

Ho, Kate; Rosen, Adam

Working Paper

Partial identification in applied research: Benefits and challenges

cemmap working paper, No. CWP64/15

Provided in Cooperation with:

The Institute for Fiscal Studies (IFS), London

Suggested Citation: Ho, Kate; Rosen, Adam (2015) : Partial identification in applied research: Benefits and challenges, cemmap working paper, No. CWP64/15, Centre for Microdata Methods and Practice (cemmap), London,
<https://doi.org/10.1920/wp.cem.2015.6415>

This Version is available at:

<https://hdl.handle.net/10419/130077>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Partial identification in applied research: benefits and challenges

Kate Ho
Adam Rosen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP64/15



An ESRC Research Centre

Partial Identification in Applied Research: Benefits and Challenges*

Kate Ho[†]

Adam M. Rosen[‡]

Columbia University and NBER

UCL and CEMMAP

October 7, 2015

Abstract

Advances in the study of partial identification allow applied researchers to learn about parameters of interest without making assumptions needed to guarantee point identification. We discuss the roles that assumptions and data play in partial identification analysis, with the goal of providing information to applied researchers that can help them employ these methods in practice. To this end, we present a sample of econometric models that have been used in a variety of recent applications where parameters of interest are partially identified, highlighting common features and themes across these papers. In addition, in order to help illustrate the combined roles of data and assumptions, we present numerical illustrations for a particular application, the joint determination of wages and labor supply. Finally we discuss the benefits and challenges of using partially identifying models in empirical work and point to possible avenues of future research.

1 Introduction

The goal of identification analysis is to determine what can be learned using deductive reasoning through the combination of models (sets of assumptions) and data. Standard approaches to econometric modeling

*This paper was prepared for an invited session at the 2015 Econometric Society World Congress in Montreal. We thank Richard Blundell, Andrew Chesher, Alon Eizenberg, Charles Manski, Francesca Molinari, and Ariel Pakes for helpful comments and suggestions. Adam Rosen gratefully acknowledges financial support from from the UK Economic and Social Research Council through a grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (CeMMAP) and from the European Research Council (ERC) grant ERC-2009-StG-240910-ROMETA.

[†]Address: Kate Ho, Department of Economics, Columbia University, 420 West 118th Street, New York, NY 10027, United States. kh2214@columbia.edu.

[‡]Address: Adam Rosen, Department of Economics, University College London, Gower Street, London WC1E 6BT, England. adam.rosen@ucl.ac.uk.

for applied research make enough assumptions to ensure that parameters of interest are point identified. However it is still possible to learn about such quantities even if they are not.¹

Econometric models that allow for partial identification, or partially identifying models, make fewer assumptions and use them to generate bounds on the parameters of interest. Such models have a long history, with early papers including Frisch (1934), Reiersol (1941), Marschak and Andrews (1944), and Frechet (1951). The literature on the topic then remained fragmented for several decades, with some further notable contributions such as Peterson (1976), Leamer (1981), Klepper and Leamer (1984), and Phillips (1989). It wasn't until the work of Charles Manski and co-authors beginning in the late 1980s that a unified literature began to emerge, beginning with Manski (1989, 1990).² Several influential papers by Elie Tamer and co-authors applying partial identification to a variety of econometric models (e.g., Haile and Tamer (2003), Honore and Tamer (2006), Ciliberto and Tamer (2009)) have helped to bring these methods into more common use. A collection of papers by Ariel Pakes and co-authors (e.g. Pakes (2010, 2014), Pakes, Porter, Ho, and Ishii (2015), Pakes and Porter (2015)) have shown how structural econometric models in industrial organization naturally produce moment inequalities that partially identify model parameters, and that these parameters can be used to make useful inferences in practice. For a more complete account and more detailed historical references on partial identification, we refer to the useful survey articles Manski (2008) and Tamer (2010).

A feature of the partial identification approach is that it can be used to assess the informational content of different sets of assumptions in a given application, and to weigh the trade-off between (i) adding precision by imposing more assumptions and (ii) reducing the credibility of the results as the assumptions become more stringent and therefore less plausible.³ This can be done without requiring that any given set of assumptions suffices for point identification. As methods for identification and inference have been developed and improved over the last ten to fifteen years, the recent applied literature has used partial identification to address particular econometric issues where stringent assumptions would otherwise be needed. These papers often vary the number and type of identifying assumptions and investigate the impact of these perturbations on the size of the estimated set. They address a wide range of economic questions, across fields such as labor, industrial organization, and international economics.

While partial identification can be used to relax assumptions required to obtain point identification,

¹Throughout this article we use “parameters” to denote the objects of interest on which researchers focus in any particular application. These could for instance be finite dimensional vectors comprising elements of a parametric model, or infinite dimensional objects, such as nonparametrically specified demand functions or unknown distributions across their domains.

²See also, for example, Manski (1994, 1997), Manski and Pepper (2000), Manski and Tamer (2002), and the monograph Manski (2003).

³This trade-off has been coined *The Law of Decreasing Credibility* by Manski (2003).

assumptions continue to play a central role. Indeed, a goal of partial identification analysis is to explore the estimates delivered by different sets of assumptions without the need to make enough, potentially unwarranted, assumptions to achieve point identification. Two particular categories of assumptions are those regarding functional forms for agents' response or utility functions, and those made on the distribution of unobserved variables conditional on observed variables.

Functional form assumptions can range from nonparametric smoothness or shape restrictions to parametric restrictions. These may be motivated directly from economic theory, for example imposing downward-sloping demand or Slutsky symmetry. They may also offer some degree of convenience and mathematical or computational tractability.

Distributional assumptions on unobserved variables may also be nonparametric, for example imposing stochastic or mean independence restrictions, or parametric, for example imposing normality. Likewise, they may offer some degree of tractability. However, care should be taken in specifying them in the context of the application at hand. For instance, as discussed in Pakes (2010, 2014), there is an important distinction between (i) errors due to differences between agents' expected payoffs at the time their decisions are taken and the payoff realized ex-post, and (ii) structural errors known to agents when their decisions are made, but not observed by the econometrician. Expectational errors are mean independent of variables in the agent's information set. Structural errors are not, and may result in endogeneity, for example due to selection. Alternatively, unobservables may be due to measurement error in observed variables, or approximation error due to the use of convenient functional forms, such as a linear projections. The intended role of the unobserved variables in the model can be used to motivate the distributional restrictions placed on them. In conjunction with the functional form restrictions, these generate observable implications in the form of the moment equalities and inequalities that partially identify model parameters.

Econometric models have historically been largely limited to applications where point identification obtains or is assumed. Consequently, partial identification may remain unfamiliar to many applied researchers. To assist practitioners in using partially identifying models, we present a sample of applied papers that have used these techniques. We discuss the features of the models and data in each paper that are combined to produce empirical results. An important step at the beginning of any empirical project is to contemplate (1) what model or models are appropriate for the task at hand, and (2) given those models, what variation in the data—i.e. what properties of the distribution of observed variables in the population under study—can allow the researcher to learn about quantities of interest. We consider these questions in the context of each paper. We focus on applications that in our view have worked well and have enabled researchers to address

questions of substantive economic importance.

Specifically, we consider the following questions:

1. What are the parameters of interest in the application? What do the researchers hope to learn?
2. What elements of theory are brought to bear? What is the economic content of the maintained assumptions invoked throughout the analysis?
3. What additional assumptions could be used or have been used by others to achieve point identification? What is the additional content of these assumptions, and are there reasons why one might not want to make them?
4. What are the implications of the maintained assumptions? That is, how do the assumptions translate into observable implications – typically (conditional) moment equalities and inequalities – that can be used for estimation and inference?
5. Given the maintained assumptions, what features of the data can be helpful for generating informative bounds?

In order to address these questions in some level of detail within space constraints, we focus on a selection of papers, noting that there are several other excellent examples that could be discussed more closely. Our paper summaries are necessarily brief. Our primary focus lies in addressing the questions above. To do this we focus on identification analysis, rather than statistical inference. In practice researchers must also account for sampling variation. This is an important but conceptually distinct consideration that we come back to in Section 7.3.

We categorize the papers discussed by the type of application considered, noting commonalities where they occur. The papers are drawn from the labor economics and industrial organization literatures. In Section 2 we begin by considering applications that feature a simultaneous discrete or “entry” game that may allow multiple equilibria. We summarize the development of this literature to a point where detailed models of particular markets have been used to answer important policy questions. Section 3 considers papers that estimate models of demand using the same sorts of revealed preference assumptions that are often used in classical models to achieve point identification (e.g., logit models), but relaxing other, potentially questionable, assumptions with tenuous theoretical foundations. In section 4 we cover auctions. Here inequalities that come directly from auction theory have been used to bound the distribution of bidders’ valuations and other quantities of interest, such as optimal reserve prices and maximal seller profit, when

strong assumptions on bidding strategies or the independence of valuations are relaxed. Finally, in section 5 we consider the literature that estimates bounds to deal with selection, for example in the study of wage distributions and treatment effects. In section 6 we investigate the use of bounds in recovering wage distributions more closely, using numerical illustrations based on the National Longitudinal Survey of Women (NLSW) 1967 to compare the identifying content of different assumptions under different degrees of variation in the observable data. We illustrate how different models exploit exogenous variation in the data to learn about unconditional wage distributions, paying particular attention to the use of exclusion restrictions. In section 7 we summarize some of the findings from the literature to date. Section 8 suggests directions for future research and concludes.

2 Multiple Equilibria in Entry Models

In this section we consider the use of bounds to recover estimates of payoff functions from observed behavior in discrete games. Such games often admit multiple equilibria.⁴ Multiple equilibria can sometimes be an important feature of the phenomenon being studied, and may thus be undesirable to rule out. For example, in a binary action entry game with two potential entrants, the data may indicate that firm A entered a particular market and firm B did not. However, the opposite configuration, namely that firm B entered and firm A did not, may also have been an equilibrium, even though it was not played and therefore not observed in the data. This multiplicity raises a barrier to estimating the model using traditional methods such as maximum likelihood because, in contrast to games with a unique equilibrium, there is no longer a unique mapping from the parameter vector and unobserved payoff shifters to observed outcomes. Thus, the presence of multiple equilibria complicates – although does not rule out – the possibility of point identification.

Researchers have traditionally imposed additional restrictions to solve this problem. Heckman (1978) showed that these problems arose in a simultaneous equations probit model, among others, and proposed what he termed a principal assumption on the parameters to guarantee a unique solution. Another approach is to assume that when there are multiple equilibria, one is selected at random with some fixed probability (Bjorn and Vuong (1984)) or by way of an equilibrium selection mechanism explicitly specified as part of the model (Bajari, Hong, and Ryan (2010)). Alternatively, one can make assumptions on the nature of firm heterogeneity to ensure that the *number* of entrants is unique even though their identities are not (Bresnahan

⁴Another consideration in the literature on econometric models of discrete games is the possibility of certain configurations of the parameters resulting in non-existence of equilibrium. Due to space constraints we do not cover this possibility here, but for an overview of the various ways this problem has been dealt with in conjunction with the possibility of multiple equilibria, we refer to Chesher and Rosen (2012).

and Reiss (1990)); or one can assume that the firms make decisions sequentially, perhaps with a random ordering where the probabilities are to be estimated (Berry (1992)). These assumptions can be *ad hoc* and may be unrealistic, see Berry and Tamer (2006) for a more detailed discussion. Tamer (2003) showed that in the simultaneous binary game, equilibrium behavior allowing for multiple equilibria implies moment equalities and inequalities that can be used as a basis for estimation. Consequently it is now recognized that in general one can avoid the additional assumptions used to guarantee a unique equilibrium when they are not credible, and instead partially identify payoff parameters.⁵

Ciliberto and Tamer (2009) (Henceforth CT). Our first example considers a complete information, static entry game with application to the airline industry. The authors use cross-sectional data where each market observation is a unique city pair. They assume that each firm is present if and only if it makes positive profit from being in the market, given rivals' actions. The equilibrium condition is thus a revealed preference condition, since firms are assumed to make zero profit in the market if they are not active, but with the added caveat that each firm's profit depends on its rivals' actions. This equilibrium assumption produces a collection of inequalities that must be satisfied in each market, that are interdependent and must therefore be taken as a system of inequalities to be satisfied simultaneously. These inequalities are sufficient to place bounds on the parameters of the firm payoff functions.

To see the intuition behind the estimator, consider a simplified version where two firms have the following profit functions:

$$\begin{aligned}\pi_{1,m} &= \alpha'_1 X_{1,m} + \delta_2 Y_{2,m} + X_{2,m} \phi_2 Y_{2,m} + \epsilon_{1,m}, \\ \pi_{2,m} &= \alpha'_2 X_{2,m} + \delta_1 Y_{1,m} + X_{1,m} \phi_1 Y_{1,m} + \epsilon_{2,m},\end{aligned}\tag{2.1}$$

where $X_{1,m}$ and $X_{2,m}$ contain observed market and firm characteristics for firms 1 and 2, respectively. For each firm j , the binary variable $Y_{j,m}$ indicates whether firm j operates in market m . The unobserved variables $\epsilon_{j,m}$ are assumed to be structural errors, that is, components of firm j 's profits that are observed by the firms but not by the econometrician. The terms (δ_j, ϕ_j) are the focus of the study. They capture the effect firm j has on firm i 's profits. The objective here is not to specify or estimate the particular mechanism by which firm j 's presence has an effect on firm i 's profit, but instead to measure that effect.

The assumption that variables $\epsilon_{j,m}$ are known to the firms when making their decisions renders the

⁵Tamer (2003) also provided sufficient conditions for point identification in simultaneous equation binary outcome models allowing for multiple equilibria. We discuss this point further below.

game one of complete information. It is motivated by the idea that firms in this industry have settled at a long-run equilibrium. The industry has been in operation for a long time, and it is reasoned that the firms therefore have detailed knowledge of both their own and their rivals' profit functions. However, alternative information structures corresponding to different assumptions about the firms' knowledge of unobservables are possible. For instance, in models of incomplete information games, it is assumed that each player knows the unobservable components of its own payoff function, but not the unobservable components of its rivals' payoff functions. The players then form expectations over rivals' actions, and are typically assumed to maximize their expected payoffs in a Bayesian Nash Equilibrium.⁶ In addition, expectational errors could also be introduced through a component of payoffs unobserved to both the econometrician and the player, as in for example Dickstein and Morales (2013) who study a partially identifying single agent binary choice model with both structural and expectational errors. In the entry model considered by CT, both incomplete information and expectational errors could have been added, but these introduce the possibility that firms regret their entry decision after learning the realization of unobserved variables, which seems at odds with the idea that the market is in a long-run equilibrium. Alternatively, one could introduce approximation error, as considered by Pakes (2010, 2014). Then agents are still assumed to maximize their payoffs (or expected payoffs), but the functional form used for payoffs is interpreted as an approximation to the true payoff function, with approximation error comprising the difference between the two. Different applications and interpretations of such models motivate different assumptions with regard to unobserved heterogeneity.

The equilibrium and complete information assumptions of CT together imply that firm decisions satisfy:

$$\begin{aligned} Y_{1,m} &= 1[\alpha'_1 X_{1,m} + \delta_2 Y_{2,m} + X_{2,m} \phi_2 Y_{2,m} + \epsilon_{1,m} \geq 0], \\ Y_{2,m} &= 1[\alpha'_2 X_{2,m} + \delta_1 Y_{1,m} + X_{1,m} \phi_1 Y_{1,m} + \epsilon_{2,m} \geq 0]. \end{aligned} \tag{2.2}$$

Consider the case where $\delta_i + X_{i,m} \phi_i < 0$ for $i = 1, 2$ (the economically interesting case where each firm has a negative effect on the other firm's profits). It is straightforward to see that multiple equilibria in the identity of firms will exist when $-\alpha'_i X_{i,m} \leq \epsilon_{i,m} \leq -\alpha'_i X_{i,m} - \delta_{3-i} - X_{3-i,m} \phi_{3-i}$ for $i = 1, 2$, since in this range both $(Y_{1,m}, Y_{2,m}) = (0, 1)$ and $(1, 0)$ will satisfy condition 2.2. Thus the probability of the outcome $(Y_{1,m}, Y_{2,m}) = (1, 0)$ cannot be written as a function of the parameters of the model, even given distributional assumptions on the unobserved payoff shifters. This problem can be circumvented by specifying

⁶Multiple equilibria and the ensuing complications for identification may still arise. See however Aradillas-López (2010) for sufficient conditions for a unique equilibrium in binary outcome models, and Seim (2006) for an application to the video retail industry additionally allowing for endogenous product choice.

an equilibrium selection rule that uniquely determines the outcome, perhaps as a function of observed and unobserved exogenous variables. Yet theory often provides little guidance on how to specify such a selection mechanism. Moreover, the existence of multiple equilibria may be a salient feature of reality, and may therefore be undesirable to assume away. The problem becomes more complex – and the configurations of action profiles that may possibly arise as multiple equilibria greater – with more than two firms or more than two actions.

