An Instrumental Variable Approach to Dynamic Models *

Work in progress. Click here for the latest version

Steven T. Berry† and Giovanni Compiani ‡

July 17, 2019

Abstract

We present a new class of methods for the identification and estimation of dynamic models with serially correlated unobservables, which typically imply that state variables are econometrically endogenous. In the context of Industrial Organization, these state variables often reflect econometrically endogenous market structure. We propose the use of Generalized Instrument Variables (GIV) methods to identify those dynamic policy functions that are consistent with instrumental variable (IV) restrictions on unobserved states. Extending popular "two-step" dynamic estimation methods, these policy functions then identify a set of structural parameters that are consistent with the dynamic model, the IV restrictions and the data. We provide computed illustrations to both single agent and oligopoly examples. We also present a simple empirical analysis that, among other things, supports the counterfactual study of an environmental policy entailing an increase in fixed costs.

1 Introduction

We propose an instrumental variable (IV) approach to the identification and estimation of dynamic models in the presence of serially correlated unobservables. Such serial correlation typically leads to dynamic state variables that are econometrically endogenous, which creates problems for identification and estimation.

We particularly consider applications to dynamic models of Industrial Organization (IO). These models often feature state variables that measure various kinds of "market structure," such as the number of firms, the number of retail outlets, the vector of current productivity levels of firms and so forth. In the dynamic context, IV methods provide a natural and

*We are grateful to Allan Collard-Wexler for generously providing the data, and to Xiaohong Chen, Liran Einav, Phil Haile, Jesse Shapiro, Paulo Somaini for helpful comments and suggestions. We are also grateful to seminar participants at NAMES 2017, Northwestern, Harvard/MIT, Stanford, Duke, 2018 and 2019 Conference on Dynamic Models, Boston University, ASSA 2019.

†Yale University, Cowles Foundation and NBER. Email: steven.berry@yale.edu.

‡University of California, Berkeley. Email: gcompiani@berkeley.edu.

economically meaningful solution to the familiar IO problem of endogenous market structure. More broadly, this paper is part of the research agenda that relates the formal identification of IO models to classic IV intuition, as in standard equilibrium models of supply and demand. The goal is to address a persistent critique of IO models that claims they are typically not well-identified. Specific examples of this agenda include Cournot style models, as in Bresnahan (1989), differentiated products demand and supply market equilibrium, as in Berry and Haile (2014), cross-sectional market structure ("static entry" models), as in Tamer (2003), and auction cost heterogeneity, as in Somaini (2015).[1]

Our paper builds on the intuition of classic "two-step" methods, following on Hotz and Miller (1993) (henceforth, HM), that distinguish between the identification of [i] the structural parameters of an underlying dynamic model and [ii] the dynamic policy function that results from the solution of that dynamic model evaluated at the true value of the structural parameters. It is this dynamic policy function that (according to the model) generates the data.

The task of identification and estimation is made much easier by an assumption that unobservable shocks are distributed independently over time. This is made clear in Rust (1987) and exploited in the HM "conditional choice probability" or "CCP" approach. In the related IO literature, the shocks are then typically assumed to also be private information.[2] Under these assumptions, the dynamic policy function is often point-identified "directly from the data." For instance, in the case of dynamic discrete choice models, estimating the policy boils down to estimation of conditional probabilities. The structural parameters are then identified as those that are consistent with the observed policy function.

However, the simplicity of these methods depends critically on the econometric exogeneity of dynamic states. Once unobservables are allowed to be serially correlated, the dynamic states become econometrically endogenous. This is because the dynamic states reflect past values of the unobservables which, due to serial correlation, are typically *not* independent of the current unobservable entering the policy function. The econometric endogeneity problem here is classic in its form: the "right hand side" state variables in the dynamic policy function are correlated with the unobservables that enter the same function.

In order to tackle the endogeneity of the dynamic states, we rely on instrumental variables. These instruments have the classic features that they [1] do not directly enter today's policy decision, [2] are assumed to be exogenous (independent of the unobservables) and yet [3] are correlated with the current state, likely because they effected past policy decisions that are correlated with present states. In a dynamic entry model, an example would be past market size or past regulatory environments that influenced past decisions to enter a market. In the presence of sunk costs, these past decisions will continue to be correlated with current market structure, even if current entry decisions are only driven by current market size and

[1] In addition, as in much of the auction literature, there are many formal IO identification arguments do not so clearly involve instrumental variables.

[2] See Pesendorfer and Schmidt-Dengler (2008), Bajari, Benkard, and Levin (2007) and Pakes, Ostrovsky, and Berry (2007) for a discussion of Hotz-Miller style methods, with pure i.i.d. private information shocks, extended to an IO dynamic oligopoly context with possibly multiple equilibria. An early review of this approach is in Ackerberg, Benkard, Berry, and Pakes (2007).

current regulations. We discuss further examples of possible instrumental variables after we have formally defined key features of the model.

Traditional IV and panel data methods face a difficult problem in our context: the dynamic policy function is derived from the "structural" dynamic model and this typically implies that the policy function is nonseparable in the serially correlated unobservable(s). The nonseparability of the policy function in unobservables creates difficulties for both identification and inference. Luckily, there is a large recent literature on the nonparametric identification of functions with nonseparable unobservables and econometrically endogenous right-hand side variables, sometimes mixed with a classic panel data structure. In the easiest possible examples for us, the dynamic policy function will be point-identified even in the presence of serial correlation, but more general cases may lead only to set identification. To consider more general cases, we leverage an existing large literature on identification and inference in partially identified models, including Manski and Tamer (2002), Tamer (2003), Manski (2003), Chernozhukov, Hong, and Tamer (2007), Berry and Tamer (2007), Ciliberto and Tamer (2009), Beresteanu, Molinari, and Molchanov (2011), Galichon and Henry (2011), Chesher (2010) and Andrews and Shi (2013).

