breakfast Cereal Industry," *Review of Industrial Organization*, 1989, *4* (2), 31–58.

List, John A., "Does Market Experience Eliminate Market Anomalies," *Quarterly Journal of Economics*, 2003, *118* (1), 41–71.

Lucas, Robert E., "Econometric Policy Evaluation: A Critique," *Carnegie-Rochester Conference Series on Public Policy*, 1976, *1*, 19–46.

Makowski, Louis, "Are 'Rational Conjectures' Rational?," *Journal of Industrial Economics*, 1987, *36* (1), 35–47.

Manski, Charles F., "Nonparametric Bounds on Treatement Effects," *American Economic Review: Papers and Proceedings*, 1990, *80* (2), 319–323.

_ , *Identification Problems in the Social Sciences*, Cambridge, MA: Harvard University Press, 1995.

_ , "Learning about Treatment Effects from Experiments with Random Assignment of Treatments," *Journal of Human Resources*, 1996, *31* (4), 709–773.

_ , *Partial Identification of Probability Distributions*, New York: Springer-Verlag, 2003.

Marschak, Jacob, "Econometric Measurements for Policy and Prediction," in "Studies in Econometric Methods," Wiley: New York, 1953, pp. 1–26.

Matzkin, Rosa L., "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 1993, *58* (1–2), 137–168.

_ , "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 2003, *71* (5), 1339–1375.

_ , "Nonparametric Identification," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6B, Elsevier B. V.: Amsterdam, Holland, 2007.

Mazur, Barry and E. Glen Weyl, "Persuasion by Argument," 2010. This paper is currently being prepared; email me at weyl@fas.harvard.edu for notes.

Milgrom, Paul and Chris Shannon, "Monotone Comparative Statics," *Econometrica*, 1994, *62* (1), 157–180.

_ **and Ilya Segal**, "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 2002, *70* (2), 583–601.

Mirrlees, J. A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38* (2), 175–208.

Moresi, Serge X., Steven C. Salop, and John R. Woodbury, "Implementing the Hypothetical Monopolist SSNIP Test With Multi-Product Firms," *The Antitrust Source*, 2008, *7* (3).

Nathanson, Charles and E. Glen Weyl, "Psychic Income, Taxes and the Allocation of Talent," 2010. This paper is currently being prepared; a earlier version is avaliable at http://www.glenweyl.com.

Nevo, Aviv, "Identification of the Oligopoly Solution Concept in a Differentiated-Products Industry," *Economics Letters*, 1998, *59* (3), 391–395.

Panzar, John C. and James N. Rosse, "Testing for "Monopoly" Equilibrium," *Journal of Industrial Economics*, 1987, *35* (4), 443–456.

Popper, Karl R., *The Logic of Scientific Discovery*, New York: Hutchinson, 1959.

Reiss, Peter C. and Frank A. Wolak, "Structural Econometric Modeling: Rationales and Examples from Industrial Organization," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6A, Elsevier B. V.: Amsterdam, Holland, 2007, chapter 64.

Resnick, Paul, Richard Zeckhauser, John Swanson, and Kate Lockwood, "The Value of Reputation on eBay: A Controlled Experiment," *Experimental Economics*, 2006, *9* (2), 79–101.

Rochet, Jean-Charles and Jean Tirole, "Two-Sided Markets: A Progress Report," *RAND Journal of Economics*, 2006, *37* (3), 645–667.

Roehrig, Charles S., "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 1988, *56*, 433–447.

Rosse, James N., "Estimating Cost Function Parameters Without Using Cost Data: Illustrated Methodology," *Econometrica*, 1970, *38* (2), 256–275.

Rust, John, "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 1987, *55* (5), 999–1033.

Rysman, Marc, "Competition Between Networks: A Study of the Market for Yellow Pages," *Review of Economic Studies*, 2004, *71* (2), 483–512.

_ , "An Empirical Analysis of Payment Card Usage," *Journal of Industrial Economics*, 2007, *55* (1), 1–36.

Saez, Emmanuel, "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 2001, *68* (1), 205–229.

Salinger, Michael A., "Vertical Mergers in Multi-Product Industriesl and Edgeworth's Paradox of Taxation," *Journal of Industrial Economics*, 1991, *39* (5), 545–556.

Schultz, Theodore W., *Transforming Traditional Agriculture*, New Haven, CT: Yale University Press, 1964.

Shimer, Robert and Ivan Werning, "Reservation Wages and Unemployment Insurance," *Quarterly Journal of Economics*, 2007, *122* (3), 1145–1185.

Silberberg, Eugene, "The Le Chatelier Principle as a Corollary to a Generalized Envelope Theorem," *Journal of Economic Theory*, 1971, *3* (2), 146–155.

Slutsky, Eugen, "Sulla Teoria del Bilancio del Consummatore," *Gionale degli Economisti e Riista di Statistica*, 1915, *51*, 1–26.

Sumner, Daniel A., "Measurement of Monopoly Behavior: An Application to the Cigarette Industry," *Journal of Political Economy*, 1981, *89* (5), 1010–1019.

Tinbergen, Jan, *Economic Policy: Principles and Design*, Amersterdam: North Holland, 1956.

Varian, Hal R., "The Nonparametric Approach to Demand Analysis," *Econometrica*, 1982, *50* (4), 945–973.

_ , "Non-Parametric Tests of Conusmer Behavior," *Review of Economic Studies*, 1983, *50* (1), 99–110.

_ , "The Nonparametric Approach to Production Analysis," *Econometrica*, 1984, *52* (3), 579–597.

_ , "Non-parametric Analysis of Optimizing Behavior with Measurement Error," *Journal of Econometrics*, 1985, *30* (1–2), 445–458.

Villas-Boas, Sofia Berto and Rebecca Hellerstein, "Identification of Supply Models of Retailer and Manufacturer Oligopoly Pricing," *Economics Letters*, 2006, *90* (1), 132–140.

Werden, Gregory J., "A Robust Test for Consumer Welfare Enhancing Mergers Among Sellers of Differentiated Products," *Journal of Industrial Economics*, 1996, *44* (4), 409–413.

Weyl, E. Glen, "A Simple Theory of Scientific Learning," 2007. http://www.glenweyl.com.

_ , "Double Marginalization in Two-Sided Markets," 2008. http://www.glenweyl.com.

_ , "The Price Theory of Two-Sided Markets," 2009. http://www.glenweyl.com.

_ , "A Price Theory of Multi-Sided Platforms," *American Economic Review*, Forthcoming.

_ **and Michal Fabinger**, "Pass-Through as an Economic Tool," 2009. http://www.glenweyl.com.

Willig, Robert D., "Consumer's Surplus Without Apology," *American Economic Review*, 1976, *66* (4), 589–597.

Wolpin, Kenneth I., "Estimating a Structural Search Model: The Transition from School to Work," *Econometrica*, 1987, *55* (4), 801–817.

Yatchew, Adonis, "Nonparametric Regression Techiniques in Economics," *Journal of Economic Literature*, 1998, *36* (2), 669–721.