Despite the multiplicity issue, and without making further assumptions to remove the multiplicity of equilibria, the model implies lower and upper bounds on the outcome probabilities that can be used for estimation. In the two-firm case above, for example, the lower bound on the probability of observing outcome $(1, 0)$ is the probability that $(1, 0)$ is the unique outcome of the game. The upper bound is the probability that $(1, 0)$ is an equilibrium outcome, either uniquely or as one of multiple equilibria, as for example when both $(1, 0)$ and $(0, 1)$ are equilibria. That is, suppressing market subscripts for brevity and defining θ to include (α, δ) as well as any parameters governing the distribution of ϵ :

$$LB_{(1,0)}(\theta; x) \leq \Pr(Y = (1, 0)|X = x) \leq UB_{(1,0)}(\theta; x) \quad (2.3)$$

where

$$LB_{(1,0)}(\theta; x) \equiv \Pr(\epsilon_1 > -\alpha'_1 x_1, \epsilon_2 < -\alpha'_2 x_2) + \Pr(\epsilon_1 > -\alpha'_1 x_1 - \delta_2 - x_2 \phi_2, -\alpha'_2 x_2 \leq \epsilon_2 < -\alpha'_2 x_2 - \delta_1 - x_1 \phi_1)$$

$$UB_{(1,0)}(\theta; x) \equiv LB_{(1,0)}(\theta; x) + \Pr(-\alpha'_1 x_1 \leq \epsilon_1 < -\alpha'_1 x_1 - \delta_2 - x_2 \phi_2, -\alpha'_2 x_2 \leq \epsilon_2 < -\alpha'_2 x_2 - \delta_1 - x_1 \phi_1).$$

Similar expressions can be derived for the probability $\Pr(Y = (0, 1)|X = x)$. In the simple two-firm case it may be possible to derive these functions analytically given an assumed distribution of the error terms. Otherwise they can be computed *via* numerical integration or simulation.

The authors extend these ideas to show that, in general games with many players, if the joint distribution of ϵ is known up to a finite dimensional parameter vector which is part of θ , then equilibrium conditions can again be used to derive lower and upper bounds on the choice probabilities. This generates a conditional moment inequality model which can generally be written:

$$\forall x \in \text{Supp}(X), \quad \mathbf{H}_{\text{LB}}(\theta; \mathbf{x}) \leq \Pr(\mathbf{y}|X = x) \leq \mathbf{H}_{\text{UB}}(\theta; \mathbf{x}) \quad (2.4)$$

where $\Pr(\mathbf{y}|X = x)$ is a 2^K vector of conditional choice probabilities for the different possible outcomes in each market, and $\text{Supp}(X)$ is the support of X . Any parameter vector θ that could have possibly generated the distribution of observed variables in the data must satisfy these inequalities, and the set of θ values that satisfy these inequalities therefore provides bounds on θ .

CT use methods from Chernozhukov, Hong, and Tamer (2007) to perform estimation and inference based on the inequalities in (2.4). They estimate two versions of the profit equation: one where ϕ_i is constrained to equal 0 for $i = 1, 2$ and a second without this constraint. The estimates provide evidence of heterogeneity across airlines in their profit equations. Market size is found to have similar effects on all firms' profit, but market distance has a positive effect on the profit of large national carriers, and a negative effect on the profit of Southwest. Competitive effects of each firm's entry decision on their rivals' profit are also found to be heterogeneous. For example, the competitive effect of large firms on other large firms is found to be smaller than that of large firms on low cost firms. The estimates further suggest that the (negative) competitive effect of an airline is increasing in its airport presence, a measure of the number of markets served by a firm out of an airport. CT also conduct a policy experiment, using their model to predict the effect of repealing the Wright Amendment, which restricted flights out of Dallas Love airport in order to encourage growth at Dallas/Forth Worth. They find that repealing the amendment would lead to a large increase in the number of markets served out of Dallas Love.

We now consider what features of the data in entry games such as that used by CT can help to obtain relatively tight bounds. The method generates a pair of bounds for each x and for every potential outcome; in the two-firm case these are $((0, 0), (1, 0), (0, 1), (1, 1))$. Tighter bounds will be generated if we can rule out a larger number of possible values for θ : if for more values of θ , an appropriate estimator of $\Pr(\mathbf{y}|X = x)$ is above $\mathbf{H}_{\text{UB}}(\theta, \mathbf{x})$ or below $\mathbf{H}_{\text{LB}}(\theta, \mathbf{x})$. Since the inequalities are required to hold for every value of x on the support of X there is some intuition that the larger is the support of X , the greater the collection of inequalities that may be violated. However the intuition is complicated by the simultaneous nature of the game and the fact that varying x may affect both $\Pr(\mathbf{y}|X = x)$ and the simulated bounds. The question therefore is: what type of variation should be expected to be particularly helpful?

We can gain some insight from Tamer (2003), who provides sufficient conditions for point identification in a two-player entry model. These include (i) rank conditions as well as (ii) a condition that the profit function of one of the firms contains an exogenous payoff shifter with non-zero coefficient, that has full support on the real line conditional on all other exogenous variables, and that does not enter the other firm's payoff function. This is similar to the use of exclusion restrictions in the classical linear simultaneous equations

model. The support condition guarantees that the conditional probability of this firm entering can be made arbitrarily close to zero or one by making this particular payoff shifter arbitrarily big or small. In the limit as this variable approaches $\pm\infty$ the entry decision of the other firm follows a simple binary probit specification. This reasoning can then be formalized to establish point-identification of the parameters of that firm's payoff function, and subsequently the other firm's payoff parameters.

Unfortunately, the occurrence of firm-specific payoff shifters with support ranging from $-\infty$ to $+\infty$ is rare. Also, with $J > 2$ firms an analogous point identification proof would require $J - 1$ such variables and rely on the possibility that they could all take arbitrarily large or small values simultaneously. Nonetheless, the idea behind the proof can provide some guidance as to what particular kind of exogenous variation should be helpful in generating smaller identified sets. First, as CT show, variation in excluded variables, i.e. x_i that affect one firm's profit function and not the others', can help shrink the identified set. They use measures of airport presence and firm-specific costs for each market. The intuition is again that, as the variable changes, it affects the probability that the particular firm enters the market but impacts other-firm choices only indirectly through firm i 's action. Second, if there are observed values of the exogenous variables that make entry by some firms very likely and/or entry by some firms very unlikely, this can be particularly helpful. Variables that could generally have these types of properties include firm-specific cost shifters. In chain store entry models, examples may include distance to nearby stores of the same chain, or distance to a store's distribution center, see for example Jia (2008), Holmes (2011), Aradillas-López and Rosen (2013), and Aradillas-López and Gandhi (2013).

CT were able to draw meaningful economic inference in their application. Yet, there is scope for further refinement of the bounds estimated in CT, and more generally in the literature on econometric models of discrete games. Heuristically, this is because the equilibrium assumptions imposed imply additional inequalities involving the probability that any particular collection of action profiles are equilibria simultaneously. This leads to additional inequalities further to those in (2.4) that bound the sum of conditional probabilities that different action profiles y are played. Beresteanu, Molchanov, and Molinari (2011) provide a characterization of all the inequalities resulting from the restrictions, i.e. they characterize sharp bounds on model parameters, in a class of econometric models that includes entry models of this sort. They show in specific numerical examples the extent to which these inequalities can lead to tighter bounds. Thus, there is potential to further exploit variation in the data in particular applications to obtain tighter set estimates in models of discrete games. Nonetheless, the CT bound estimates proved sufficiently informative to draw useful empirical conclusions with their data. This is a more general point that we come back to in our discussion of sharp

bounds in Section 7.2.

Several related papers draw on the intuition developed in CT to analyze discrete entry games. Grieco (2014) makes the interesting addition of assuming that the players observe part but not all of the error term ϵ . He thus considers two errors, one observed by all the players and another observed by only one player, but neither of which is observed by the econometrician. Aradillas-López and Rosen (2013) study an extension of the simultaneous binary outcome model to simultaneous ordered outcome models. That is, the authors consider models where players simultaneously choose among ordered alternatives such as how many stores to open in a market, or how many daily flights to operate on a given route. As in CT, each player's action is allowed to have an effect on its rivals' payoffs, the game considered is one of complete information, and pure strategy Nash Equilibrium is used as the solution concept. They apply their model to study the number of store fronts operated by home improvement chains in the United States (Home Depot and Lowes'). Their application employs a parametric version of the model for which they establish point identification of coefficients on demographic variables in the firms' payoff functions. They obtain point estimates for these, as well as bounds on strategic interaction parameters that capture the effect of each firm's action on its rival's profit. They implement a novel inference method for the parameters of their model, and use their results to bound objects such as the probability with which an equilibrium is selected when it is one of multiple equilibria, and to investigate what would happen under counterfactual scenarios. For instance, their results indicate that if Home Depot were to operate as a monopolist, many markets in which Lowes' currently operates would go underserved, while if Lowes' were a monopolist, it would operate in nearly all markets where at present only Home Depot operates.

Other Models for Multiple Equilibria. We next review a body of papers that use a different approach to analyze discrete games. Building on the methods developed in Pakes (2010) and Pakes, Porter, Ho, and Ishii (2015), these papers use revealed preference inequalities as one input into a carefully constructed structural model of both demand and supply in particular markets. This requires detailed attention to the error structure, and in particular assumptions about the extent of agents' knowledge of them at the time their decisions are taken. Given the carefully specified model, the assumption that agents make decisions to maximize their expected payoffs produces revealed preferences inequalities that serve to partially identify some of the model parameters, while point-identifying others. The use of carefully specified structural models in this way can lead to results of direct policy relevance.

Our first example is Eizenberg (2014) which investigates the welfare implications of rapid innovation in

central processing units (CPUs) in the personal computer (PC) market. The objective is to assess whether the rapid removal of existing basic products from the market, when new products are introduced, reduces surplus. This is a question that cannot be answered by estimating a profit equation like that in CT. Instead the author writes down and estimates a full model of the market. PC makers are assumed to play a two-stage game. In the first stage they observe realizations of shocks to fixed costs and then simultaneously choose which of a discrete number of vertically differentiated CPUs to offer to consumers. In the second they observe realizations of demand and marginal cost shocks to the PCs chosen in stage 1 and engage in a simultaneous price-setting game. The author estimates a static random coefficient logit model of household demand for PCs based on Berry, Levinsohn, and Pakes (1995) and uses a Nash Bertrand pricing equilibrium assumption to estimate a marginal cost function in a way that is standard in this literature.

The remaining step is to use the observed product line configurations to obtain estimates of the fixed costs associated with different products. This is where, as in CT, the existence of a multiplicity of equilibria makes it impossible to write down a well-defined likelihood function without making further, possibly unpalatable assumptions. Again the issue is resolved through a revealed preference assumption. However, rather than using these conditions to generate bounds on the *probability* of the observed outcome, Eizenberg works directly with the revealed preference conditions themselves as in Pakes, Porter, Ho, and Ishii (2015).

The author assumes that the fixed cost of offering product j is $F_j = F^d + \nu_j$ with $E[\nu_j | j \in I^d] = 0$, where I^d is the set of firm d 's potential products: those that are offered and those that are not. That is, fixed costs are the sum of a mean (to be estimated) and a mean-zero error term. The assumptions regarding the timing of the model imply that the error ν_j is known to firms when making choices but not observed by the econometrician (it is a structural error). The revealed preference conditions are generated from an assumption that no firm can increase its expected profit by unilaterally changing its first-stage product choices, taking into account the impact of that deviation on second-stage prices (but not on rivals' product choices because of the simultaneous-moves assumption in stage 1). An upper bound on the fixed cost of product j that is offered by firm d as part of its overall observed product set A_d is generated from the condition that a deviation which eliminates one of d 's observed products must not be profitable in expectation:

$$F_j = F^d + \nu_j \leq E_e[VP_d(A_d; e, \theta_0) - VP_d(A_d - 1_d^j; e, \theta_0)] = \bar{F}_j(\theta_0), \forall j \in A_d \quad (2.5)$$

where the expectation is over the shocks to demand and marginal costs that are realized in stage 2, 1_d^j is a

vector of the same length as A_d in which the j^{th} entry is equal to 1 and the others are 0, and VP_d is the firm's variable profit. An analogous *lower* bound can be defined for every product j that was *not* offered since a deviation that adds j to the firm's portfolio must not be profitable in expectation:

$$F_j = F^d + \nu_j \geq E_e[VP_d(A_d + 1_d^j; e, \theta_0) - VP_d(A_d; e, \theta_0)] = \underline{F}_j(\theta_0), \forall j \notin A_d. \quad (2.6)$$

Note that, in contrast to earlier papers, the model has direct implications for the properties of the unobservables. The shocks to demand and marginal costs (e_j) are not known to firms when they make their product choices and are therefore non-structural errors that are mean-independent of the firm's information set in this first stage of the model. In contrast, ν_j is a structural error. It introduces a selection issue because products that were introduced were likely to have had different ν_j terms than those that were not. Thus applying a conditional expectation to 2.5, for example, generates:

$$F^d + E[\nu_j | j \in A_d] \leq E[\bar{F}_j(\theta_0) | j \in A_d] \quad (2.7)$$

where the second term is likely to be non-zero. The author addresses the problem by imposing a bounded support assumption, under which bounds are obtained on the entry cost for *every* potential product j in the choice set, which hold whether the product is offered or not. He then applies an unconditional expectation to obtain:

$$ELB_j(\theta_0) \leq F^d \leq EUB_j(\theta_0), \forall j \in I^d. \quad (2.8)$$

The estimated set for the fixed cost parameter is then obtained by replacing the true parameter vector θ_0 with its estimator $\hat{\theta}$ from estimation of the demand and pricing models, and using the sample average in place of the population mean. Confidence intervals are generated following Imbens and Manski (2004). Counterfactual analyses indicate that, in part due to elimination of older products, the bulk of the welfare gains from innovation are enjoyed by a limited group of price-insensitive consumers. However, had a social planner reinstated the older products, consumer welfare gains would be largely offset by producer losses, given the estimated range for fixed costs.

There are other recent papers that use moment inequalities to estimate firm fixed costs in discrete games. Like Eizenberg (2014) several of these papers make the point that particular changes in an industry (caused,

for example, by new innovations or by firm mergers) can have indirect effects on firms' choices of product mix that are as important for consumer welfare as their direct effects. Quantifying these indirect effects requires estimating the fixed costs of introducing or re-positioning a product, and these are estimated using moment inequalities implied by revealed preference. Nosko (2014) considers the CPU market (upstream from the PC manufacturers examined by Eizenberg), and examines how the two dominant firms (Intel and AMD) use product line decisions to increase their returns to innovation. His two-stage model mirrors that of Eizenberg, although he avoids the selection problem by assuming that the fixed costs of product repositioning do not vary across products. The unobservables in the model are assumed to be non-structural: demand and cost shocks that are not known to firms when product choices are made and an additional expectational error. His counterfactual simulations indicate that the innovator, Intel, benefits from its innovation both because of its new high-end products and because the innovation reduces the marginal cost of producing CPUs at a given quality level, enabling the firm to restructure the remainder of its product line and move its competitor to lower portions of the quality spectrum. A monopolist would have fewer incentives to innovate than the duopolists in the data because of this business-stealing effect.

Wollman (2014) considers the commercial vehicle segment of the auto industry, focusing on the potential implications of the \$85B government bailout of the industry in 2008-9. The estimation framework is again a two-stage model of supply that uses an estimated demand model as an input. In this case the estimated fixed costs are the costs of adding and retiring models. The model is used to simulate firms' product choices had GM and Chrysler been liquidated, or acquired by a major competitor, rather than being bailed out. The assumptions on the unobservables are similar to those in Eizenberg (2014). Sunk costs include an additively separable disturbance, part of which is not known to firms (a non-structural error), and part of which is known (a structural error); the latter is again handled by taking an unconditional expectation to avoid any issues with selection bias. The simulations indicate that, had the firms been liquidated, the surviving firms' profit increases would have been high enough to induce them to introduce new products. The probability of not purchasing a vehicle in the case of liquidation is predicted to rise by 50% for the most affected buyers when product entry is ignored, but only 14% when it is accounted for.

Earlier papers with a similar focus on firms' product selection choices include Ho (2009), which uses revealed preference conditions as an input to modeling the formation of the hospital networks offered by U.S. health insurers. Holmes (2011) uses data on rollout of Wal-Mart store openings, and the now-familiar revealed preference approach, to investigate the trade-offs between the costs and benefits of high store density. Ho, Ho, and Mortimer (2012) and Lee (2013) both use moment inequalities as a means to estimate bounds

on firm fixed costs, as an input to modeling product choices in a particular industry, without making the assumptions needed to ensure point identification of supply.

Taken together, these papers demonstrate that the industrial organization literature using moment inequalities to estimate firm fixed costs as one input to a model of a market has the potential to substantially advance the debate on the welfare effects of market structure changes. In many industries, product mix can be changed more easily than price, and these changes should clearly be taken into account in any policy analysis. The methods developed in these papers allow practitioners to use revealed preference to bound fixed costs, an important input to models that consider these issues.

3 Demand, Utility Maximization, and Revealed Preference

The empirical literature that estimates discrete choice models of demand, for example using a logit framework (e.g. McFadden (1974) at the individual level and Berry (1994) and Berry, Levinsohn, and Pakes (1995) for market demand), has a revealed preference foundation. The standard approach is to write down a utility equation for the decision-making agent and assume that agents make utility-maximizing choices. Authors then add further assumptions, for example regarding the distribution of unobservables and the information structure of the model, in order to achieve point identification. These additional assumptions are not always palatable for particular applications. Several recent papers have demonstrated that a lot can often be learned by maintaining the revealed preference approach based on theory, relaxing the more questionable assumptions, and generating bounds on the parameters of interest.

One example of such a paper is Ho and Pakes (2014) which studies hospital referrals for labor and birth episodes in California in 2003. In this setting some commercial insurers utilized particular types of contracts (capitation contracts) under which referring physician groups faced incentives to reduce patients' hospital costs, either because they received a fixed payment to cover the medical costs of their patients (including hospital costs), or because they shared in any savings made relative to some pre-agreed benchmark for hospital costs. The extent to which these contractual arrangements generated cost reductions is an important policy question given their similarity to contracts being implemented by Accountable Care Organizations in the wake of the Patient Protection and Affordable Care Act (2010).