One paper that sums up and extends an IV style literature on this topic is Chesher and Rosen (2017) (henceforth, CR), who discuss a class of "Generalized Instrumental Variable" (henceforth, GIV) methods. In addition to emphasizing an appropriate IV framework for the identification of a very broad class of dynamic policy functions, CR closely build on the work of Galichon and Henry (2011) and Beresteanu, Molinari, and Molchanov (2011) to characterize the sharp identified set. This characterization will help us build intuition about how instruments serve to (set) identify policy functions.

The identifying power of these instrumental variable methods is increased by the presence of multiple periods of data. We note that even in the absence of any instrumental variables, nonseparable policy functions can be usefully restricted purely from the presence of multiple periods of data, as in the nonseparable error, nonparametric panel data papers of Altonji and Matzkin (2005) and Athey and Imbens (2006).[3]

While we make a number of analogies to the two-step literature, there are closely related papers that emphasize full computational approaches to identification and estimation, and that also frequently allow for some form of persistent unobservables. A classic single-agent example is Keane and Wolpin (1997), who allow for persistent unobserved discrete "types" of decision-makers. A classic oligopoly example is the full computation oligopoly approach of Ericson and Pakes (1995) and Pakes and McGuire (2001), who emphasize that serially correlated unobservables are an important feature of realistic dynamic models in IO.

Full computational approaches do not often feature formal identification arguments. Work on the identification of mixture models, as in Kasahara and Shimotsu (2009) and Hu and Shum (2012),[4] does provide some formal results on identification of discrete dynamic policy functions with persistent unobservables. We show that our framework subsumes the

[3]These papers do not explicitly consider the fully dynamic problems that we consider here, but instead focus on nonparametric analogues of non-dynamic panel data-style arguments.

[4]See also Hu, Shum, Tan, and Xiao (2015).

class of models they consider as a special case. In addition, as with other panel-data style approaches, a key restriction in Kasahara and Shimotsu (2009) is that the variation in unobservables is in some well-defined sense "lower-dimensional" than the variation in the observed data. In contrast to that, our approach is applicable to settings with as many as two time periods irrespective of the dimension of the unobservable. Arcidiacono and Miller (2011) provide maximum likelihood computational methods for the structural parameters of dynamic models with discrete persistent heterogeneity, but do not address identification beyond the suggestive results of Kasahara and Shimotsu (2009). Pesendorfer (2009) proposes a Bayesian estimation method for dynamic discrete choice models with serially correlated unobservables. Additional full-solution approaches allowing for serial correlation include Blevins (2016) and Reich (2018). Neither discusses identification formally. Benkard, Bodoh-Creed, and Lazarev (2010) highlight that some interesting counterfactuals are identified directly from the policy function without having to identify the underlying structural parameters. Our approach is complementary to theirs in that it allows one to characterize the identified set for the policy function in presence of serially correlated unobservables and to perform counterfactuals that do require recovering the structural parameters in addition to the policy. An example of the latter is a change in fixed or sunk costs induced by a policy change—e.g. new environmental regulation—as in Ryan (2012). Heckman, Humphries, and Veramendi (2016) study identification and estimation of dynamic treatment effects allowing for time-invariant unobserved heterogeneity. A recent paper by Kalouptsidi, Scott, and Souza-Rodrigues (2018) shows that, in a class of dynamic discrete choice models with serially correlated unobservables, one can obtain Euler equations that point-identify flow payoffs. This approach leads to computationally light linear IV estimators of payoff parameters that are robust to serial correlation in the unobservables. However, it does not address identification of the joint distribution of the unobservables over time and thus cannot be used to perform counterfactuals requiring that as an input.

Finally, our paper is related to the large literature on distinguishing between state dependence and unobserved heterogeneity (see, e.g., Heckman and Singer (1984) and Israel (2005)). We also face the challenge of disentangling the roles of past actions and persistent unobservables in driving current outcomes. However, unlike many of these papers, we fully specify the dynamic problem solved by the agents and impose the restrictions embedded in Bellman's equation in the estimation procedure.

The rest of this paper is organized as follows. Section 2 introduces the general model and identification framework for both single-agent problems and games. Section 3 presents several Monte Carlo simulations. Section 4 contains the empirical application. Finally, Section 5 concludes. Appendix A shows that the GIV framework subsumes more general forms of persistent unobserved heterogeneity relative to the case considered in the main body of the paper. Appendices B and C provide additional details on the computed examples for the single-agent and oligopoly case, respectively.

2 Model and Identification

In this section, we present the formal model. After introducing variables and notation, we focus on the single-agent case and illustrate the GIV approach through a simple monopolist entry example. We then extend the analysis to settings with multiple agents in each market.

2.1 Variables and Notation

We consider identification of a model that generates data on a large set of markets, with one or more agents[5] per market, and a fixed (perhaps small) number of time periods denoted by $t = 1, ..., T$. We may additionally have access to some subset of variables for prior periods, $t < 1$. In the general oligopoly model, markets are indexed by i and firms within markets are indexed by j. We do not model cross-market interactions. Our simpler examples will involve a single firm per market.

In each market in each time period, each firm takes an action (or actions) denoted a_{ijt}. These actions contribute over time to the firms observed current state(s), denoted x_{ijt}. The set of feasible actions for a firm is denoted $\mathcal{A}(x_{ijt})$. As one example, in an entry model there might be a scalar action a_{ijt}, equal to one or zero, that indicates the decision to operate in the market in period $t+1$. A scalar state x_{ijt} might then be whether firm j operates in market i in period t.

There are also observed exogenous states, w_{ijt}, that evolve separately from the firms' actions. Some or all of the exogenous states may be shared across firms. In some cases, we may observe some partial information on exogenous variables from before the beginning of our full panel dataset. We denote these variables, which may later prove useful as instruments, as r_i.

In addition, there are unobserved (to us) state(s) u_{ijt} that also evolve exogenously from the actions of firms. Within market, the unobservables may be correlated both across time and firms. The u_{ijt} are the only variables that the firms observe but we do not. In the oligopoly context, we treat the serially correlated component of u_{ijt} as commonly observed by all firms. In some cases, it is also useful to model an independent (over time and firms) component that that is private information to the firm.[6]

Suppose that there are a maximum of J firms within each market. We define market-time vectors as

$$a_{it} \equiv (a_{i1t}, a_{i2t}, \ldots, a_{iJt}) \tag{1}$$

and we define the market-time vectors x_{it}, w_{it} and u_{it} in a similar fashion. As further notation, we let across-time within-market vectors of variables (and their respective supports) be denoted $a_i = (a_{i1}, \ldots, a_{iT}) \in \mathbb{A}$, $x_i = (x_{i1}, \ldots, x_{iT}) \in \mathbb{X}$, $w_i = (w_{i1}, \ldots, w_{iT}) \in \mathbb{W}$, and $u_i = (u_{i1}, \ldots, u_{iT}) \in \mathbb{U}$.

[5]Since many IO dynamic models involve firms making decisions over time, we use the words "firms" and "agents" interchangeably throughout the paper.

[6]The distinction between the full information serially correlated unobservable and the private independent unobservable is similar to the distinction between the variables ν^1 and ν^2 in Pakes, Porter, Ho, and Ishii (2015). See also Appendix A for more details on this point.

The probability that the vector u_i of unobservables (across time and firms within market) lies in the set $\mathcal{S} \subset \mathbb{U}$ is denoted by

$$\Phi(\mathcal{S}; \theta_u), \tag{2}$$

where the vector θ_u parameterizes the distribution of the vector of market unobservables across time and firms. The parameter θ_u will often, *inter alia*, control the degree of serial correlation in the unobservables. The single-period profit of firm j in market i in period t is given by the function

$$\pi_j \left(a_{it}, x_{it}, w_{it}, u_{it}; \theta_\pi\right). \tag{3}$$

The subscript j on the single-period profit function indicates the natural property that firm j's profits depend differently on its own elements of $(a_{ijt}, x_{ijt}, w_{ijt}, u_{ijt})$ as opposed to its rivals'. The unknown parameters of the single period profit function are θ_π. The full vector of unknown parameters then includes the unknown parameters of the single-period profit function and of the distribution of unobservables: $\theta = (\theta_\pi, \theta_u)$.

2.2 Single Firm per Market

We begin with the single-agent case, returning to dynamic oligopoly in section 2.5. In this special case, we treat each firm (agent) as operating in its own "market" and so we drop the j firm subscripts in $(a_{ijt}, x_{ijt}, w_{ijt}, u_{ijt})$, leaving (for example) a_{it} as the action of the firm in market i at time t. In the single-firm case, we will shorthand the phrase "firm in market i" as "firm i."

As is classic in much of the literature following on Rust (1987), we assume that the observed endogenous states of the firm evolve according to the transition probability function

$$\Gamma(x_{it+1}|a_{it}, x_{it}, w_{it}), \tag{4}$$

where Γ gives the probability of each possible future state conditional on the firm's own action and observable states. As a special case, this could describe deterministic state transitions, where some state occurs with a conditional probability of one. For instance, in a dynamic entry model, the current state (whether the firm is in or out of the market) is equal to the action taken last period. Table 1 gives some examples of actions, states and transition processes that might occur in the IO context.

Table 1: Some Single Agent IO Examples

State, x_{it}	Action, a_{it}	$\mathcal{A}(x_{it})$	Transition
Capital	Investment	$\mathbb{R}^+$	$x_{it+1} = \tilde{\lambda} x_{it} + a_{it}$
Out/In	Entry/Exit	$\{0, 1\}$	$x_{it+1} = a_{it}$
Retail	# of Stores	$\mathcal{I}^+$	$x_{it+1} = a_{it}$
Quality	R&D	$\mathbb{R}^+$	$x_{it+1} \sim f\left(x_{it}, a_{it}\right)$

Similarly, the observable exogenous states are first-order Markov, with conditional probabilities given by

$$Q(w_{it+1}|w_{it}). \tag{5}$$

In our single-firm examples, we will focus on the special case where the firm-level unobservable u_{it} follows a first-order Markov process. Our leading example here will be a model of first-order serial correlation where u_{it} is a scalar that obeys

$$u_{it} = \rho u_{it-1} + \nu_{it}. \tag{6}$$

In the simplest case, the period t vector ν_{it} innovation might be assumed to have a simple parameterized distribution. The parameter θ_u then includes those parameters plus the serial correlation parameter ρ. In Appendix A, we consider more general forms of unobserved heterogeneity, including cases with a persistent component and an idiosyncratic component in the spirit of Heckman and Singer (1984), Keane and Wolpin (1997), Arcidiacono and Miller (2011) and related literature.