Ho and Pakes analyze hospital referral choices for patients enrolled in six California health insurers that use capitation contracts to different extents. The objective is to understand whether patients enrolled in high-capitation insurers tend to be referred to lower-priced hospitals, all else equal, than other same-severity

patients, and whether quality of care was affected. Answering these questions requires the authors to estimate a hospital referral function (separately for each insurer) that is analogous to the utility equation estimated in a standard discrete choice demand analysis. The referral function whose maximum determines the hospital (h) to which patient i of insurer π is allocated is assumed to take the additively separable form:

$$W_{i,\pi,h} = \theta_{p,\pi}p(c_i, h, \pi) + g_\pi(q_h(s), s_i) + \theta_d d(l_i, l_h) \quad (3.1)$$

where $p(c_i, h, \pi)$ is the price insurer π is expected to pay at hospital h for a patient who enters in condition (or price group) c_i ; s_i is a measure of the severity of the patient's condition; $q_h(s)$ is a vector of perceived qualities of hospital h , one for each severity; $g_\pi(q_h(s), s_i)$ is a plan- and severity-specific function which determines the impact of the quality of a hospital for the given severity on the choice of hospitals; and $d(l_i, l_h)$ is the distance between patient and hospital locations. The severity groups s_i are aggregates of the c_i (so there is variance in price conditional on severity). The first coefficient of interest is $\theta_{p,\pi}$; a more negative coefficient implies a larger response to price by the referring physician. The authors also evaluate the trade-offs made between price, quality and distance that are implied by the overall estimated equation.

A standard discrete choice approach would add an additively separable error term $\varepsilon_{i,\pi,h}$ and assume that it was a structural error, known to the decision-makers when hospital choices were made and distributed i.i.d. Type 1 extreme value. A parametric assumption would then be made on the form of $g_\pi(\cdot)$, and the (now point-identified) equation would be estimated via maximum likelihood. However the authors wish to relax the usual distributional assumption on $\varepsilon_{i,\pi,h}$, which is not based on consumer theory. In addition there are two other impediments to the standard point-identifying approach in this setting. First, in its most general form, the $g_\pi(\cdot)$ term should be allowed to differ arbitrarily across insurers, across sickness levels, and across hospitals. This allows particular hospitals to have higher quality for some sickness levels than for others. If this variation is not fully accounted for in $g_\pi(\cdot)$, the residual variance will create an additional unobservable representing unobserved quality which, if it is correlated with price, is likely to cause a positive bias in the price coefficient. However there are over 100 patient severity groups and almost 200 hospitals, so estimating $g_\pi(q_h(s), s_i)$ as a fully flexible interaction between severity and hospital fixed effects generates an incidental parameters problem similar to that described in Neyman and Scott (1948) which makes coefficient estimates very unreliable. Chamberlain (1980)'s conditional likelihood estimator is not suitable for the problem because its computational burden grows with the combinatorial formula for the number of ways patients in a severity group can be divided across hospitals conditional on the given number of patients and hospitals. The number

of patients and hospitals in the study is too large to make this feasible.

The second issue is that the expected price that generates hospital choices is inherently unobservable. The variable needed for the analysis is the price that the decision-makers expect the insurer to pay for a patient entering the hospital with a given condition c_i . The authors assume that expected prices are on average correct, making the average realized price for the hospital-insurer-condition triple an appropriate estimator of the expected price. However these predictions are only assumed to be correct on average, so the estimation methodology must allow $\varepsilon_{i,\pi,h}$ to be interpreted as non-structural, mean-zero measurement error in price. The logit model does not admit this interpretation.

The authors resolve these issues using a partially identifying model based on a revealed preference inequality. The inequality follows precisely the same logic as the inequalities that define the logit model. It is implied by assuming that the chosen hospital is preferred to feasible alternative hospitals. The authors define $g_\pi(q_h(s), s_i)$ as a fully flexible set of interactions between hospital and severity fixed effects, and given this flexibility, assume that the only remaining unobservable to be added to (3.1) is price measurement error.⁷ They consider all couples of same-insurer, same-severity (s_i) patients whose chosen hospitals differ but both of whose choices were feasible for both agents. Within each couple they sum the inequalities obtained from the fact that each patient's hospital is preferred to the hospital attended by the other. Since the severity-hospital interactions (the $g_\pi(\cdot)$) from the two inequalities are equal but opposite in sign, when they sum the inequalities the interaction terms difference out. Revealed preference implies that this sum is positive, and this constrains the remaining parameters.

More formally: for notational simplicity let $\Delta x(i, h, h') = x_{i,h} - x_{i,h'}$ for any variable x , and $\Delta W(i, h, h') = W_{i,\pi,h} - W_{i,\pi,h'}$. Let the average realized price for group c_i at hospital h be $p^o(c_i, h, \pi)$, the agents' expected price be $p(c_i, h, \pi)$, and the difference between them generate the measurement error term $\varepsilon_{i,\pi,h}$. Substituting into equation (3.1) for a same-plan same-severity couple (i, i') who could have chosen each other's hospital and are in different price groups, normalizing the distance coefficient (a free parameter) to equal -1 and dropping π subscripts for simplicity, the revealed preference inequality becomes

$$0 \leq \Delta W(i, h, h') + \Delta W(i', h', h) = \tag{3.2}$$

⁷They also allow for classification error in the quality-severity interactions $g_\pi(\cdot)$; we omit this from our discussion for simplicity. There is a question of whether an additional structural error, beyond the $g_\pi(\cdot)$ term, could be useful to account for remaining factors observed by decision-makers but not by the econometrician. However, the authors conducted multiple tests for the presence of such an error, and found no evidence of its importance. We note that a special case of the method developed in Pakes and Porter (2015) allows for a structural error but no measurement or approximation error, and does not require distributional assumptions on the structural error.

$$\theta_p \left[\Delta p^o(c_i, h, h') + \Delta p^o(c_{i'}, h', h) \right] - \left[\Delta d(l_i, l_h, l_{h'}) + \Delta d(l_{i'}, l_{h'}, l_h) \right] - \Delta \varepsilon_{i,h,h'} - \Delta \varepsilon_{i',h',h}.$$

Using mean independence of the measurement error $\varepsilon_{i,\pi,h}$ with choice of hospital as well as distance between patients and hospitals, this translates to a conditional moment inequality:

$$\forall (h, h'), \quad E_p[\Delta W(i, h, h'; \theta_p) + \Delta W(i', h', h; \theta_p) | \Delta d_{ihh'}, \Delta d_{i'h'h}, h, h'] \geq 0, \quad (3.3)$$

where $\Delta d_{ihh'}$ and $\Delta d_{i'h'h}$ denote $\Delta d(l_i, l_h, l_{h'})$ and $\Delta d(l_{i'}, l_{h'}, l_h)$. $E_p[\Delta W(i, h, h'; \theta_p) | \Delta d_{ihh'}, \Delta d_{i'h'h}, h, h']$ is the expected utility difference (over price) for patients who were referred to hospital h but could have visited some other h' , $E_p[\Delta W(i', h', h; \theta_p) | \Delta d_{ihh'}, \Delta d_{i'h'h}, h, h']$ is the analogous expression for patients who visited h' instead of h , and the argument θ_p has been added to the function $\Delta W(\cdot)$ to make the dependence explicit.

The authors use a sample analog method for estimation. They take averages of equation (3.2) across switches between patients who chose hospital h and those who had the same severity and insurer, but a different price group, and who could have chosen h but chose another hospital. The assumptions on the price measurement error term - that it is mean zero conditional on the patient's chosen plan and hospital and independent of hospital and patient location - imply that these errors average out when they take averages across switches, generating a simple inequality that defines a bound on the price coefficient in terms of observables (price and distance differences). The inequality can be written:

$$\forall h, \quad \sum_{s, h' > h} w(h, h', s) \left[\theta_p \left(\Delta \bar{p}^o(h, h', s) + \Delta \bar{p}^o(h', h, s) \right) - \left(\Delta \bar{d}(h, h', s) + \Delta \bar{d}(h', h, s) \right) \right] \geq 0 \quad (3.4)$$

where $\Delta \bar{x}(h, h', s)$ is the average of any difference Δx over switches for the relevant (h, h', s) and the weights $w(\cdot)$ are the fractions of switches for that h that are contributed by each (h, h', s) triple. The moments generate lower bounds if the expected value of $\left(\Delta \bar{p}^o(h, h', s) + \Delta \bar{p}^o(h', h, s) \right)$ is positive and upper bounds otherwise.

Additional inequalities, and therefore bounds, can be generated by multiplying each inequality by an instrument z that has the same sign for all observations, that is known by the agents when their decisions are made, and that is mean independent of the measurement error. The authors use as instruments the positive and negative parts of the distance differences: $\Delta d(l_i, l_h, l_{h'})_+ \equiv \max\{d(l_i, l_h, l_{h'}), 0\}$, $\Delta d(l_i, l_h, l_{h'})_- \equiv \min\{d(l_i, l_h, l_{h'}), 0\}$, and analogously, $\Delta d(l_{i'}, l_{h'}, l_h)_+$, $\Delta d(l_{i'}, l_{h'}, l_h)_-$.

The question of data properties that assist with generating informative bounds is fairly transparent

from (3.4), which defines a moment for every h . Exogenous variation in patients' distances from different hospitals, and in severity, are both potentially important. Different pairs of hospitals may have patients with different sickness levels, living at different relative distances and facing different relative prices; and these factors may be correlated with hospital referrals. If so, this will provide variation in the expected values of $\left(\Delta\bar{p}^o(h, h', s) + \Delta\bar{p}^o(h', h, s)\right)$ and $\left(\Delta\bar{d}(h, h', s) + \Delta\bar{d}(h', h, s)\right)$ as we move across hospitals, which will be potentially useful in generating more informative bounds. The authors demonstrate in the paper that in fact there is a good deal of variation in prices as they condition on different values of severity and distance, and severities and distances vary across hospitals h .

It turns out that no value of θ_p satisfies all the inequality constraints in the data, so the estimation algorithm produces a point estimate: the value of the parameter that minimizes the sum of squares of the negative part of the moments. This is unsurprising given the large number of inequalities used (between 78 and 285 per insurer); sampling error can easily lead to the greatest of the lower bounds being above the smallest of the upper bounds in finite samples. The method used for inference (taken from PPHI) allows for set identification and the authors emphasize the resulting confidence intervals rather than the associated point estimate. The results are informative. The confidence intervals on the price coefficients are negative and are ordered with respect to insurers' capitation rates. That is, the price paid by insurers to hospitals does seem to affect referrals, and the price response is more elastic for insurers whose physician groups are more highly capitated.

In a second step the authors return to equation (3.1) and use the estimated bounds on the price coefficient to generate bounds on the quality terms $g_\pi(\cdot)$. They examine how the trade-offs between price, quality and distance vary with capitation rates. They find that highly capitated more price-sensitive plans tend to send their patients longer distances to obtain the same quality service at a lower price. The authors argue that these results suggest the use of capitation contracts in Accountable Care Organizations is likely to reduce costs and unlikely to result in a reduction in quality of care.

This paper demonstrates how partial identification methods can be applied to discrete choice models of consumer demand. The authors could achieve point identification by making more assumptions: for example in a robustness test they parameterize $g_\pi(\cdot)$ as a function of observables and interpret the error term as an i.i.d. Type 1 extreme value variable observed by the decision-maker but not the econometrician, rather than price measurement error, and thereby generate a multinomial logit specification that is straightforward to estimate. However, this requires a strong distributional assumption on the unobservables and also assumes away the potential endogeneity and price measurement issues described above, producing suspect results

with estimated price coefficients that are negative but much smaller in magnitude than in the main analysis.

Partial identification has been usefully applied in several other demand-related settings by relying on revealed preference arguments based on the assumption that agents make an optimal choice from a menu of options. Manski (2007) considers a setting of discrete choice, where individuals are observed to choose their most preferred alternative from various choice sets, and the task is then to predict what choice frequencies would occur under some counterfactual choice set. Distributional restrictions on unobserved tastes are not imposed, and bounds are obtained which are sensitive to both how much structure is placed on the determination of individuals' preference orderings (e.g. through hedonic assumptions), and how much variation is observed in individuals' choice sets. Similar logic has also been used to nonparametrically bound demand responses to counterfactual changes in prices or income using revealed preference arguments, see Blundell, Browning, and Crawford (2008), Hausman and Newey (2011), Blundell, Kristensen, and Matzkin (2014), and Adams (2014) for some examples. In these papers, the revealed preference inequalities are shown to have useful identification content, in that tight bounds can be obtained using the available data. Partial rather than point identification typically obtains due to the use of nonparametric functional forms for demand and/or a desire to leave unobserved heterogeneity in preferences unrestricted. Both of these concerns are important practical considerations, because economic theory does not give guidance on functional forms or the distribution of unobserved heterogeneity. For related papers on testing various implications of rationality through the use of revealed preference inequalities, see for example Hoderlein and Stoye (2014) and Kitamura and Stoye (2013).

4 Auctions

Several recent papers have estimated partially identified parameters in econometric models of auctions, including for example Haile and Tamer (2003), Tang (2011), and Aradillas-López, Gandhi, and Quint (2013). They have each focused on quantities of economic interest in the auction literature such as optimal reserve prices, seller profit, and bidder surplus, as well as questions of optimal auction format. They use partial identification to address important questions of substantive empirical content.

The essential components of these models comprise two types of assumptions, name those on (i) the mapping from bidders' unobservable information about their valuations to observable bidding data, such as equilibrium bidding assumptions, and (ii) bidders' information about their valuations for the object at auction, and the stochastic relation between components of this information. Assumptions about bidder

information in the auction model – for example whether bidders know their own valuations or only imprecise signals of their valuations, and whether these valuations include common components or may be otherwise correlated – dictate assumptions placed on unobserved heterogeneity. As was the case in previous sections, the assumptions made on the agents’ behavior, and on the properties of the unobservables and their roles in agents’ information sets, both play a role in deriving observable implications (typically inequalities) that can bound parameters of interest.

The first paper to consider bounds using an econometric model of auctions is Haile and Tamer (2003), henceforth HT, who focus on open outcry English auctions. They maintain a familiar information structure from the auction literature, in which symmetric bidders are assumed to have independent private values for the object being auctioned. That is, each bidder’s valuation is assumed to be an independent draw from common distribution $F(\cdot)$. Bidders know their own valuations when placing bids, but these are not observed by the econometrician.

HT motivate their study by pointing out that in real life ascending open outcry English auctions differ from the stylized models used to characterize equilibrium behavior. A common idealized model used to study such auctions is the button auction in which each bidder presses a button as the auction price rises continuously on a display in front of her, lifting her finger from the button when she is no longer willing to win at the given price. In practice we do not observe precisely the point at which each bidder drops out. Some bidders may not bid at all and there may be discrete bid increments. As HT put it, “...while the standard auction model serves well in illuminating strategic forces that arise in an English auction, it may serve poorly in providing an exact interpretation of bidding behavior.”

Thus HT adopt a very simple and appealing strategy that relaxes the equilibrium restrictions derived from the idealized auction model. Specifically, they instead assume:

Assumption 1: Bidders do not bid more than they are willing to pay.

Assumption 2: Bidders do not allow an opponent to win at a price they are willing to beat.

These assumptions are satisfied in the equilibrium of the standard auction model. But they also hold in other equilibria and are robust to small perturbations in the rules of the game, such as discrete rather than continuous bid increments.

The theory that these assumptions impose is essentially no more than revealed preference. Let bidder j have valuation V_j for the item being auctioned, and suppose her final bid at the auction was B_j . Assumption

1 gives us the revealed preference inequality

$$B_j \leq V_j.$$

The second assumption places an upper bound on bidder j 's valuation only if j is not the winning bidder,

$$V_j \leq \bar{B} + \Delta,$$

where \bar{B} denotes the highest bid and Δ the minimum bid increment.

HT then show that taken together these inequalities imply inequalities on the *order statistics* of observed bids $B \equiv (B_1, \dots, B_n)$ and $V \equiv (V_1, \dots, V_n)$, where n denotes the number of bidders in an auction, which is assumed to be independent of the bidder valuation distribution $F(\cdot)$. Let $B_{i:n}$ and $V_{i:n}$ denote the i^{th} order statistics of bids and valuations in an auction of n bidders, i.e. the i^{th} lowest bid and valuation, respectively. The implied inequalities on the order statistics are

$$B_{i:n} \leq V_{i:n}, \text{ all } n \text{ and all } i = 1, \dots, n, \quad V_{n-1:n} \leq B_{n:n} + \Delta, \text{ all } n.$$

Importantly, these inequalities are required to hold for any number of bidders n . The first of these additionally applies for any order statistic i from 1 to n . As a result, as will be shown below, the first inequality will deliver a moment inequalities for all possible (i, n) combinations, and the second inequality will provide a moment inequality for every possible n . HT then transform these inequalities to bounds on the distribution of order statistics at any given point v :

$$F_{i:n}(v) \leq G_{i:n}(v), \text{ all } i, n, \quad F_{(n-1):n}(v) \geq G_{n:n}(v - \Delta), \text{ all } n,$$

where $F_{i:n}(\cdot)$ and $G_{i:n}(\cdot)$ denote the CDF of the i_{th} order statistic of valuations and bids, respectively, in an n bidder auction.