In this case, the firm's dynamic problem is given by the classic Bellman equation:

$$\begin{aligned} &V\left(x_{it}, w_{it}, u_{it}\right) = \\ &\max_{a_{it} \in \mathcal{A}(x_{it})} \left(\pi\left(a_{it}, x_{it}, w_{it}, u_{it}; \theta_\pi\right) + \delta E_{\theta_u}\left[V\left(x_{it+1}, w_{it+1}, u_{it+1}\right) | a_{it}, x_{it}, w_{it}, u_{it}\right]\right). \end{aligned} \tag{7}$$

where δ denotes the discount factor and V the value function. The expected value function in this expression is

$$\begin{aligned} &E_{\theta_u}\left[V\left(x_{it+1}, w_{it+1}, u_{it+1}\right) | a_{it}, x_{it}, w_{it}, u_{it}\right] = \\ &\int\int\int V\left(x_{it+1}, w_{it+1}, u_{it+1}\right) d\Gamma(x_{it+1}|a_{it}, x_{it}, w_{it}) dQ(w_{it+1}|w_{it}) d\tilde{\Phi}(u_{it+1}|u_{it}; \theta_u) \end{aligned} \tag{8}$$

where $\tilde{\Phi}(u_{it+1}|u_{it}; \theta_u)$ denotes the conditional distribution of u_{t+1} given u_t. Note that this is similar to Rust (1987), but we do not make Rust's full conditional independence assumption: we do not drop the conditioning on the past unobservable in $\tilde{\Phi}(u_{it+1}|u_{it}; \theta_u)$. In the single-agent case, there is a unique solution for the value function and we assume conditions such that there is a unique policy function consistent with that value function.[7] We let σ denote this policy function, so that

$$a_{it} = \sigma\left(x_{it}, w_{it}, u_{it}\right), \quad \sigma \in \mathcal{F}. \tag{9}$$

In many cases, reasonable assumptions on the single-period return function and the transition processes imply that the policy function must obey certain qualitative restrictions, such as monotonicity. These restrictions can then be imposed on the set of possible policy functions $\mathcal{F}$.

For the identification argument presented next, it will also be useful to define the different (counterfactual) policy functions that would be generated by any other possible parameter vectors $\theta = (\theta_\pi, \theta_u)$. In the single-firm case, these policy functions, generated by the model and the unique solution to Bellman's equations, will be denoted by

$$a_{it} = \sigma_\theta\left(x_{it}, w_{it}, u_{it}\right). \tag{10}$$

[7]For the technical conditions guaranteeing a unique policy function, see Stokey, Lucas, and Prescott (1989).

2.3 Identification of the Single-Agent Model

We focus first on identification in the single-agent case. For purposes of identification, we assume that we observe the true distribution of the data, which we denote by

$$P(a_i, x_i, w_i, r_i)$$

This is equivalent to seeing a T period panel on a very large (in fact, infinite) cross-section of firms or agents. We look to identify (possibly set-identify) the parameters θ. Nothing in our general discussion of identification requires these to be finite dimensional, but in practice we consider only finite-dimensional parametric models. Note that in the case where (a_i, x_i, w_i) are discrete, the single-period profit function may be fully flexibly characterized by a finite number of parameters.

The potential instruments in the model consist of the exogenous variables

$$z_i = (r_i, w_i)$$

The critical assumption that allows for our instrumental variable approach is independence of the instrument and the unobservables:[8]

$$z_i \perp u_i$$

Table 2 gives some ideas of possible instruments in different contexts. As is usual with discussions of potential instruments, the required independence assumption may be more or less appropriate in different real-world cases.

Table 2: Examples of Possible Instruments

State	Example Instruments
Capital	Past investment cost
Out/In of Market	Past market population, past regulation
# of Stores	Distance from headquarters, interacted with time
Quality	Past R&D shocks, age of firm

Identification in the case of no serial correlation and discrete actions is considered in Magnac and Thesmar (2002) and related papers. Problems with low-dimensional discrete persistent unobservables can sometimes make use of the results on the identification of mixture models in Kasahara and Shimotsu (2009) and related papers. These results aid in the identification of the policy function and of the transition matrices in (4)-(5) (especially if these are extended to depend on discrete persistent unobservables). Even without persistent unobservables,

[8]While we focus on this restriction throughout the paper, CR show that the GIV approach may also be applied under weaker assumptions, such as mean or quantile independence.

there are not typically formal point-identification results for models with no serial correlation and mixtures of continuous and discrete variables, which is why Bajari, Benkard, and Levin (2007) ("BBL") relies on set-identification, as do we. Unlike BBL, we characterize the sharply identified set for the structural parameters in the presence of persistent unobservables using the "generalized instrumental variable (GIV)" framework of CR.[9]

The broad idea is to (set-)identify the policy function from classic instrumental variables conditions, extended to cases where the policy function is highly nonlinear in the states. The GIV framework allows us to deal with the following complications arising in many dynamic models of interest:

1. the incompleteness of the model, i.e. the fact that the exogenous variables do not uniquely pin down the endogenous variables (see Tamer (2003));
2. the fact that, if the dynamic states and actions are discrete—as in entry/exit models—the policy function is known to be generally only partially identified in the absence of a model for the endogenous explanatory variables (see Chesher (2010));
3. lack of point-identification of the parameters, even in the absence of problems 1 and 2.

In applications, we may have all or none of these problems. If the model and data generating process in fact imply point-identification, then the sharply identified set will collapse to the true parameter value.

In the single agent case, an incomplete model can follow from the presence of unknown initial conditions.[10] Traditionally, solutions to the initial conditions problem include either [i] parameterizing the initial joint distribution of states and unobservables or [ii] specifying some process for the past history of the firm that uses the model parameters to construct that same initial joint distribution.[11] We argue that if the parameterization in method [i] is so flexible as to not impact the resulting identified parameter set, then we might just as well look for the sharply identified set that does not restrict the initial distribution.

As in Tamer (2003), discrete actions naturally lead to conditions on sets of unobservables that give a particular policy a_{it}. In particular, following CR, and using similar notation, if the sequence (a_i, x_i, w_i) occurs, then u_i must be in the inverse image set

$$\mathcal{U}(a_i, x_i, w_i, \sigma) = \{u_i \colon \sigma(x_{it}, w_{it}, u_{it}) = a_{it}, \forall t\}.$$

The condition

$$\{u_i \in \mathcal{U}(a_i, x_i, w_i, \sigma)\}$$

is then a necessary condition for observed event (a_i, x_i, w_i). If the model is incomplete, however, that condition is not sufficient for the event. A pair of policies and parameters for

[9]Recall from the references above that these authors in turn build on very large literature in set-identification and in the use of random sets to characterize sharply identified regions of parameters.

[10]See Anderson and Hsiao (1981), Arellano and Bond (1991) and Blundell and Bond (1998), among others. Honoré and Tamer (2006) emphasize how the initial conditions problem leads to partial identification in nonlinear dynamic panel data models.

[11]Collard-Wexler (2014) employs both solutions. While the results of his counterfactuals are robust, a few parameter estimates vary substantially across the two methods, suggesting that the way in which the initial conditions problem is addressed matters in general.

the unobservables, $(\sigma(x_{it}, w_{it}, u_{it}), \theta_u)$, is then in the identified set if and only if for all closed sets $\mathcal{S} \in \mathbb{U}$ and for all z

$$\Pr\big(\mathcal{U}(a_i, x_i, w_i, \sigma) \subseteq \mathcal{S} \mid z\big) \leq \Phi(\mathcal{S}; \theta_u). \tag{11}$$

In this last equation, the left-hand side is the conditional probability of the outcomes $y_i = (a_i, x_i)$, which, according to σ, have $\{u_i : u_i \in \mathcal{S}\}$ as a necessary condition. For a given σ and z, this probability is observed in the data. The right-hand side is the probability of that necessary condition wrt the distribution of u_i, which by assumption does not depend on z.

CR shows that to obtain the sharp identified set for θ we only need to check certain sets, $\mathcal{S} \in Q(\sigma, z_i)$. The collection of sets $Q(\sigma, z_i)$ is the "core determining set" as defined in CR and earlier work, i.e. the minimal collection of closed sets $\mathcal{S} \in \mathbb{U}$ that yields the sharp identified set for θ. These include the unions of overlapping sets of the form $\mathcal{U}(a_i, x_i, w_i, \sigma)$, excluding cases of strict subsets. For simple low-dimensional discrete problems, $Q(\sigma, z_i)$ can be easy to compute and not "too big," but it can otherwise grow very (indeed infinitely) large.

The CR result characterizing the core determining sets in the space of the unobservables builds on earlier similar results that apply to the space of observables (see Galichon and Henry (2011)). For us, the result is useful because it completely characterizes the inequality restrictions that define the sharply identified set of policy functions. In practice, however, the dimension of these restrictions can grow quite large in realistic problems. The CR characterization is therefore perhaps most useful because we can compute sharply identified sets in smaller illustrative examples, which in turn builds some intuition for more complex cases. When the dimension of the unobservables is relatively large—as in the empirical application in Section 4 where we have $T = 12$ time periods—using simulation methods to approximate the right-hand side of (11) is often useful.[12]

When necessary conditions are actually *necessary and sufficient* for particular (sets of) actions, then the associated inequalities become strict equalities. In a complete model, all of the necessary conditions are equalities. However, as usual, this does not guarantee that the parameters are point-identified and so, in the absence of a proof of point-identification, we might still want to consider set-identification.

We next define the set of policy functions that are identified exclusively by the IV conditions and the data, with no use of the dynamic model. In particular, for a given θ_u and a given data generating process, we define the set of σ functions that satisfy condition (11) $\forall \mathcal{S} \in Q(\sigma, z)$ and $\forall z$ as

$$\Sigma^{IV}(\theta_u) \subseteq \mathcal{F}. \tag{12}$$

These identified policy functions then help us to define the identified set for the structural dynamic parameters. As noted, for any $\theta = (\theta_\pi, \theta_u)$, we can use the Bellman equation to compute the implied policy $\sigma_\theta(x_{it}, w_{it}, u_{it})$. For $\theta = (\theta_\pi, \theta_u)$, this policy is

$$\sigma_\theta(x_{it}, w_{it}, u_{it}) \equiv$$

[12] See Berry (1992) and Ciliberto and Tamer (2009) for other uses of simulation in models characterized by moment conditions.

$$\underset{a_{it}\in\mathcal{A}(x_{it})}{\text{argmax}}\Big(\pi(a_{it},x_{it},w_{it},u_{it},\theta_\pi)+\delta E_{\theta_u}\big[V(x_{it+1},w_{it+1},u_{it+1})|a_{it},x_{it},w_{it},u_{it}\big]\Big).$$

The sharply identified set of parameters is then

$$\Theta_{ID}\equiv\{\theta=(\theta_\pi,\theta_u)\colon \sigma_\theta(x_{it},w_{it},u_{it})\in\Sigma^{IV}(\theta_u)\} \tag{13}$$

This imposes both the dynamic model and the GIV restrictions. This is the sharply identified set because any θ in this set generates a policy function that cannot be rejected by the data plus the IV condition.