Finally, using results from the order statistics literature and exploiting independent private values, there is for any i a one-to-one relationship between $F_{i:n}(v)$ and the primitive of interest, $F(v)$. Applying this transformation, HT obtain the following bounds on the valuation distribution at each v :

$$\phi(G_{n:n}(v - \Delta); n - 1, n) \leq F(v), \text{ all } n, \tag{4.1}$$

$$F(v) \leq \phi(G_{i:n}(v); i, n), \text{ all } i, n, \tag{4.2}$$

where $\phi(\cdot; i, n)$ is a strictly monotonic function for each (i, n) .⁸ Because these inequalities must hold for all n on the one hand, and all (i, n) on the other, there is the following representation of (4.1) and (4.2):

$$\max_n \phi(G_{n:n}(v - \Delta); n - 1, n) \leq F(v) \leq \min_{i,n} \phi(G_{i:n}(v); i, n). \quad (4.3)$$

That is, of the collection of bounds in (4.1) and (4.2) across all values of n and (i, n) , respectively, the implied bounds on $F(v)$ can be represented as simply the tightest lower and upper bounds.

The valuation distribution is the key primitive driving the answers to economically meaningful questions, such as the choice of optimal reserve price or how much is the maximal achievable expected seller profit. HT demonstrate the use of their method to estimate upper and lower bounds on the valuation distribution using data on U.S. timber auctions. The upper and lower bound estimates and corresponding confidence bands are quite tight. The bounds are consistent with the possibility that the shape of the underlying distribution is close to log normal, an assumption that had been used previously in the empirical auction literature for parametric estimation via maximum likelihood. To map bounds on the valuation distribution to bounds on the optimal reserve price additionally requires knowledge of the seller's cost, for which HT consider a range of plausible values. They find the implied bounds on the optimal reserve price are fairly wide, but the corresponding bounds on counterfactual seller profit can still be informative. The results suggest that the actual reserve price set in these auctions is considerably below the optimal reserve price. HT discuss the range of implied percentage gains in seller profit as a function of seller cost, while noting that revenue maximization is not the sole objective of the Forest Timber service that sells these timber tracts. Finally, HT use a simple semiparametric model in conjunction with their estimated bounds on valuations to bound the effect of auction characteristics on valuations. The estimated bounds are tight, of reasonable magnitudes, and provide support for the independent private values assumption.

It may seem surprising that HT were able to relax the standard equilibrium assumptions, replace them with what seem like much weaker behavioral assumptions, and still obtain empirically useful results in their application. What were the keys to the success of their approach?

First, auction theory has strong behavioral implications. The rules of the game are well understood, and much is known from the theory literature. This can and has been exploited in the empirical literature. Athey and Haile (2002) showed that in the symmetric independent private values model, under the standard equilibrium conditions, the valuation distribution is identified from knowledge of just one order statistic

⁸Specifically, $\phi(\cdot; i, n)$ is the quantile function of the beta distribution with parameters i and $n - i + 1$.

of the bid distribution. This would seem to suggest that there may indeed be scope to relax the usual equilibrium assumption while retaining a good amount of identifying power.

The second component of the success of the HT model is the use of variation in their data. Consider again the bounds (4.1) and (4.2). The lower bound, (4.1), derives from the implication of Assumption 2 that the second highest bidder would not prefer to have outbid the winning bidder. This logic holds no matter how many bidders n there are in each auction. In principle this bound could be more or less restrictive for different values of n . Indeed, which values of n make the inequality the tightest may differ for different points on the valuation distribution v . Thus we should expect variation in the number of bidders to be potentially helpful in producing a more informative lower bound. This could be true in particular if the relationship between the highest and second-highest observed bids were to differ significantly across different n .

The upper bound in (4.2) changes with the rank of the bid order statistic under consideration, i , as well as the number of bidders n . The idea here is that for any given n , each bid order statistic $i = 1, \dots, n$ provides an upper bound on $F(v)$ through Assumption 1, that no bidder bids more than her valuation. But this assumption says nothing about *how much* each bidder's bid lies below her valuation. Bids observed from different parts of the bid distribution may lie below the underlying bidder valuation by different amounts. That is, if there is heterogeneity in the bidding strategy used by bidders at different parts of the distribution, this may provide more or less restrictive upper bounds for different order statistics i . Furthermore, the magnitude of the difference between bids and valuations may change with n . As was the case for the lower bound, variation across number of bidders n may be useful in providing tighter bounds on $F(v)$.

The independent private values setup that is the context of HT has been widely used in applications. In practice though, the assumption that bidder valuations are mutually independent may not always seem reasonable. However, allowing for joint dependence of private values is more challenging. For instance, Athey and Haile (2002) Theorem 4 shows that even if we adopt the standard button auction model for bidder behavior, the joint distribution of valuations is not identified, even if a symmetry assumption is imposed. Aradillas-López, Gandhi, and Quint (2013), henceforth AGQ, thus take up the challenge of constructing an econometric model that allows for correlated private values.

Given the aforementioned nonidentification result of Athey and Haile (2002), even under the standard equilibrium assumption, it is perhaps unsurprising that AGQ perform a bounds analysis. They derive results that are both novel and empirically useful. As a starting point, they take two central quantities as their object of interest: seller profit and bidder surplus.

This is an example where defining one's objective from the outset and seeking to learn about those

quantities directly seems a particularly good approach. With correlated private values, the underlying valuation distribution can no longer be conveyed by a univariate CDF, but rather by a multivariate CDF of dimension equal to the number of bidders. Seller profit and bidder surplus on the other hand are one dimensional objects, and focusing on them directly obviates any potential complications of estimating bounds on multivariate functions. Moreover, in their model AGQ show that their parameters of interest depend only on the marginal distributions of each of the two highest order statistics. Thus, bounds on each of these two marginal distributions translate directly into bounds on the parameters of interest.

The theory that AGQ use regarding bidder behavior lies somewhere between that of HT and the standard button auction paradigm. The key implication that AGQ assume follows from bidder behavior is that the auction transaction price is determined by the maximum of the reserve price and the second highest bidder's valuation. This holds true in the standard button auction model equilibrium, but does not require equilibrium behavior.⁹ For example, AGQ point out that in a modified version of the HT model where there is no minimum bid increment and there are no jump bids, this assumption is also satisfied. AGQ begin without restricting the dependence between the valuation distribution and the number of bidders, and subsequently consider the impact of such restrictions at a later point.

The main problem these authors then face is that observed bids do not reveal the valuation of the highest bidder. The bids reveal that the winning bidder's valuation exceeds the second highest bid, but not by how much. Thus, the distribution of the highest valuation, $F_{n:n}(\cdot)$, will in general not be point identified. The authors show however that the correlated private values setup allows them to bound $F_{n:n}$ in an elegant way. The joint dependence of valuations is not known, but in the polar cases of independent valuations or perfectly correlated valuations, $F_{n:n}$ can be backed out from knowledge of $F_{n-1:n}$, the distribution of the second highest valuation. The authors show that these two cases in fact provide lower and upper bounds on $F_{n:n}$ in their correlated private values model. Since seller profit and bidder surplus are known functions of $F_{n-1:n}$ and $F_{n:n}$, this maps directly to bounds on each of these quantities.

However, the authors state that the implied bounds on seller profit and bidder surplus may be wide. They thus consider placing restrictions on the joint distribution of valuations and the number of bidders, for example by assuming that participation is exogenous, or that the valuation distribution increases in a stochastic dominance sense with n . These restrictions relate directly to and in some cases are implied or imposed by previously used models in the auction literature. Using such restrictions on how the distribution of valuations can change across auctions with different numbers of bidders, the bounds can then be tightened

⁹In the Appendix, AGQ show they can still attain bounds when they instead adopt the weaker HT assumptions.

by exploiting variation in the number of bidders across auctions.

AGQ consider again the timber auction data used by HT. They nonparametrically estimate $F_{n-1:n}$ conditional on auction characteristics X , which can then be mapped into bounds on profit and bidder surplus. They focus on auctions where each component of X is set to its median value, but they find similar results across other values of the conditioning variables. Their bounds on expected profit as a function of reserve price allowing for correlation across valuations indicate that the optimal reserve price is generally much lower than point estimates delivered by the IPV model. Moreover, their bounds indicate that the IPV model significantly overstates expected bidder surplus.

Finally, AGQ consider the reserve price policy used by the United States Forest Service. In particular they study the adoption of a “rollback factor” that was applied to a predicted sale price for each tract used to determine the reserve price, while taking account of a government mandated no-sale frequency of no more than 15%. Estimates using IPV lead one to conclude that a rollback factor of 10% would have yielded higher profit than larger rollback factors, while abiding by the no-sale percentage mandate. The bound estimates provided by AGQ contrast this finding, in that the lower bound on the no-sale frequency would exceed 15% at a rollback factor of 10%, and in fact for rollback factors as high as 30%. Their lower and upper bound estimates are more in-line with the actual average rollback factor of 31.1% in the region they study. Their bound estimates thus support the possibility that the seller’s policy was in accord with profit maximization subject to the no-sale mandate, while the IPV point estimates do not.

Given the wide range of results from the theory literature on auction models and their use in practice, there seems a good deal of scope to extend these applications to other, possibly partially-identifying contexts. Indeed other papers that use partial identification to deal with important issues in auction models other than those studied by HT and AGQ include Tang (2011), Armstrong (2013), and Gentry and Li (2014). Tang (2011) studies first-price auctions, where equilibrium behavior is assumed in a model of affiliated values that nests both common and private value auctions as special cases. He shows that nonetheless seller revenue under counterfactual reserve prices and auction rules can be usefully bounded, applying his approach to U.S. municipal bond auctions. Armstrong (2013) also studies first-price sealed bid auctions, where the key complication is heterogeneity in valuations that is unobservable to the econometrician. He shows that observing bids of just one bidder in each auction, or one order statistic across auctions, can be used to bound the mean of the bid or valuation distribution, and other functionals of the valuation distribution, and applies his results to highway procurement auctions. Gentry and Li (2014) study a two-stage auction model to account for selection, where in the first stage potential bidders choose whether or not to bid in the

auction. Even with some excluded instruments that affect selection into the auction, the model primitives are in general only partially identified.

5 Selection, Missing Data, and Treatment Effects

In this section we consider the use of bounds to address issues of selection, for instance in the study of wage distributions or treatment effects. Dealing with sample selection has been recognized as an important consideration in empirical work since at least the 1970s, and many models have been developed to account for it. In some contexts, the assumptions required to guarantee point identification may be called into question. Allowing for partial identification enables us to relax such assumptions and investigate what can be learned in their absence.

In the terminology of classical selection models, point identification is well-understood to be sensitive to the presence of an *instrumental variable* (IV) that affects participation but is entirely excluded from having a direct effect on the outcome of interest. In partially identifying models, identifying assumptions may still include restrictions on the selection process. For instance, a common source of identifying information is variation in the participation decision with respect to variables whose effect on the outcome of interest is somehow limited, for example through a monotone instrumental variable restriction (see (5.5) and the bounds (5.6) below) rather than an IV restriction. In other cases these restrictions take the same form as the instrumental variable and exclusion restrictions used in point-identifying models, but the partially identifying model relaxes other restrictions required for point identification.

We begin by considering Blundell, Gosling, Ichimura, and Meghir (2007), henceforth BGIM, a paper that uses partial identification to address selection into employment. The numerical examples in Section 6 that follows illustrate application of some of the bounds from that paper.

BGIM study changes in the distribution of both male and female wages, and in wage inequality, over time in the U.K. The problem is complicated by the fact that only the conditional wage distribution—that is, the distribution for those choosing to enter the workforce—is observed in the data. As the composition of the workforce changes over time, the observed distribution of wages will in general also change, obscuring any changes in the actual (uncensored) wage distribution, and hindering inference on the underlying changes in inequality over time. Censoring of the population wage distribution due to selection into the workforce has been a known problem for some time, having been pointed out in papers such as Gronau (1974) and Heckman (1974), which addressed the problem with parametric models of individual selection into the labor

force. A subsequent literature considers semiparametric and nonparametric selection models, see for example Powell (1987), Newey (1988), Heckman (1990), Heckman and Honore (1990), Ahn and Powell (1993), and Das, Newey, and Vella (2003), among others.

In much of the prior literature on selection models, sufficient conditions have been established for point identification of various parameters. Why then is there partial identification in BGIM? There are two reasons, both of which are important, and both of which are motivated by careful consideration of the application. First, rather than impose parametric restrictions on the determination of wages and labor force participation that may be hard to justify, BGIM instead sought to impose restrictions based on economic theory. Second, their goal was to learn about the shape of various conditional wage distributions, in order to learn about changes in inequality and gender wage differences over time. Semiparametric and nonparametric selection models previously developed in the literature achieve point identification of some parameters of interest, such as certain conditional mean functions or average treatment effects, but they do not generally point identify conditional distribution functions.

Thus, BGIM develop bounds on conditional wage distribution and quantile functions using restrictions motivated by economic theory. Indeed, in their context, there is partial - rather than point - identification even when they use an IV. They use out-of-work income as an instrument, but they have concerns that it might not be fully independent of unobservables that play a role in determining wages. They use insights from the literature on nonparametric bounds on treatment effects, e.g. Manski (1989, 1990, 1994, 1997) and Manski and Pepper (2000), to nonetheless exploit variation in observable variables, and household asset data in particular, in order to deliver nonparametric bounds on wage distributions while allowing for selection into the labor market. For example, they consider treating out-of-work income as a monotone instrumental variable. They begin by identifying the bounds implied by clear theoretical restrictions. They then consider the use of exclusion restrictions, and show how progressively stronger assumptions can generate increasingly informative bounds.

To illustrate, let W and X denote the log wage and a vector of observed exogenous variables that includes gender, age, education and year, respectively. The binary variable D indicates whether the individual is employed: $D = 1$ when W is observed and 0 otherwise. The probability that $D = 1$ given that $X = x$ is written as $P(x)$. Define $F(w|x)$ to be the cumulative distribution function (CDF) of W given $X = x$, $F(w|x, D = d)$ to be the CDF of W given $X = x$ and $D = d$. $F(w|x)$ is the object of interest. It is not

identified due to nonrandom selection into employment. However it can be written as:

$$F(w|x) = F(w|x, D = 1)P(x) + F(w|x, D = 0)[1 - P(x)]. \quad (5.1)$$

The problem is illuminated by observing that although $F(w|x, D = 1)$ and $P(x)$ are identified, $F(w|x, D = 0)$ is not, since wages are not observed for individuals who do not work.

The implied *worst-case bounds* come directly from Manski (1994). Since the conditional CDF $F(w|x, D = 0)$ is bounded between 0 and 1, it follows that:

$$F(w|x, D = 1)P(x) \leq F(w|x) \leq F(w|x, D = 1)P(x) + 1 - P(x). \quad (5.2)$$

BGIM then derive tighter bounds under a *first order stochastic dominance assumption*, which strengthens a monotone treatment selection (MTS) assumption of Manski and Pepper (2000) for the purpose of studying distribution functions. This is essentially an assumption that people with higher wages are more likely to work, i.e. that there is positive selection into the labor market. Formally, when $0 < P(x) < 1$, the assumption is that $F(w|x, D = 1) \leq F(w|x, D = 0) \forall w, x$. Substituting into 5.1 generates:

$$F(w|x, D = 1) \leq F(w|x) \leq F(w|x, D = 1)P(x) + 1 - P(x). \quad (5.3)$$

BGIM note that if selection into employment were determined by a Roy Model, the first order stochastic dominance assumption could fail if the wage and reservation wage were sufficiently positively correlated. This motivates them to consider a weaker assumption that the median wage offer for those not working is weakly lower than the median of observed wages. This tightens the bounds in (5.2) only for quantiles at the median and above.

They also consider the impact of *exclusion restrictions*. In the absence of the aforementioned MTS assumption, Manski (1994) showed that if W is independent of Z conditional on X then the bounds on the conditional distribution $F(w|x)$ become:

$$\max_z \{F(w|x, z, D = 1)P(x, z)\} \leq F(w|x) \leq \min_z \{F(w|x, z, D = 1)P(x, z) + 1 - P(x, z)\}. \quad (5.4)$$

Tighter bounds can also be obtained by combining this exclusion restriction with one of the above MTS assumptions. Alternatively the exclusion restriction can be weakened to a monotone instrumental variable

assumption, similar to that considered by Manski and Pepper (2000). This assumption imposes that the conditional distribution of wages is weakly decreasing in the value of the instrument:

$$\forall w, x, z < z' : F(w|x, z') \leq F(w|x, z). \quad (5.5)$$

This restriction offers an alternative way to tighten the worst-case bounds in (5.2). Under this restriction, $F(w|x, z')$ is bounded from below by the largest worst-case lower bound on $F(w|x, z)$ over the range $z \geq z'$, and $F(w|x, z')$ is bounded from above by the lowest worst-case upper bound on $F(w|x, z)$ over the range $z \leq z'$. Then z' may be integrated out to obtain MIV bound $F(w|x)$ as follows:

$$E_Z[\max_{z \geq Z}\{F(w|x, z, D = 1)P(x, z)\}] \leq F(w|x) \leq E_Z[\min_{z \leq Z}\{F(w|x, z, D = 1)P(x, z) + 1 - P(x, z)\}], \quad (5.6)$$

see Manski and Pepper (2000) for further details.

For estimation, BGIM proceed by constructing nonparametric estimates of the objects $P(x)$, $F(w|x, D = 1)$, and their values at percentile ranks of the instrument Z at particular values $X = x$. These are then substituted into the relations above to obtain estimates of the bounds under different sets of assumptions. Confidence intervals are constructed by using a bootstrap procedure in combination with methods developed by Imbens and Manski (2004). The authors' main focus is on the bounds on the quantiles, which are obtained in a straightforward way from the bounds on the wage distribution.