The identifed set in (13) is related to the classic HM style two-step approach. Note that we could

1. identify a (set of) policy function(s) that are consistent with the data and the IV restrictions and then
2. see which structural parameters are consistent with the identified policy function(s). These are the identified θ.

In contrast to HM, complications in our case include the presence of parameters for the unobservables (necessary at least to model the degree of serial correlation) and the use of IV methods in the first step, as opposed to directly "fitting the policy to data." It is possible that, in some cases, one might point-identify the policy function in step 1. In this case, our procedure is just like a 2-step HM method, except for running a GIV first stage. For example, when the choice variable is continuous, point identification of the policy might be attained under the assumptions in Chernozhukov and Hansen (2005), even using only one transition.

However, there is no need to use a HM-style two-step approach. As a computational alternative, one could search over the space of the structural parameters. Specifically, for each candidate θ, one could

1. compute $\sigma_\theta(x_{it},w_{it},u_{it})$ via the contraction mapping and then
2. test whether σ_θ survives the IV restrictions applied to the data. If so, that particular θ is in the identified set, otherwise not.

2.4 A Simple Example

In this subsection, we consider a minimal single-agent model that illustrates identification via GIV restrictions. In this example, the state is whether a firm is "In" or "Out" of the market in the prior period, $x_{it}\in\{0,1\}$, and the action today is whether to be active in the market today, $a_{it}\in\{0,1\}$.

We specify single-period payoffs as follows:

$$\begin{aligned}\tilde{\pi}(0,1)&\equiv\tilde{\pi}(0,0)=0\\ \tilde{\pi}(1,1)&=\bar{\pi}(w_{it})-\epsilon_{it}\\ \tilde{\pi}(1,0)&=\tilde{\pi}(1,1)-\gamma\end{aligned} \tag{14}$$

where $\tilde{\pi}(a_{it},x_{it})$ denotes the deterministic part of the profit function as a function of the action and current state, ϵ_{it} is a shock which we might think of as reflecting variation in

per-period fixed costs, and γ represents entry costs. In the example, we assume that fixed costs follow a first-order autocorrelation process,

$$\epsilon_{it} = \rho\epsilon_{i,t-1} + \nu_{it}\sqrt{1-\rho^2}, \tag{15}$$

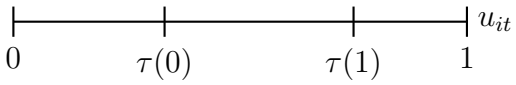
where ν_{it} is distributed standard normal. In the simplest case, we assume away any variation in exogenous profit shifters w_{it}. The resulting model then has three structural parameters: $\bar{\pi}, \gamma$, and ρ. The policy function that generates the data is

$$a_{it} = \sigma(x_{it}, u_{it}) \tag{16}$$

where $u_{it} \sim \text{Unif}(0,1)$ can be viewed as the quantile of ϵ_{it}. We assume that the dynamic model generates the monotonicity results that σ is weakly increasing in x_{it} and weakly decreasing in u_{it}.

If we focus on only one period of data, the policy function in (16) is a nonparametric binary choice model with endogeneity and monotonicity restrictions, similar to Chesher (2010). Given monotonicity in u_{it}, the policy function is fully described by two policy cutoffs, $\tau(x)$, for $x \in \{0,1\}$, as illustrated in Figure 1.

Figure 1: Policy Cutoffs in the One Period Case



As in Chesher (2010), even one period of data will generate nontrivial bounds on the policy function. As a simple example of GIV restrictions, Table 3 illustrates the inverse image sets associated with the example (in column 2) as well as the inequalities implied by the GIV restrictions (in the last columns of the table). Note that [1] there are nontrivial bounds even

Table 3: Inverse Image Sets & Inequalities for the One-Period Example

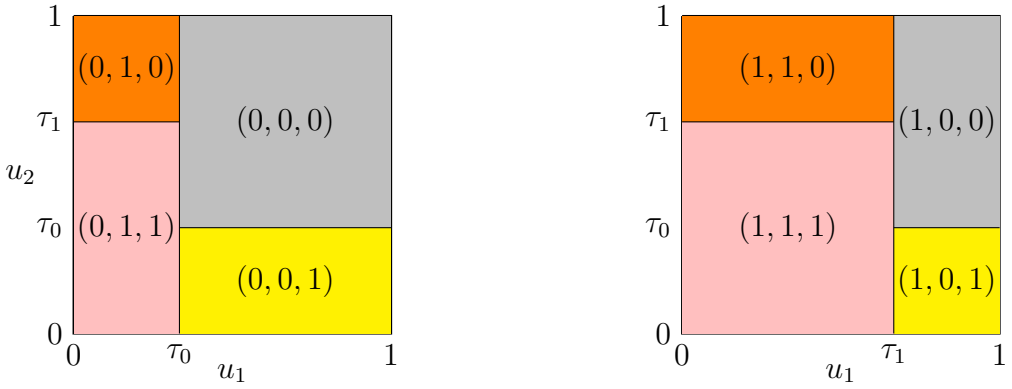
a	x	$\mathcal{S} = \mathcal{U}(a,x,\sigma)$	$\Pr\big(\mathcal{U}(a_i,x_i,\sigma) \subseteq \mathcal{S} \mid z\big)$	$\leq \Phi(\mathcal{S};\theta_u)$
1	1	$(0,\tau(1))$	$\Pr((1,1)\|z) + \Pr((1,0)\|z)$	$\leq \tau(1)$
1	0	$(0,\tau(0))$	$\Pr((1,0)\|z)$	$\leq \tau(0)$
0	1	$(\tau(1),1)$	$\Pr((0,1)\|z)$	$\leq 1-\tau(1)$
0	0	$(\tau(0),1)$	$\Pr((0,0)\|z) + \Pr((0,1)\|z)$	$\leq 1-\tau(0)$

in the absence of IVs, but [2] instrumental variable variation is helpful to tighen those bounds. Note also that, by themselves, the restrictions in Table 3 place no restrictions on θ_u. It is not surprising that it is impossible, in the example, to learn anything about serial correlation

from restrictions on the single-period policy function. However, with multiple periods of data, restrictions on the policy function may rule out some values of serial correlation, even without reference to the structural model.