This paper provides a useful illustration of the increasingly tight bounds generated by increasingly strong assumptions. The worst case bounds alone indicate that male wage inequality rose from 1980-1998 and that inequality as measured by the interquartile range must have risen by at least 0.089 log points. Adding the median restriction increases that estimate to 0.127 log points; adding both median and monotonicity restrictions generates an estimate of 0.252 log points. The worst case bounds are uninformative regarding changes in gender wage differentials over time because of the lower employment rates for women. The combination of the monotonicity, median, and an additional additivity restriction indicate that the male/female wage differential declined by at least 0.23 log points between 1978 and 1998.

Restrictions similar to those used by BGIM have also been used in the study of treatment effects or program evaluation. For example, Kreider, Pepper, Gundersen, and Jolliffe (2012), henceforth KPGJ, study the effects of the Supplemental Nutrition Assistance Program (SNAP, formerly known as the Food Stamp Program) on child health. The previous empirical literature generated little evidence that the program

promoted food security or reduced health problems. However, as KPGJ point out, these effects are difficult to identify for two reasons. First there is a selection problem because the decision to participate is unlikely to be exogenous: families may choose to participate precisely because they expect or are already experiencing poor health. Second there is an issue of non-random measurement error because many families do not report SNAP participation in household surveys. Both issues can be addressed using partial identification methods, with weaker and potentially more credible assumptions than would be needed under standard parametric approaches.

KPGJ use three sets of restrictions to generate inequalities. They begin by focusing on the selection problem, abstracting away from measurement error until later in the paper. They consider the *Monotone Treatment Selection* (MTS) restriction (Manski and Pepper (2000)) that the decision to enter SNAP is monotonically related to poor latent health outcomes. They add the *Monotone Instrumental Variable* (MIV) assumption that the latent probability of a poor health outcome is non-increasing in household income (adjusted for family composition). Finally they consider a *Monotone Treatment Response* (MTR) assumption that participation in SNAP does not worsen health status. This last restriction assumes an answer to part of the question being considered but allows them to tighten the bounds on the magnitude of the effect. Finally they introduce measurement error to the model and develop a method to address it with additional data (administrative data on the size of the caseload) and an assumption of no false-positive reports of participation.

As in the previous paper, the inputs to the bounds are estimated nonparametrically. Inference follows Kreider and Pepper (2007). The estimates from the MTS restriction alone do not allow the authors to sign the impact of SNAP on health outcomes. When MIV is added the bounds become tighter and the confidence intervals almost always exclude zero, generating new evidence that SNAP participation reduces negative health outcomes. Adding MTR makes the bounds tighter still. When the authors allow for measurement error, they can identify strictly negative effects on poor health outcomes under the MTS and MIV assumptions for sufficiently small degrees of food stamp reporting error. Under joint MTS-MIV-MTR, SNAP is found to lead to a decline in food insecurity rates and in poor health outcomes even when allowing for high rates of measurement error.

Our goal in reviewing BGIM and KPGJ is to give the flavor of the sets of assumptions that have proven useful in applications. Yet there have been several other applications employing bounds on treatment effects or counterfactual outcome distributions in the presence of selection. Some further examples include Heckman, Smith, and Clements (1997), Heckman and Vytlacil (1999), Manski and Nagin (1998), Ginther

(2000), Gonzalez (2005), and Bhattacharya, Shaikh, Vytlacil (2008, 2012), Manski and Pepper (2013), and Siddique (2013). In related work Honore and Lleras-Muney (2006) estimate bounds in competing risks models, focusing attention on changes in cancer cardiovascular disease mortality since the 1970s, where selection is due to the structure of the competing risks setup rather than treatment assignment.

6 An Illustration: Modeling Wages and Labor Supply

We now revisit the important problem of selection into the labor market in the analysis of wage distributions, using numerical illustrations to demonstrate the interplay between assumptions and data in partial identification analysis. Our starting point is a parametric probit selection model studied in the pioneering work of Heckman (1976) and related to models in Heckman (1974) and Gronau (1973, 1974). Such models have been used by many authors since to study the wage distribution—and in particular issues like the determinants of female labor force participation and wages—allowing for non-trivial selection into the labor market.

We first estimated the probit selection model using data from the 1967 National Longitudinal Survey of Women aged 30-44 (NLSW67), the same data set used by Heckman (1974, 1976). For the purpose of these illustrations we use the conditional distributions of employment and observable wages corresponding to the NLSW67 parameter estimates taken as population values. We then examine the identifying power of different models for the *unconditional* distribution of female wages, working or not working. We focus on identification, that is the extent to which the bounds implied by our assumptions provide information on the objects of interest given the variation in the data, rather than estimation or inference.

6.1 The Probit Selection Model

The probit selection model comprises two equations, one for the determination of the individual’s log wage W and the other for employment, indicated by the binary variable D as follows:

$$W = \beta_0 + X_1\beta_1 + X_2\beta_2 + U_1, \tag{6.1}$$

$$D = 1[\gamma_0 + Z\gamma_1 + X_2\gamma_2 + U_2 > 0], \tag{6.2}$$

where $U = (U_1, U_2)$ is a bivariate normal unobservable representing individual specific heterogeneity. The variance of U_2 is normalized to one. The variance of U_1 and the correlation of U_1 and U_2 are denoted by the parameters σ^2 and ρ , respectively. Log wage is only observable when $D = 1$, for which we define the

random variable

$$Y \equiv D \cdot W.$$

The vector $X \equiv (1, X_1, X_2)$ comprises covariates that enter into the determination of wages, while $(1, Z, X_2)$ affect employment. The vector X_2 consists of covariates common to both equations, while Z comprises instruments excluded from the wage equation, and X_1 contains exogenous variables excluded from the selection equation. X_1 may be empty, but Z should not be. Variables in Z are the instrumental variables that effect selection into employment, but do not otherwise effect wages. Unobserved heterogeneity U is restricted to be independent of the exogenous variables (X_1, X_2, Z) . The researcher is presumed to have a random sample of observations of (Y, D, X, Z) denoted $\{(y_i, d_i, x_i, z_i) : i = 1, \dots, n\}$.

The implied conditional distribution of log wage given covariate values (x, z) is $\mathcal{N}(x\beta, \sigma^2)$. Following Heckman (1976) the model can be estimated either *via* a two-stage procedure or by maximum likelihood. We use the NLSW67 to construct a sample of 2,263 white, married women with spouse present from the original sample of 5,083 women; further information on the dataset is provided in Shea et al (1970) and in Heckman (1976). We set

$$\begin{aligned} X_1 &= \text{YearsWorked}, & X_2 &= \text{YearsEducation}, \\ Z &= (\text{HusbandAnWage}, \text{HHAssets}, \text{KidsUnder6}). \end{aligned}$$

We use the Heckman command in Stata to estimate the model via maximum likelihood, with log hourly wage taken as the outcome variable. Details of the sample, variable definitions and estimates are provided in the Appendix.

6.2 Numerical Illustration of BGIM Bounds

We begin by computing the BGIM *worst-case bounds* (Manski (1994)) implied by setting the “true” population parameter values to equal our estimated parameters. The bounds are given in BGIM and derived in Section 5. Conditioning on realizations of (X, Z) in place of X in (5.2):

$$F(w|x, z, D = 1) P(x, z) \leq F(w|x, z) \leq F(w|x, z, D = 1) P(x, z) + 1 - P(x, z). \quad (6.3)$$

When the population data generation process follows the probit selection model,

$$P(x, z) \equiv \Pr(D = 1 | X = x, Z = z) = \Phi(\gamma_0 + Z\gamma_1 + X_2\gamma_2).$$

Solving for $F(w|x, z, D = 1)$ we obtain

$$F(w|x, z, D = 1) = \frac{1}{P(x, z)} \int_{-\gamma_0 - z\gamma_1 - x_2\gamma_2}^{\infty} \Phi\left(\frac{w - x\beta - \sigma_1\rho t}{\sigma_1\sqrt{(1 - \rho^2)}}\right) \phi(t) dt.$$

These quantities can be computed using standard software (we used matlab) and numerical integration, which can in turn be used to compute the worst case bounds (6.3) for all values of w , for any given (x, z) and parameter vector θ .

We then compute IV bounds. Under the assumption that $F(w|x, z)$ does not vary with z (i.e., W is independent of Z conditional on X), then for any x , and each w ,

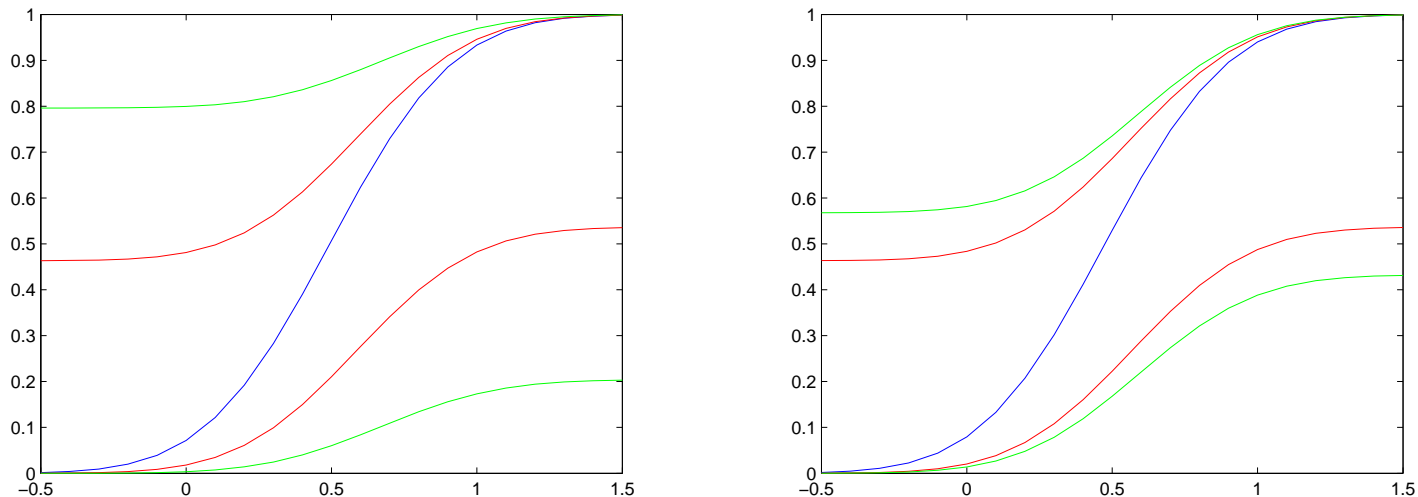
$$\max_z \{F(w|x, z, D = 1)P(x, z)\} \leq F(w|x) \leq \min_z \{F(w|x, z, D = 1)P(x, z) + 1 - P(x, z)\}, \quad (6.4)$$

With functions that compute the worst case bounds already in hand, we can maximize and minimize those bounds over a range of value for z numerically to compute the resulting bounds on $F(w|x)$.

Figure 1 plots the worst-case and IV bounds (in green and red respectively), and the CDF of log hourly wage implied by the estimated probit selection model (in blue), for women with 12 years of education and 3 years' experience. We hold the support of z fixed across the two panels: we take the first 400 z combinations in the data where the woman has 12 years of education and assume they define the conditional support of Z for every experience level. Because no exclusion restriction is made for the worst-case bounds, they are illustrated conditional on particular values of z (husband's annual wage, household assets, and kids under 6). In the left-hand panel we impose $z = (8000, 12600, 1)$. This value is observed in the first 400 z combinations we consider and is fairly close to the empirical mean of $(7444, 17503, 1)$ and the empirical median of $(7500, 10250, 1)$. In the right-hand panel we set $z = (5000, 1675, 0)$.

Of the three sets of restrictions, the worst-case bounds are the widest because they impose the weakest assumptions: random sampling of observables, but no further restrictions on the data generating process. The derivation of these bounds does not make use of linearity of the conditional log wage function in the covariates, nor does it impose a selection equation, Gaussian distribution, or exclusion restrictions. The

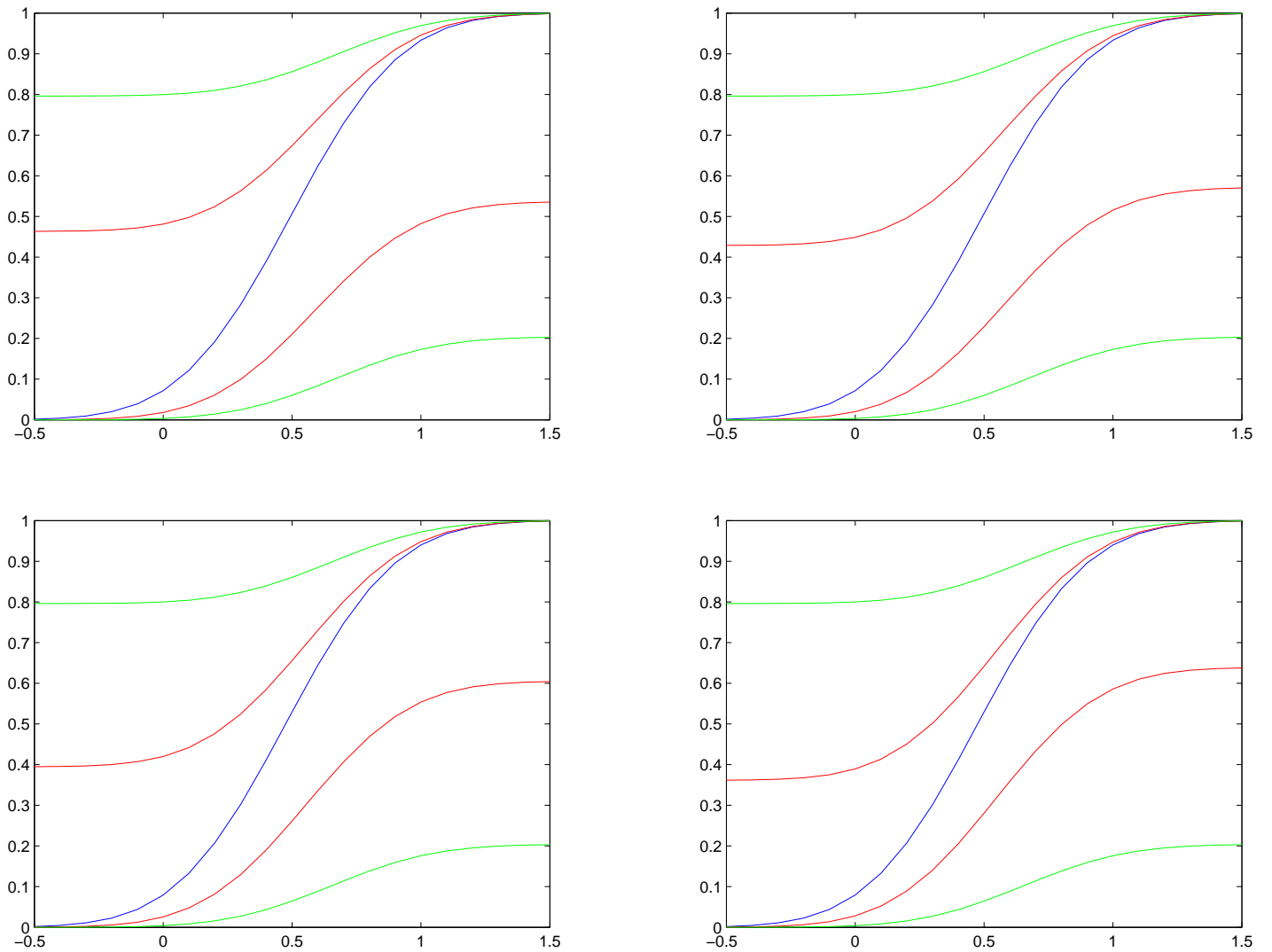
Figure 1: Bounds on log hourly wage distributions conditional on 12 years of education and 3 years' work experience. The left-hand panel defines worst-case bounds at $z = (8000, 12600, 1)$; the right-hand panel uses $z = (5000, 1675, 0)$. Log hourly wages are displayed on the x-axis. In each figure the CDF implied by the estimated probit selection model is drawn in blue. Worst-case bounds described in (6.3) are shown in green, and IV bounds described in (6.4) are shown in red.



IV bounds are tighter because they add the restriction that the conditional distribution of log wages given (x, z) does not vary with the excluded instruments z . The probit selection model additionally imposes the parametric structure of (6.1) and (6.2), as well as joint normality of unobservables. Hence it imposes the strongest assumptions and point identifies the conditional wage distribution.

The bounds in Figure 1 illustrate the role of the instrumental variable restriction relative to the restrictions imposed for the worst-case bounds. The IV bounds always improve upon the worst case bounds, but the size of the improvement varies with (x, z) . For instance, in the left panel corresponding to $z = (8000, 12600, 1)$, the worst case bounds are extremely wide. This reflects the fact that in this population only a small fraction of women with the given covariate values work. The worst-case bounds assume nothing about the distribution of wages of those who do not work, and this makes the bounds wide. The IV bounds are much tighter than the worst case bounds for these values of z . The right panel, corresponding to $z = (5000, 1675, 0)$, is different. At these lower values of the husband's annual wage and household assets, a larger proportion of women with the given education and experience levels work, and this makes the worst-case bounds more informative. For this z the IV restriction also clearly helps to tighten the bounds, but the IV bounds are only slightly tighter than the worst-case bounds at this z .

Figure 2: Bounds on log hourly wage distributions. The four panels illustrate CDFs conditional on 12 years of education and 3 years' work experience. The top left column uses the support of Z in the first 400 observations in the data. The top-right panel concatenates the original support with $2 * z$ for each z on the original support; the bottom left also adds $3 * z$ for each such initial value, and the bottom right additionally includes every such z multiplied by 4. Each successive panel therefore both introduces an additional 400 possible combinations of Z and also increases their support. In each figure the CDF implied by the estimated probit selection model is drawn in blue. Worst-case bounds described in (6.3) are shown in green, and IV bounds described in (6.4) are shown in red.



These two panels illustrate the variation in worst-case bounds across realizations of Z that is exploited to generate the IV bounds. They support the intuition that, the larger the support of Z , the greater the possible variation in worst-case bounds across particular values of z , and the more informative we might expect the IV bounds to be. That is, additional values of Z that can be conditioned upon serve the purpose of increasing the potential variation in the data useful for identification. Figure 2 illustrates this dynamic by plotting the IV bounds as we increase the support of Z beyond the values observed in the data. Again we consider women with 12 years of education and 3 years of work experience, and plot the CDF of log hourly wage implied by the estimated probit selection model, the worst-case bounds (at the z values in the left-hand panel of Figure 1), and the IV bounds. The top left panel defines the support of Z as in Figure 1, based on the first 400 observations in the data. The top right panel appends to this an additional set of possible z combinations where every observed value is multiplied by 2. The bottom left panel also appends the triple of every observed z combination and the bottom right also multiplies every combination by 4. Each successive panel thus both introduces an additional 400 combinations of z and increases their range. Consistent with the intuition suggested by Figure 1, every successive increase in the support of Z generates more informative IV bounds.

However there are caveats to these findings. First, the intuition that increasing the support of Z tightens the IV bounds requires the worst-case bounds to be more or less restrictive for different values of z , which may not always be the case. For example, in our setting, if the entire support of household assets was high, e.g. every value was above \$100,000, then variation across that support might have little effect on the selection probability, since only a relatively small fraction women with these asset levels choose to work. This would imply little change in the worst-case bounds across z and little difference between worst-case and IV bounds, even if the support was large. A second, related caveat is that a broader support of z , particularly across values that affect the selection probability, may make the IV exclusion restriction less plausible. Concerns about this issue motivated BGIM to move to the weaker monotone IV assumptions. These issues carry over to other settings; they are reasons why applied researchers may generally benefit from assessing the power of different identifying assumptions in their particular applications.

Finally note that nothing in this section is estimated. We simply characterize the bounds implied by the different theoretical restrictions when, unknown to the researcher, the data generation process (DGP) follows the probit selection model of Heckman (1976). We set the population parameter values for the underlying DGP to be parameter estimates obtained using the NLSW67 data. Our objective is to illustrate the different amounts of information provided by the different bounds as we change the underlying DGP,

namely the values and support of the conditioning variables.

7 Discussion

The examples discussed in this paper illustrate a variety of ways in which partially identifying models have been used in applied work. The methods have been used to address problems posed by multiple equilibria; to infer demand using restrictions that derive from revealed preference arguments; to bound bidder valuations, optimal reserve prices, and surplus in auctions; and to deal with sample selection and missing data. Our list is by no means exhaustive. Our objective is simply to provide guidance to applied researchers with an interest in these methods by pointing out some of the areas where they have been implemented successfully. Although there is a wide diversity of applications to which the methods have been applied, there are some notable common themes. In this section we discuss these common themes, and then briefly consider some further important aspects of partial identification analysis in applied work.

7.1 Common Themes in Applications

The first theme we wish to emphasize is the use of restrictions motivated from economic theory. In each area of application, there is consideration of an underlying model of economic agents' behavior that produces meaningful restrictions. The theory of revealed preference used to produce bounds in the papers cited in Section 3 has strong theoretical foundations dating back to Samuelson (1938) and a rich history in economics, see e.g. McFadden (2005). The entry models considered in Section 2 also impose optimizing behavior, but feature economic agents taking decisions that affect not only themselves, but also others. That is, the agents are players in a game. Authors typically use an equilibrium solution concept, again with established theoretical foundations, e.g. Nash Equilibrium in complete information games (Nash (1950)) and Bayesian Nash Equilibrium in incomplete information games (Harsanyi (1967)). In the HT analysis discussed in section 4 the restrictions on bidder behavior are essentially revealed preference conditions that relax the typical equilibrium assumptions. More broadly, the literature on auction theory also offers a rich history of analyses of another class of game theoretic setting. The typical solution concept is that of Bayesian Nash Equilibrium, although others are possible. Selection models of the type discussed in Section 5 also have their foundations in models of agent's optimizing behavior, see e.g. Heckman (1974) where women optimally choose whether to work and, if so, how many hours to work to maximize their utility. The shape restrictions and monotone IV restrictions employed by BGIM and KGPJ have clear interpretations that are consistent

with certain models of individual behavior.

Second, the papers considered also make use of assumptions regarding the properties of unobserved heterogeneity in their models. These assumptions are often motivated by thinking carefully about the potential sources of unobserved heterogeneity in the application at hand, and the implications of whether or not economic agents observe econometric unobservables at the time their decisions are made. The timing of the assumed underlying model can be an important factor that determines the relationship between unobserved and observed exogenous variables, as in for instance Eizenberg (2014), Nosko (2014) and Wollman (2014). This is reminiscent of the careful discussion regarding timing, and the resulting properties of unobservables, in Olley and Pakes' (1996) study of production functions. Furthermore, the assumptions made on the properties of unobservables have a direct impact on the form of the inequalities used for estimation.

Third, as in point-identifying models, the restrictions imposed on functional forms or the distribution of unobserved variables in partially identifying models do not always have a clear tie to economic theory or lend themselves to a clean interpretation. Nonetheless, depending on the problem at hand, the strength of the theoretical restrictions imposed, and the data available, these restrictions may still have a role in making the model more tractable. The key point here is that partial identification is not a panacea for using assumptions. By allowing for partial identification, we give ourselves the freedom to compare the implications of different models – point identifying or not – and to more richly assess what conclusions may be drawn from different assumptions. This allows researchers who disagree about which assumptions are more or less palatable to understand which assumptions may lead them to more or less clear conclusions.

Fourth, the applications on which we've focused combine theoretical modelling restrictions with data. The data used in each application allow the researcher to identify a particular distribution of observed variables. In most of our examples the model's implications for this distribution, which are used for estimation, can be written as a collection of conditional moment inequalities of the form:

$$E[m(Y, X, \theta)|Z = z] \geq 0, \tag{7.1}$$

which must hold for almost every value of the conditioning variable Z . Here Y denotes a vector of outcome variables, and X a vector of variables that have a role in the determination of Y . Z are exogenous variables, which may contain elements of X as well as additional variables excluded from having a role in determining

Y , that is instrumental variables. The identified set comprises the set of parameter values θ such that

$$P_Z [\{z \in \text{Supp}(Z) : E[m(Y, X, \theta)|Z = z] < 0\}] = 0. \quad (7.2)$$

In other words, the bounds comprise those values of θ for which there is no positive measure set of z -values with respect to the distribution of Z that violate (7.1).

The inequality (7.1) and the corresponding bounds characterization (7.2) point the way towards addressing the question of what kind of variation in the observed data can help provide restrictive parameter bounds. The bounds are tighter when more values of θ can be excluded from the identified set. Inequality (7.1) shows that observing realizations $Z = z$ that induce higher values of the conditional moment function $E[m(Y, X, \theta)|Z = z]$, and in particular that make this expression positive, generates inequalities that can exclude θ from the identified set. Mechanically, we see that all else equal, observing a wider range of values of Z helps in obtaining a tighter identified set.

In practice, with actual data, one can contemplate exactly which variables Z it is that can provide useful exogenous variation. These are typically different kinds of instrumental variables, variation in which changes the corresponding conditional distribution of endogenous variables Y . Likewise, variation in Z could also induce changes in the conditional distribution of X given Z . Both effects of variation in Z can induce a change in the value of the conditional moment $E[m(Y, X, \theta)|Z = z]$ at a given value of θ . We saw instances of such variables in each of the applications discussed above. Consideration of the underlying economic processes and the mechanisms that generate one's data can be used to reason which variables play such a role. The use of exogenous variation and instrumental variable restrictions in applied work has been commonplace for some time; it is not unique to the partial identification literature. Conceptually, partially identifying models exploit exogenous variation in observed data in the same way as point-identifying models, but this variation is not required to pin down the parameters of interest uniquely.

7.2 Sharp Bounds and Tight Bounds

A concern in applied work using partial identification is that bound estimates – either set estimates or confidence regions – could be too wide to answer a question of practical interest. We refer to such situations as those in which bounds are not “tight”. What is meant by this is both application-dependent and somewhat subjective. While we believe the concern of obtaining bounds that are not tight is valid, in our view it should not stop researchers from using these methods, but rather should help to guide them in the research process

and in the interpretation of results.

How can such concerns help to guide empirical research? First, partial identification analysis broadens our ability to consider alternative menus of assumptions. Given a particular data set, some sets of assumptions may produce bound estimates that are tight and others may not. If certain assumptions are simply not strong enough to generate tight bounds given the data available, this is useful to know! It tells us that in order to get informative bounds we either need better data (in the sense of observing more variables, or greater variation in exogenous variables) or stronger assumptions. Further assumptions can of course be added. Researchers can then examine the trade-off between adding assumptions and having less informative bounds, and can debate the validity of the assumptions used. Alternatively, the researcher may consider how to obtain better data with the further exogenous variation that would be helpful to learn about the questions of interest.

A related but different concept is the notion of “sharp” bounds. Bounds for a parameter vector θ are said to be sharp if, given the assumptions made, they comprise *only* those values of θ that arise in conjunction with a data generation process that could have produced the distribution of observed variables, and no others. Bounds that are not sharp are valid in the sense that they include these values of θ , but may additionally include values of θ for which there is no data generation process satisfying the modelling restrictions that is capable of producing the distribution of observable variables. Sharp bounds comprise *the* “identified set” for θ , using all of the restrictions of the model to obtain the smallest possible bounds. The question of whether a given collection of a model’s observable implications – typically moment equalities and inequalities – characterize sharp bounds is thus a property of the those implications and the model itself, addressable without reference to data.

Whether sharp bounds are sufficiently tight to address a question in a given application is an empirical matter. It can be easier to characterize non-sharp bounds, or more convenient to base estimation and inference on non-sharp bounds for computational reasons. Indeed, some of the bounds on which estimation is based in the applications already discussed are known not to be sharp. These non-sharp bounds produced sufficiently tight estimates in these applications to deliver interesting and useful empirical results. But suppose that in another application, non-sharp bounds are used, and the bound estimates for parameters of interest are found to not be sufficiently tight to address questions of empirical interest. Why are the bounds not tight? There are two possible reasons, which the researcher cannot distinguish. One reason could be that there is simply not enough variation in the data to answer the empirical question(s) of interest. The other possibility is that there is enough variation in the data, but the researcher has not used all of the implications

of the model to the fullest extent. That is, it could be that if the researcher had based estimation on sharp bounds, the resulting bound estimates would have been sufficiently tight to address questions of empirical interest.

Characterization of sharp bounds rather than merely valid or “outer” bounds is therefore important for informing applied work. Manski and co-authors clearly distinguish sharp and non-sharp bounds. As discussed above, non-sharp bounds can be useful for applied work, but it is still important to recognize when they are not sharp in order to fully understand the mapping from the distribution of observable data to information about parameters of interest.

Establishing that bounds are sharp can be difficult. Until very recently, this question had been addressed on a case-by-case basis, often with model-specific constructive arguments. Fortunately, recent advances have been made in developing general tools to address whether bounds are sharp, for example by Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011), and Chesher and Rosen (2014). These papers use alternative approaches to characterize sharp bounds in a variety of different models. Additionally, in those situations in which more than one apply, they can offer different representations of the identified set that can be used to motivate estimation and inference.

Beresteanu, Molchanov, and Molinari (2011) use properties of the set of possible outcomes Y produced by incomplete structural econometric models in which the identified set can be characterized by a finite number of conditional moment equalities involving an unknown and possibly infinite-dimensional function. This function is referred to as a selection mechanism, since it plays the role of selecting from among the set of possible outcomes. Beresteanu, Molchanov, and Molinari (2011) develop a tractable characterization of sharp bounds in such models by establishing an alternative representation of the identified set based on conditional moment inequalities, from which the unknown selection mechanism is absent. They illustrate the applicability of their methods to models with multiple equilibria such as that of CT, allowing for either pure strategy or mixed strategy Nash Equilibrium, as well as correlated equilibrium or Bayesian Nash Equilibrium if there is incomplete information. They additionally show how their characterization can be applied to best linear prediction or multinomial choice models with interval data. See also Beresteanu, Molchanov, and Molinari (2012) for characterizations of sharp bounds for the distribution of response functions in various treatment effect models.

Galichon and Henry (2011) provide sharp bounds on the parameters of incomplete models that allow for multiple equilibria, when the distribution of unobserved variables is parametrically specified. They show how tools from optimal transportation theory may be usefully applied, and they introduce the important concept

of core-determining sets, which help to determine which of a collection of moment inequalities are necessary for obtaining sharp bounds, and which can be helpful for making such characterizations tractable.

Chesher and Rosen (2014) study partial identification in structural econometric models, and focus their analysis on obtaining sharp bounds on structural functions and distributions of unobserved heterogeneity by using properties of the set of possible values of unobservables that may occur given observed variables. Using the inverse mapping from outcomes Y to unobserved heterogeneity U in this way enables application of their analysis in models imposing a variety of different restrictions on unobservables, of the sort commonly used in structural econometrics. They demonstrate for example how sharp bounds can be obtained under conditional mean, conditional quantile, or independence restrictions. The developments deliver novel results in particular for models with nonparametrically specified distributions of unobservable heterogeneity and continuously distributed endogenous variables.

For instance, in Chesher, Rosen, and Smolinski (2013) the authors show how their approach can be used to obtain sharp bounds on model parameters and distributions of unobserved heterogeneity, as well as bounds on counterfactual choice probabilities, in unordered discrete choice models with endogenous covariates and instrumental variable restrictions. The analysis can be applied to the classical setup of McFadden (1974), with or without the Type I Extreme Value assumption on the distribution of unobserved heterogeneity, where some of the observed individual characteristics can be allowed to be correlated with unobserved heterogeneity in preferences. For example, in the classical example of choice of transportation to work, individuals could choose in part where to live based on their preferences for mode of transport. If so, then distance to work will be correlated with the unobservable components of utility from different transportation options. Chesher, Rosen, and Smolinski (2013) show how an instrument that is excluded from the utility functions and independent of unobserved components of utility can be used to address this issue. Point identification generally does not obtain, even if parametric functional forms are assumed for the utility functions. The resulting identified set is characterized by a collection of conditional moment inequalities.

Another setting of practical interest to which the Chesher and Rosen (2014) analysis applies is the auction model of HT, where the question of the sharpness of the bounds derived was left open. Chesher and Rosen (2015) resolve this question, applying their analysis to a slightly simplified version of the HT auction model. They show that, in addition to the inequalities used to bound valuation distributions in HT, there are additional inequalities that further refine the identified set, bounding not only the valuation distribution at each point on the distribution, but also the shape of the distribution function as it passes through multiple points. Chesher and Rosen (2015) show with numerical examples that the additional inequalities can be

binding, and can carry information about objects of economic interest, such as the optimal reserve price. The identified set obtained comprises a dense collection of a continuum of conditional inequalities involving the valuation distribution. In ongoing research the authors study the application of their identification analysis to auctions with unobserved heterogeneity and affiliation in bidder valuations.

7.3 Inference, Specification Testing, and Computation

So far we have focused solely on identification analysis for the purpose of illuminating the link between (1) the strength of maintained assumptions and (2) variability in observable data. We now briefly discuss further considerations that arise in the construction of set estimates in practice.

First, partially identifying models typically lead to bound characterizations by way of conditional or unconditional moment inequalities of the form (7.1) for some collection of moment functions m , for which there are currently a variety of different methods that can be used to construct confidence regions. Conceptually these confidence sets serve the same purpose as confidence sets for point-identified parameters. As with point-identified parameters, confidence sets may be constructed that are guaranteed to include the true population parameter of interest with high probability asymptotically (e.g. 0.95) in repeated samples.¹⁰

In cases where there is interval identification of a univariate parameter of interest, with asymptotically normal estimators for the interval endpoints, the methods developed by Imbens and Manski (2004) and Stoye (2009) are applicable and easy to compute. If this is not the case, for example if there is interval identification with interval endpoints defined by intersection bounds, or if the identified set does not take the form of an interval, then the situation is more complex. If the bounds are given by moment functions of the form (7.1), but with a discrete conditioning set of values z and a finite number of moment functions, then one can use inference methods that employ a finite number of moment inequalities, such as Chernozhukov, Hong, and Tamer (2007), Beresteanu and Molinari (2008), Romano and Shaikh (2008, 2010), Rosen (2008), Stoye (2009), Andrews and Soares (2010), Bugni (2010), Canay (2010), Romano, Shaikh, and Wolf (2014), and Pakes, Porter, Ho, and Ishii (2015), among others. If instead the conditioning variable is continuous with a continuum of values for its support then inference methods such as those developed by Andrews and Shi (2013), Chernozhukov, Lee, and Rosen (2013), Armstrong (2014), Chetverikov (2011), or Lee, Song, and Whang (2013a,b) can be used.

¹⁰There is also a conceptually different criterion that has been considered in the literature on partial identification, namely construction of confidence regions that asymptotically contain the entire identified set with prespecified probability under repeated samples. See e.g. Imbens and Manski (2004) or Chernozhukov, Hong, and Tamer (2007) for discussion of this type of confidence region.