We next consider two periods of data. Now there are eight elemental[13] inverse image sets $\mathcal{U}(a_i, x_i, \sigma)$, in the space of (u_{i1}, u_{i2}), that depend on $(x_{i1}, a_{i1} = x_{i2}, a_{i2})$. These are illustrated in Figure 2.

Figure 2: Elemental Inverse Image Sets Labeled as (x_{i1}, a_{i1}, a_{i2})



The core determining collection of sets also includes unions of partially overlapping elemental sets, which we show in Appendix B.

The core determining sets for this simple example allow us to build some intuition about identification in this class of models. The left panel of Figure 2 gives the four elemental sets associated with the initial condition $x_{i1} = 0$, while the right panel gives the sets associated with $x_{i1} = 1$. Conditional on the initial condition, the model is complete (the sets do not overlap), but across initial conditions the sets do overlap, reflecting incompleteness. For example, there are values of (u_{i1}, u_{i2}) that are consistent with both the sequence $(1, 1, 1)$ and the sequence $(0, 0, 0)$. If the initial x_{i1} was exogenous, the model would be complete.

Recall that the probability of the each of the eight events associated with different (x_{i1}, a_{i1}, a_{i2}) must be less than the probability weight placed by the distribution of (u_{i1}, u_{i2}) over the regions of the elemental sets, with further restrictions from the additional core determining sets. In Figure 2, the joint density of (u_{i1}, u_{i2}), which varies with the serial correlation parameter ρ, places the relevant probability weight over the various regions. Note that in this example, with two transitions, we can rule out some values of ρ without any use of the dynamic model. For example, perfect correlation, $\rho = 1$, collapses the joint density down to a straight line across the diagonal of each box. This is rejected by any value of an instrument

[13]We define a set to be "elemental" if it associated with a single value of the outcomes (a_i, x_i).

that places any weight on the observed outcome $(0,1,0)$ or $(1,0,1)$.

As another piece of intuition, consider an instrument associated with a probability equal to one for initial condition $x_{i1} = 1$. The event probabilities associated with the right-hand side panel of Figure 2 then sum to one and all of the associated inequality restrictions hold with equality. These equalities are exactly the same as those that would be implied by maximum likelihood applied to the model with an exogenous initial condition $x_{i1} = 1$. Thus, if the model is complete and MLE point-identifies the parameters $(\tau(0), \tau(1), \rho)$, then GIV identifies the same parameter values.

2.5 Dynamic Oligopoly

IIn the oligopoly case, each firm's equilibrium policy is its single-agent best reply to its rivals' equilibrium strategies. The firm still solves a value function problem similar to (7), but its expectations of the future evolution of endogenous market states depend on its action and the equilibrium actions of its rivals.

In the oligopoly case, we require our original notation of i for the market and j for the firm. If the equilibrium policies of firm j's rivals are given by the function σ_{-j}, then the firm's expected equilibrium state transition probabilities are given by

$$\tilde{\Gamma}_j \left(x_{it+1} | a_{ijt}, x_{it}, w_{it}, \sigma_{-j}(x_{it}, w_{it}, u_{it}) \right). \tag{17}$$

This notation allows for a rich set of possible state transitions models, including oligopoly variations on our earlier single firm examples.

Firm j's equilibrium Bellman equation then depends on the equilibrium strategies of its rivals:

$$V_j \left(x_{it}, w_{it}, u_{it}, \sigma_{-j}, \theta \right)) = \tag{18}$$

$$\max_{a_{ijt} \in \mathcal{A}(x_{ijt})} \left(\pi_j \left(a_{it}, x_{it}, w_{it}, u_{it}; \theta_\pi \right) + \delta E_{\theta_u} \left[V_j \left(x_{it+1}, w_{it+1}, u_{it+1}, \sigma_{-j} \right) | a_{it}, x_{it}, w_{it}, u_{it} \right] \right).$$

The expected Bellman's equation is

$$E \left[V_j \left(x', w', u' \right) | a, x, w, u, \sigma_{-j} \right] = \tag{19}$$

$$\int \int \int V_j \left(x', w', u', \sigma_{-j} \right) d\tilde{\Gamma}(x' | a_j, x, w, \sigma_{-j}(x, w, u)) dQ(w_{it+1} | w_{it}) d\tilde{\Phi}(u' | u; \theta_u). \tag{20}$$

Associated with this dynamic program is a "best response" strategy for firm j, which we assume is unique, denoted by $\bar{\sigma}_j(\sigma_{-j}, \theta)$. The vector of "best response" strategies is then the J-vector

$$\bar{\sigma}(\sigma, \theta) = (\bar{\sigma}_1(\sigma_{-1}, \theta), \ldots, \bar{\sigma}_J(\sigma_{-J}, \theta)).$$

Any equilibrium strategy, σ^*, then must satisfy the fixed point

$$\sigma^* = \bar{\sigma}(\sigma^*, \theta). \tag{21}$$

We can then define the set of possible equilibrium policy functions (strategies) as

$$\Sigma^{EQ}(\theta) = \{\sigma^* \colon \sigma^* = \bar{\sigma}(\sigma^*, \theta)\}.$$

We adopt the same approach as in earlier papers and assume that, even if the underlying model admits multiple equilibria, the firms themselves always play the same policy function when at the same state vectors. The true policy function that generates the data is then an element of the set $\Sigma^{EQ}(\theta^0)$, where θ^0 is the true parameter that generates our data.