Second, in practice it is possible that analog estimators of the identified set admit no values of θ . That is, it could very well be that there is no value of θ such that the sample analogs of the moments $E[m(Y, X, \theta)|Z = z]$ in (7.1) are all nonnegative. The analog set estimator is then empty. If the moment estimators were equal to the population moments, this would indicate that the model was misspecified. However, since the estimators suffer from sampling error, they are only approximations to the population moments. If the empirical moment inequalities are close to being satisfied for some values of θ , it is quite possible that the population moments do satisfy these inequalities at some θ , with the analog set estimator being empty only due to sampling error. Indeed, this was the case in the application previously discussed in Ho and Pakes (2014). They found however that their corresponding confidence sets, which allowed for the possibility of partial identification, were in fact not empty. The empty analog estimator was likely to be solely due to sampling variation. On the other hand, BGIM found that with their data some of the estimated bounds on wage distributions were empty. This persisted when they took account of sampling variation via a simulation procedure, and they reasoned that the IV assumption may not have been appropriate in their sample.¹¹ They argued that a weaker MIV restriction could be more reasonable, and found that the estimated MIV bounds did not cross. The logic behind this reasoning has since led to the development of specification tests in moment inequalities models, as considered by Bugni, Canay, and Shi (2015) and references therein.

Third, computing set estimators and confidence sets for partially identified parameters can be challenging. This is due in part to the relative novelty of inference methods, which have been primarily developed within the last ten years, and some much more recently. As has historically been the case for new econometric methods, there is much scope for computational advances to simplify their implementation. That said, some code has already been made publicly available, including Beresteanu and Manski (2000a,b), Beresteanu, Molinari, and Steeg Morris (2010), and Chernozhukov, Kim, Lee, and Rosen (2015). These implement some of the methods from papers such as Manski (1990, 1997), Manski and Pepper (2000), Beresteanu and Molinari (2008), and Chernozhukov, Lee, and Rosen (2013).

One difficulty with presenting set estimators arises when identified sets correspond to high dimensional parameter vectors. This can present computational as well as presentation issues. In terms of computation, it is costly to scan over even a moderate-dimensional parameter space to check whether each candidate parameter value passes a given criterion used for estimation or inference. This is one area where the potential for computational gains seems promising, particularly if one is willing to focus on models that exploit

¹¹The IV assumption here was that out of work income had no effect on the wage distribution, except through selection. This assumption may also be suspect based on economic reasoning, see BGIM for discussion.

some sort of common structure, such as index restrictions. The recent literature suggests the potential for computational gains using approaches pioneered in other literatures, such as the slice sampling method of Neal (2003) used by Kline and Tamer (2012), and methods from machine learning used by Bar and Molinari (2015).

As a practical matter, the presentation of estimators or confidence intervals for identified sets of more than three dimensions is not straightforward. A natural approach is to report projections of such sets along certain dimensions, for example by reporting confidence intervals for particular components of a partially identified parameter vector. Projecting confidence sets for high dimensional parameter vectors into lower dimensions generally results in conservative confidence sets for the lower dimensional objects. A current area of research in the literature seeks to address this by considering inference on projections of the identified set directly, rather than projection of confidence intervals for higher dimensional parameter sets. See for example Bugni, Canay, and Shi (2014) and Kaido, Molinari, and Stoye (2015) for approaches with sets defined by unconditional moment inequalities, and the Bayesian approach of Kline and Tamer (2012) for which inference on functionals of parameters is straightforward.

Fourth, an important point arises when considering the use of sharp versus non-sharp bounds in practice. As discussed in Section 7.2 and exemplified in some of the applications discussed, non-sharp bounds can sometimes produce informative results, depending on the distribution of observable data. Yet, in other settings the distribution of observables can result in non-sharp bounds that do not produce sufficiently tight set estimates or confidence regions. Thus, ideally empirical research would use all possible observable implications of their data to the extent feasible in their finite sample, basing estimation and inference on sharp bounds in order to achieve the tightest possible set estimates. The recent advances in characterizing sharp bounds can help to make this feasible. In some cases, depending in particular on the modeling assumptions, these characterizations constitute inequalities of the form (7.1) that can be implemented directly using inference techniques discussed above. Yet in other cases sharp bounds are characterized by an *extremely* large collection of conditional or unconditional moment inequalities, with possibly millions or even billions of moment functions m . In one sense, this is good news. Recall that the more inequalities that are required for any given θ to belong to the identified set, the more difficult it will be for that θ to satisfy the inequalities, and the tighter will be the resulting set. There is a challenge however in deciding how best to incorporate all of these inequalities in the analysis, and there may be far fewer observations than inequalities. It may very well seem impossible to use all of the moment inequalities implied by the sharp characterization in a finite sample. In our view this should not dissuade researchers from using partial identification methods,

perhaps based on non-sharp bounds using only a subset of all the possible inequalities. Further research on the question of how best to incorporate the identifying information of a large number of inequalities in such settings seems an important avenue for future research, combining considerations from both an identification as well as an inference standpoint. Important recent advances on the use of many moment inequalities for estimation and inference include Menzel (2014), Chernozhukov, Chetverikov, and Kato (2013), and Andrews and Shi (2015).

Finally, another important question in practice is to consider from the start exactly what are the primary objects of interest. It is sometimes beneficial to bound these objects—e.g., welfare measures or elasticities—directly, rather than estimating bounds on the underlying parameters of the model. AGQ provide a good example where the quantities of interest, seller profits and bidder surplus, are simple objects that are much easier to bound than the underlying multi-dimensional valuation distribution. The focus in BGIM on the quantiles to the wage distribution is another example. In cases such as Ho and Pakes (2014), where both the underlying parameters (the price coefficient in the referral function) and functions of those parameters (the trade-off between price, quality and distance) are of interest, deriving bounds on the latter functions can be non-trivial. Eizenberg (2014) uses a clever approach to use estimated bounds on fixed costs to simulate the effect of removing the most advanced CPU, the Pentium M, from the set of technologies available for installation in a particular time period in order to address the question of cannibalization through technological advance. Counterfactual simulation is complicated by the existence of multiple equilibria. To deal with this he computes welfare measures at each possible equilibrium to produce bounds on counterfactual welfare predictions, identifying findings that hold across all possible equilibria. An important take-away is that it can be helpful for researchers to consider a priori exactly what they wish to learn, and to then construct bounds or point estimates either for those objects directly, or for parameters that will enable direct calculation of point or set estimates for the quantities of interest.

8 Conclusion

This paper has focused on a small selection of papers that use partial identification to analyze topics of substantial economic interest. We have used these examples to point out the many benefits, and also some of the challenges, inherent in the task of applying these methods in practice. There are several other promising areas where researchers are continuing to apply these methods; we mention a few of them here before concluding.

The industrial organization literature that uses moment inequalities as one input into modelling the effects of changes in market structure continues to make substantial progress. We discussed several good examples, such as Eizenberg (2014), Nosko (2014), and Wollman (2014) in Section 2. These papers make the point that, in order to understand the impact of changes in market structure (e.g. mergers) on consumer surplus, we need to predict the resulting changes in firms' product mix and product positioning. This requires an estimate of the fixed costs of product development, which can be bounded using moment inequalities motivated by revealed preference. The findings of these papers have gained the attention of the antitrust authorities, suggesting that further work in this area could be of substantial policy importance.

Morales, Sheu, and Zahler (2015) show how moment inequalities can be used to simplify estimation of dynamic structural models. They consider exporting firms' decisions regarding which new foreign markets to enter, and assume that a firm's exports depend on how similar the new market is to its own country (gravity) and to its previous export destinations (extended gravity). They write down a dynamic multi-period model in which firm choices are functions of their past histories of choices and the choices of competitor firms. Under their assumptions it is possible to use an analog of Euler's perturbation method to difference out much of the complexity introduced by the dynamic aspect of the model, and still estimate the parameters of interest and bound the importance of extended gravity. This method seems promising for other dynamic settings.

A further example is the issue of sample selection in randomized experiments. Lee (2009) analyzes the wage effects of a large federal job training program in the U.S. He notes that the impact of a training program on wages is difficult to assess even with a randomized experiment because of a variant on the sample selection issue in BGIM: wage rates are only observed for those who are employed, and employment status is itself likely to be affected by the training program. He uses a simple procedure to bound the treatment effect of the program, identifying the incremental number of people who become employed because of the treatment, and "trimming" the tails of the wage distribution by this number to generate upper and lower bounds. This approach has the potential to be applied to other methods that are used to estimate treatment effects. For example the working paper Gerard, Rokkanen, and Rothe (2015) uses similar intuition to develop bounds on treatment effects estimated using regression discontinuity analysis where the distribution of observations across the running variable cutoff is not smooth, implying that the maintained assumptions of the usual RD design are likely to be violated.

In Sections 5 and 6 we discussed how partially identifying models may be used to study wage distributions. They may additionally be useful in the study of labor supply decisions, for example to bound labor supply

elasticities or responses to tax changes, see for example Blundell, Bozio, and Laroque (2011, 2013), Chetty (2012), and Kline and Tartari (2015). The last of these employs a nonparametric revealed preference approach to evaluate labor supply responses to welfare reform experiments. The theoretical framework is related to that advocated by Manski (2014) for partial identification of income-leisure preferences and responses to tax policy. This approach uses the same type of revealed preference arguments used in neoclassical consumer theory, which have been applied to bound demand responses and test consumer rationality as discussed in Section 3. In a different context, Barseghyan, Molinari, and Teitelbaum (2014) have recently applied revealed preference arguments to the study of decision-making under risk, using household data on insurance coverage choices. They allow for departures from expected utility theory that are motivated by developments in the theoretical literature, without imposing arbitrary assumptions on the distribution of preferences in the population. This is yet another paper exemplifying the power of revealed preference analysis to usefully bound preference parameters or counterfactual choices across a variety of economic contexts.

There are several papers that derive bounds in panel data models, and there seems to us a good deal of scope for further developments and applications. Honore and Tamer (2006) use partial identification analysis to get around the initial conditions problem in a dynamic binary random effects probit model. They show how they can obtain tight bounds on model parameters without making assumptions about the initial conditions. Rosen (2012) studies a fixed effect panel data model in which a conditional quantile restriction is imposed on time-varying unobserved heterogeneity. He shows how inequalities implied by the conditional quantile restriction can be differenced across time to obtain inequalities involving conditional moments of observable quantities from which the fixed effects are absent. Li and Oka (2015) extend similar ideas to analyze short panels with censoring. Chernozhukov, Fernandez-Val, Hahn, and Newey (2013) derive bounds on average and quantile treatment effects in a variety of nonseparable panel data models, both nonparametric and semiparametric. Pakes and Porter (2015) continue the literature on nonlinear panel data (or group) models with fixed effects, relaxing distributional assumptions on unobservables, as Ho and Pakes (2014) relaxed the assumption of i.i.d. extreme value errors used in the cross-sectional logit model. They show how to use partial identification to estimate bounds on the parameters of these models semi-parametrically, without imposing commonly-used restrictions on the joint distribution of the unobservables across choices or their correlations across time or within groups. Similar ideas can also be used to analyze firm entry decisions with cross-sectional data, as considered in Pakes (2014).

Overall, our consideration of this literature leaves us in no doubt of the potential for future researchers to use partially identifying models to study important policy-relevant questions across economic fields. As

we discussed in Section 7, some challenges remain, in particular in determining how best to incorporate the identifying information of very many conditional moment inequalities in practice, as well as how to perform computations of set estimates as efficiently as possible. Nonetheless, several insightful applications have already been executed. These methods can be used to answer interesting economic questions in situations where sufficient conditions for point identification are dubious or altogether unwarranted. They can also be used for sensitivity analysis when researchers may disagree about the veracity of identifying assumptions. The utility of partial identification seems evident, and we look forward to the development of future applications and methodological advances as they progress.

References

- ADAMS, A. (2014): “Revealed Preference Heterogeneity,” Working paper, University of Oxford and Institute of Fiscal Studies.
- AHN, H., AND J. L. POWELL (1993): “Semiparametric estimation of censored selection models,” *Journal of Econometrics*, 58, 3–29.
- ANDREWS, D. W. K., AND X. SHI (2013): “Inference Based on Conditional Moment Inequalities,” *Econometrica*, 81(2), 609–666.
- (2015): “Inference Based on Many Conditional Moment Inequalities,” Working Paper, Yale University and University of Wisconsin.
- ANDREWS, D. W. K., AND G. SOARES (2010): “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 78(1), 119–157.
- ARADILLAS-LÓPEZ, A. (2010): “Semiparametric Estimation of a Simultaneous Game with Incomplete Information,” *Journal of Econometrics*, 157(2), 409–431.
- ARADILLAS-LÓPEZ, A., AND A. GANDHI (2013): “Robust Inference of Strategic Interactions in Static Games,” Working paper, University of Wisconsin and Penn State University.
- ARADILLAS-LÓPEZ, A., A. GANDHI, AND D. QUINT (2013): “Identification and Inference in Ascending Auctions with Correlated Private Values,” *Econometrica*, 81(2), 489–534.
- ARADILLAS-LÓPEZ, A., AND A. M. ROSEN (2013): “Inference in Ordered Response Games with Complete Information,” CEMMAP working paper CWP33/13.

- ARMSTRONG, T. B. (2013): “Bounds in Auctions with Unobserved Heterogeneity,” *Quantitative Economics*, 4, 377–415.
- (2014): “Weighted KS Statistics for Inference on Conditional Moment Inequalities,” *Journal of Econometrics*, 181(2), 92–116.
- ATHEY, S., AND P. HAILE (2002): “Identification of Standard Auction Models,” *Econometrica*, 70(6), 2107–2140.
- BAJARI, P., H. HONG, AND S. P. RYAN (2010): “Identification and Estimation of a Discrete Game of Complete Information,” *Econometrica*, 78(5), 1529–1568.
- BAR, H., AND F. MOLINARI (2015): “Computational Methods for Partially Identified Models via Data Augmentation and Support Vector Machines,” In preparation.
- BARSEGHYAN, L., F. MOLINARI, AND J. C. TEITELBAUM (2014): “Inference Under Stability of Risk Preferences,” working paper, Cornell University and Georgetown University.
- BERESTEANU, A., AND C. F. MANSKI (2000a): “Bounds for MatLab,” Software at http://faculty.wcas.northwestern.edu/~cfm/bounds_matlab.zip, documentation at http://faculty.wcas.northwestern.edu/~cfm/bounds_matlab.pdf.
- (2000b): “Bounds for STATA,” Software at http://faculty.wcas.northwestern.edu/~cfm/bounds_stata.zip, documentation at http://faculty.wcas.northwestern.edu/~cfm/bounds_stata.pdf.
- BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): “Sharp Identification Regions in Models with Convex Moment Predictions,” *Econometrica*, 79(6), 1785–1821.
- (2012): “Partial Identification Using Random Set Theory,” *Journal of Econometrics*, 166(1), 17–32.
- BERESTEANU, A., AND F. MOLINARI (2008): “Asymptotic Properties for a Class of Partially Identified Models,” *Econometrica*, 76(4), 763–814.
- BERESTEANU, A., F. MOLINARI, AND D. STEEG MORRIS (2010): “Asymptotics for Partially Identified Models in STATA,” Software at http://economics.cornell.edu/fmolinari/Stata_SetBLP.zip.
- BERRY, S. (1992): “Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 60(4), 889–917.

- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63(4), 841–890.
- BERRY, S., AND E. TAMER (2006): “Identification in models of oligopoly entry,” in *Advances in Econometrics. Theory and Applications, Ninth World Congress*, ed. by R. Blundell, W. Newey, and T. Persson, vol. 2, pp. 46–85. Econometric Society Monographs: Cambridge University Press.
- BERRY, S. T. (1994): “Estimating Discrete Choice Models of Product Differentiation,” *Rand Journal of Economics*, 25(2), 242–262.
- BHATTACHARYA, J., A. M. SHAIKH, AND E. VYTLACIL (2008): “Treatment Effect Bounds Under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization,” *American Economic Review, Papers and Proceedings*, 98(2), 351–356.
- (2012): “Treatment Effect Bounds: An Application to Swan-Ganz Catheterization,” *Journal of Econometrics*, 168, 223–243.
- BJORN, P., AND Q. VUONG (1984): “Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation,” CIT working paper, SSWP 537.
- BLUNDELL, R., A. BOZIO, AND G. LAROQUE (2011): “Labor Supply and Extensive Margin,” *American Economic Review: Papers and Proceedings*, 101(3), 482–486.
- (2013): “Extensive Margin and Intensive Margins of Labor Supply: Work and Working Hours in the US, the UK and France,” *Fiscal Studies*, 34(1), 1–29.
- BLUNDELL, R., M. BROWNING, AND I. CRAWFORD (2008): “Best Nonparametric Bounds on Demand Responses,” *Econometrica*, 76(6), 1227–1262.
- BLUNDELL, R., A. GOSLING, H. ICHIMURA, AND C. MEGHIR (2007): “Changes in the distribution of male and female wages accounting for employment composition,” *Econometrica*, 75, 323–63.
- BLUNDELL, R., D. KRISTENSEN, AND R. MATZKIN (2014): “Bounding Quantile Demand Functions Using Revealed Preference Inequalities,” *Journal of Econometrics*, 179(2), 112–127.
- BRESNAHAN, T. F., AND P. J. REISS (1990): “Entry in Monopoly Markets,” *Review of Economic Studies*, 57, 531–553.

- BUGNI, F. (2010): “Bootstrap Inference for Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set,” *Econometrica*, 78(2), 735–753.
- BUGNI, F., I. CANAY, AND X. SHI (2014): “Inference for Functions of Partially Identified Parameters in Moment Inequality Models,” CeMMAP Working Paper CWP22/14.
- (2015): “Specification Tests for Partially Identified Models Defined by Moment Inequalities,” *Journal of Econometrics*, 185, 259–282.
- CANAY, I. (2010): “EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity,” *Journal of Econometrics*, 156(2), 408–425.
- CHAMBERLAIN, G. (1980): “Analysis of Covariance with Qualitative Data,” *Review of Economic Studies*, XLVII, 225–238.
- CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2013): “Testing Many Moment Inequalities,” CeMMAP working paper CWP65/13.
- CHERNOZHUKOV, V., I. FERNANDEZ-VAL, J. HAHN, AND W. NEWEY (2013): “Average and Quantile Effects in Nonseparable Panel Models,” *Econometrica*, 81(2), 535–580.
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica*, 75(5), 1243–1284.
- CHERNOZHUKOV, V., W. KIM, S. LEE, AND A. M. ROSEN (2015): “Implementing Intersection Bounds in Stata,” *Stata Journal*, 15(1), 21–44.
- CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): “Intersection Bounds: Estimation and Inference,” *Econometrica*, 81(2), 667–737.
- CHESHER, A., AND A. M. ROSEN (2012): “Simultaneous Equations Models for Discrete Outcomes: Coherence, Completeness, and Identification,” CeMMAP working paper CWP21/12.
- (2014): “Generalized Instrumental Variable Models,” CeMMAP working paper CWP04/14.
- CHESHER, A., AND A. M. ROSEN (2015): “Identification of the Distribution of Valuations in an Incomplete Model of English Auctions,” CeMMAP working paper CWP30/15.

- CHESHER, A., A. M. ROSEN, AND K. SMOLINSKI (2013): “An Instrumental Variable Model of Multiple Discrete Choice,” *Quantitative Economics*, 4(2), 157–196.
- CHETTY, R. (2012): “Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply,” *Econometrica*, 80(3), 969–1018.
- CHEKVERIKOV, D. (2011): “Adaptive Test of Conditional Moment Inequalities,” CeMMAP working paper CWP36/12.
- CILIBERTO, F., AND E. TAMER (2009): “Market Structure and Multiple Equilibria in Airline Markets,” *Econometrica*, 77(6), 1791–1828.
- DAS, M., W. K. NEWEY, AND F. VELLA (2003): “Nonparametric Estimation of Sample Selection Models,” *Review of Economic Studies*, 70, 33–58.
- DICKSTEIN, M. J., AND E. MORALES (2013): “Accounting for Expectational and Structural Error in Binary Choice Problems: A Moment Inequality Approach,” working paper, Stanford University and Princeton University.
- EIZENBERG, A. (2014): “Upstream Innovation and Product Variety in the U.S. Home PC Market,” *Review of Economic Studies*, 81, 1003–1045.
- FRECHET, M. (1951): “Sur les tableaux de corrélation dont les marges sont données,” *Ann. Univ. Lyon A*, 3, 53–77.
- FRISCH, R. (1934): *Statistical Confluence*. Publ. No. 55. Oslo: Univ. Inst. Econ.
- GALICHON, A., AND M. HENRY (2011): “Set Identification in Models with Multiple Equilibria,” *Review of Economic Studies*, 78(4), 1264–1298.
- GENTRY, M., AND T. LI (2014): “Identification in Auctions with Selective Entry,” *Econometrica*, 82(1), 315–344.
- GERARD, F., C. ROKKANEN, AND C. ROTHE (2015): “Partial Identification in Regression Discontinuity Designs with Manipulated Running Variables,” working paper, Columbia University.
- GINTHER, D. K. (2000): “Alternative Estimates of the Effects of School on Earnings,” *Review of Economics and Statistics*, 82(1), 103–116.

- GONZALEZ, L. (2005): “Nonparametric Bounds on the Returns to Language Skills,” *Journal of Applied Econometrics*, 20, 771–795.
- GRIECO, P. (2014): “Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets,” *RAND Journal of Economics*, 45(2), 303–340.
- GRONAU, R. (1973): “The Effect of Children on the Housewife’s Value of Time,” *Journal of Political Economy*, 81(2), S168–S199.
- (1974): “Wage Comparisons – A Selectivity Bias,” *Journal of Political Economy*, 82(6), 1119–1143.
- HAILE, P. A., AND E. TAMER (2003): “Inference with an Incomplete Model of English Auctions,” *Journal of Political Economy*, 111(1), 1–51.
- HARSANYI, J. (1967): “Games with Incomplete Information Played by “Bayesian” Players. Part I: The Basic Model,” *Management Science*, 14, 159–182.
- HAUSMAN, J., AND W. NEWEY (2011): “Individual Heterogeneity and Average Welfare,” CeMMAP Working Paper CWP34/13.
- HECKMAN, J. J. (1974): “Shadow Prices, Market Wages, and Labor Supply,” *Econometrica*, 42(4), 679–694.
- (1976): “The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models,” *Annals of Economic and Social Measurement*, 5, 475–492.
- (1978): “Dummy Endogenous Variables in a Simultaneous Equation System,” *Econometrica*, 46, 931–959.
- HECKMAN, J. J. (1990): “Varieties of Selection Bias,” *American Economic Review: Papers and Proceedings*, 80, 313–318.
- HECKMAN, J. J., AND B. HONORE (1990): “The Empirical Content of the Roy Model,” *Econometrica*, 58, 1121–1149.
- HECKMAN, J. J., J. SMITH, AND N. CLEMENTS (1997): “Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts,” *Review of Economic Studies*, 64, 487–537.

- HECKMAN, J. J., AND E. J. VYTLACIL (1999): “Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects,” *Proceedings of the National Academy of Sciences*, 96, 4730–4734.
- HO, K. (2009): “Insurer-Provider Networks in the Medical Care Market,” *American Economic Review*, 99(1), 393–430.
- HO, K., J. HO, AND J. MORTIMER (2012): “The Use of Full-line Forcing Contracts in the Video Rental Industry,” *American Economic Review*, 102(2), 686–719.
- HO, K., AND A. PAKES (2014): “Hospital Choices, Hospital Prices and Financial Incentives to Physicians,” *American Economic Review*, 104(12), 3841–84.
- HODERLEIN, S., AND J. STOYE (2014): “Revealed Preferences in a Heterogeneous Population,” *Review of Economics and Statistics*, 96(2), 197–213.
- HOLMES, T. (2011): “The Difusion of Wal-Mart and Economics of Density,” *Econometrica*, 79(1), 253–302.
- HONORE, B., AND A. LLERAS-MUNEY (2006): “Bounds in Competing Risks Models and the War on Cancer,” *Econometrica*, 74(6), 1675–1698.
- HONORE, B., AND E. TAMER (2006): “Bounds on Parameters in Panel Dynamic Discrete Choice Models,” *Econometrica*, 74(3), 611–629.
- IMBENS, G., AND C. F. MANSKI (2004): “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 74, 1845–1857.
- JIA, P. (2008): “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry,” *Econometrica*, 76, 1263–1316.
- KAIDO, H., F. MOLINARI, AND J. STOYE (2015): “Inference for Projections of Identified Sets,” In preparation.
- KITAMURA, Y., AND J. STOYE (2013): “Nonparametric Analysis of Random Utility Models: Testing,” CeMMAP working paper CWP36/13.
- KLEPPER, S., AND E. E. LEAMER (1984): “Consistent sets of estimates for regressions with errors in all variables.,” *Econometrica*, 52, 163–183.

- KLINE, B., AND E. TAMER (2012): “Bayesian Inference in a Class of Partially Identified Models,” Working Paper, University of Texas at Austin and Northwestern University.
- KLINE, P., AND M. TARTARI (2015): “Bounding the Labor Supply Response to a Randomized Welfare Experiment: A Revealed Preference Approach,” Working Paper. NBER.
- KREIDER, B., AND J. PEPPER (2007): “Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors,” *Journal of the American Statistical Association*, 102(478), 432–441.
- KREIDER, B., J. PEPPER, C. GUNDERSEN, AND D. JOLLIFFE (2012): “Identifying the Effects of SNAP (Food Stamps) on Child Health Outcomes When Participation is Endogenous and Misreported,” *Journal of the American Statistical Association*, 107(499), 958–975.
- LEAMER, E. E. (1981): “Is it a demand curve, or is it a supply curve? Partial identification through inequality constraints,” *Review of Economics and Statistics*, 63(3), 319–327.
- LEE, D. S. (2009): “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *Review of Economic Studies*, 76(3), 1071–1102.
- LEE, R. (2013): “Vertical Integration and Exclusivity in Platform and Two-Sided Markets,” *American Economic Review*, 103(7), 2960–3000.
- LEE, S., K. SONG, AND Y.-J. WHANG (2013a): “Testing for a General Class of Functional Inequalities,” working paper, Seoul National University.
- (2013b): “Testing Functional Inequalities,” *Journal of Econometrics*, 172(1), 14–32.
- LI, T., AND T. OKA (2015): “Set Identification of the Censored Quantile Regression Model for Short Panels with Fixed Effects,” forthcoming, *Journal of Econometrics*.
- MANSKI, C. F. (1989): “Anatomy of the Selection Problem,” *The Journal of Human Resources*, 24(3), 343–363.
- (1990): “Nonparametric bounds on treatment effects,” *American Economic Review: Papers and Proceedings*, 80, 319–323.
- (1994): “The selection problem,” in *Advances in econometrics, Sixth World Congress*, ed. by C. Sims, pp. 143–70. Cambridge, UK: Cambridge University Press.

- (1997): “Monotone Treatment Response,” *Econometrica*, 65(6), 1311–1334.
- MANSKI, C. F. (2003): *Partial Identification of Probability Distributions*. Springer-Verlag, New York.
- (2007): “Partial Identification of Counterfactual Choice Probabilities,” *International Economic Review*, 48(4), 1393–1410.
- (2008): “Partial Identification in Econometrics,” in *New Palgrave Dictionary of Economics, 2nd Edition*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave MacMillan.
- (2014): “Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy,” *Quantitative Economics*, 5(1), 145–174.
- MANSKI, C. F., AND D. S. NAGIN (1998): “Bounding Disagreements About Treatment Effects: A Case Study of Sentencing and Recidivism,” *Sociological Methodology*, 28(1), 99–137.
- MANSKI, C. F., AND J. PEPPER (2000): “Monotone Instrumental Variables: with an application to the returns to schooling,” *Econometrica*, 68, 997–1010.
- (2013): “Deterrence and the Death Penalty: Partial Identification Analysis Using Repeated Cross Sections,” *Journal of Quantitative Criminology*, 29(1), 123–141.
- MANSKI, C. F., AND E. TAMER (2002): “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70(2), 519–546.
- MARSCHAK, J., AND W. ANDREWS (1944): “Random simultaneous equations and the theory of production,” *Econometrica*, 12(143-203).
- MCFADDEN, D. (1974): “Conditional Logit Analysis of Qualitative Choice Behavior,” in *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press.
- MCFADDEN, D. L. (2005): “Revealed Stochastic Preference: A Synthesis,” *Economic Theory*, 26, 254–64.
- MENZEL, K. (2014): “Consistent Estimation with Many Moment Inequalities,” *Journal of Econometrics*, 182(2), 329–350.
- MORALES, E., G. SHEU, AND A. ZAHLER (2015): “Extended Gravity,” working paper, Princeton University.
- NASH, J. F. (1950): “Equilibrium points in n-person games,” *Proceedings of the National Academy of Sciences, USA*, 36(1), 48–49.

- NEAL, R. (2003): “Slice Sampling,” *Annals of Statistics*, 31(3), 705–767.
- NEWKEY, W. K. (1988): “Two-Step Series Estimation of Sample Selection Models,” Working Paper, Princeton University.
- NEYMAN, J., AND E. SCOTT (1948): “Consistent Estimation based on Partially Consistent Observations,” *Econometrica*, 16, 1–32.
- NOSKO, C. (2014): “Competition and Quality Choice in the CPU Market,” working paper, University of Chicago.
- OLLEY, A., AND A. PAKES (1996): “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, 64(6), 1263–1297.
- PAKES, A. (2010): “Alternative Models for Moment Inequalities,” *Econometrica*, 78(6), 1783–1822.
- (2014): “The 2013 Lawrence R. Klein Lecture: Behavioral and Descriptive Forms of Choice Models,” *International Economic Review*, 55(3), 603–624.
- PAKES, A., AND J. PORTER (2015): “Moment Inequalities for Multinomial Choice with Fixed Effects,” working paper, Harvard University.
- PAKES, A., J. PORTER, K. HO, AND J. ISHII (2015): “Moment Inequalities and Their Application,” *Econometrica*, 83(1), 315–334.
- PETERSON, A. V. (1976): “Bounds for a joint distribution function with fixed subdistribution functions: application to competing risks,” *Proceedings of the National Academy of Sciences U.S.A.*, 73(1), 11–13.
- PHILLIPS, P. C. (1989): “Partially identified econometric models,” *Econometric Theory*, 5(2), 181–240.
- POWELL, J. L. (1987): “Semiparametric Estimation of Bivariate Latent Variable Models,” Working Paper 8704, Social Systems Research Institute, University of Wisconsin-Madison.
- REIERSOL, O. (1941): “Analysis by Means of Lag Moments and Other Methods of Confluence Analysis,” *Econometrica*, 9(1), 1–24.
- ROMANO, J., AND A. SHAIKH (2008): “Inference for Identifiable Parameters in Partially Identified Econometric Models,” *Journal of Planning and Statistical Inference*, 138, 2786–2807.

- ROMANO, J., A. SHAIKH, AND M. WOLF (2014): “A practical two-step method for testing moment inequalities with an application to inference in partially identified models.,” *Econometrica*, 82(5), 1979–2002.
- ROMANO, J. P., AND A. M. SHAIKH (2010): “Inference for the Identified Set in Partially Identified Econometric Models,” *Econometrica*, 78(1), 169–211.
- ROSEN, A. M. (2008): “Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities,” *Journal of Econometrics*, 146, 107–117.
- (2012): “Set Identification via Quantile Restrictions in Short Panels,” *Journal of Econometrics*, 166(1), 127–137.
- SAMUELSON, P. A. (1938): “A Note on the Pure Theory of Consumer’s Behaviour,” *Economica*, 5(17), 61–71.
- SEIM, K. (2006): “An Empirical Model of Firm Entry with Endogenous Product-Type Choices,” *RAND Journal of Economics*, 37(3).
- SHEA, J., R. SPITZ, AND F. ZELLER (1970): “Dual Careers: A Longitudinal Study of Labor Market Experience of Women,” Discussion paper, Columbus, Ohio: Center for Human Resource Research, the Ohio State University.
- SIDDIQUE, Z. (2013): “Partially Identified Treatment Effects Under Imperfect Compliance,” *Journal of the American Statistical Association*, 502(108), 504–513.
- STOYE, J. (2009): “More on Confidence Regions for Partially Identified Parameters,” *Econometrica*, 77(4), 1299–1315.
- TAMER, E. (2003): “Incomplete Simultaneous Discrete Response Models with Multiple Equilibria,” *Review of Economic Studies*, 70(1), 147–167.
- (2010): “Partial Identification in Econometrics,” *Annual Review of Economics*, 2, 167–195.
- TANG, X. (2011): “Bounds on Revenue Distributions in Counterfactual Auctions with Reserve Prices,” *Rand Journal of Economics*, 42(1), 175–203.
- WOLLMAN, T. (2014): “Trucks without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles,” working paper, Harvard University.

Table 1: **Summary Statistics, NLSW 1976**

	Mean	Std. Devn.	N
HourlyWage	\$2.06	\$0.823	694
YearsWorked	4.67	4.93	2263
YearsEducation	11.51	2.41	2263
KidsUnder6	0.567	0.821	2263
HHAssets	\$15,751.49	\$30,692.03	2263
HusbandAnWage	\$7,061.50	\$4,095.40	2263

9 Appendix: Dataset and Variable Construction

We follow Heckman (1976) in selecting white, married women with their spouse present in the household from the sample of 5,083 women in the 1967 National Longitudinal Survey of women aged 30-44. As detailed in Appendix A-1 of that paper, we drop nonwhite women; those who were not married with their spouse present in the household; those whose husband had no income; those where the wife had a job but was not working in the survey week; and those where the wife’s work experience, education or wage rate (for working women) was not available. These rejection criteria generated a working sample of 2,263 total women, 694 of whom were working in the survey year, very close to the numbers reported in Heckman (1976) (2,253 and 701 respectively). Working women are defined as those for whom the variable “Main Activity” is listed as “working”.

The variables required for our analysis are HourlyWage, YearsWorked, YearsEducation; HusbandAnWage, KidsUnder6 and HHAssets. All are provided directly in the dataset except HHAssets which we construct as the sum of the following components of debt and assets: market value less debt on house and/or farm; value minus debt of business if applicable; value minus debt of any other real estate; amount of savings, US savings bonds, market value of stocks; amount of debt owed to respondent. Summary statistics of the resulting dataset are provided in Table 1.

We use the Heckman command in Stata to estimate the probit selection model via maximum likelihood, with log hourly wage taken as the outcome variable. We obtain the following point estimates for the

parameters of the wage equation.

$$\hat{\beta}_0 = -0.3167372, \quad \hat{\beta}_1 = 0.0187277, \quad \hat{\beta}_2 = 0.062866,$$

$$\hat{\gamma}_0 = 0.00478113, \quad \hat{\gamma}_1 = -0.0000237,$$

$$\hat{\gamma}_2 = (-0.0000343, -0.00000198, -0.532556),$$

$$\hat{\sigma} = .3373301, \quad \hat{\rho} = 0.3799919.$$