The sharply identified set of parameters in the oligopoly case is the same as in the single agent case, except with the further restriction that the policy associated with θ is an equilibrium policy:

$$\Theta_{ID} \equiv \{\theta = (\theta_\pi, \theta_u) \colon \text{ there exists } \sigma^* \in \Sigma^{EQ}(\theta) \text{ such that } \sigma^* \in \Sigma^{IV}(\theta_u)\}. \tag{22}$$

That is, a parameter vector θ is in the identified set if there is a policy that both [i] is not rejected by the IV restrictions and the data (given θ_u) and [ii] is an equilibrium strategy given θ.

The connection to two-step methods is again clear. In a "first stage," we could find the set of policies that survive the IV restrictions. For any θ, we can see if any of those policies are in the equilibrium set of policies. This can be tested by checking the dynamic "best reply" condition for that policy. Importantly, this does not require us to calculate the full equilibrium set of policies for a given structural parameter vector.

Even putting issues of inference aside, finding the identified set can be considerably more computationally complicated than the single-agent case. However, if the IV restrictions point identify the policy functions, the method becomes very similar to the two-step case and becomes especially close to the framework of BBL.

As an example, consider a two-firm capital accumulation game involving policy functions that would be associated with a duopoly version of the dynamic capital accumulation model that motivates Olley and Pakes (1996). In particular, assume that the equilibrium policy functions of the two firms are given by

$$i_{ijt} = \sigma(k_{it}, w_{it}, u_{it}), \tag{23}$$

where the states x_{it} are here denoted as the capital stocks $k_{it} = (k_{i1t}, k_{i2t})$, the exogenous shifters are still w_{it}, the serially correlated unobservables are $u_{it} = (u_{i1t}, u_{i2t})$ and the actions are investments $i_{it} = (i_{i1t}, i_{i2t})$. Assume further that k_{ijt} and i_{ijt} are continuously valued variables. For the purposes of a simple example, further suppose that we can establish (or are willing to assume) that the equilibrium policy functions are continuous and injective in u_{it}, so that we can write

$$u_{ijt} = \sigma_j^{-1}(k_{it}, w_{it}, i_{it}), \tag{24}$$

where u_{ijt} takes on continuous values. This takes the form of a two equation "quantile IV model," where point identification might be achieved under the conditions of Chernozhukov and Hansen (2005). Given the continuously valued u_{ijt}, a simple specification of serial correlation could then allow one to recover the serial correlation parameter θ_U. In this illustrative lucky case, a given θ is in the identified set if and only if the identified policy satisfies the fixed-point in (21).

The procedure of the last paragraph is similar to BBL, with two differences. First, the policy functions are identified via GIV conditions. Second, we obtain the sharply identified set via

the fixed-point condition, whereas BBL (for reasons of practicality and inference) suggest the use of various inequality conditions that are motivated by the same best-response (fixed point) condition. Our method is easily modified to use BBL style inequalities, which may be useful if those create a large computational gain in a given situation.

Even if the GIV conditions do not point-identify the equilibrium policy, if the set $\Sigma^{IV}(\theta_u)$ is "small enough," it may be practically possible to test the fixed point condition for each policy in the identified set. In this case, we can note the relative computational ease of our identification approach versus the implicit identification algorithm in, for example, Pakes and Ericson (1996) and some other full-computational methods. Their algorithm searches for the equilibrium policy function at each θ, which is a particularly difficult task under possible multiple equilibrium.[14] However, in our set up the search for possible equilibrium policies is restricted to those that satisfy the IV restrictions, which may greatly aid the equilibrium computation. For example, sometimes theoretically possible equilibrium are quite extreme (for example, involving one firm that never enters), but these need not be considered if they are rejected by the data/IV conditions.

The BBL idea of forward simulating the expected future value function (in the manner of Hotz, Miller, Sanders, and Smith (1994)) may also ease the computational burden of checking the fixed-point condition. In this case, for a given θ_u and a policy vector in $\Sigma^{IV}(\theta_u)$, one can sometimes forward simulate the future expected value function and in some cases this is a linear function of θ_π. Checking the fixed-point condition for a given θ_π (holding rivals' policy and θ_u fixed), can then be a relatively simple static calculation.

In Appendix C, we consider a more complicated discrete oligopoly example, and provide a computed example of an identified set in that case. In the discrete example, we have a particular worry not just about uniqueness of equilibrium, but also about the existence of equilibrium. Thus, in the discrete oligopoly case, it may be particularly useful to introduce an explicit private-information independent shock to the profitability of an action, in addition to the serially correlated unobserved state.[15] To this end, in the example we make use of a more general structure for the unobservables which is discussed in Appendix A.

3 Monte Carlo Simulations

In this section, we illustrate the approach through several Monte Carlo simulations based on a simple single-firm entry and exit model. Appendix C contains a computed example for the oligopoly case.

We consider the same monopoly entry model as in the example of Section 2.4.[16] We set

[14] See the online appendix of Doraszelski and Satterthwaite (2010) for examples of multiplicity. Further examples are in Pesendorfer and Schmidt-Dengler (2010). Borkovsky, Doraszelski, and Kryukov (2010) provides a homotopy method for exploring a range of possible multiple equilibrium.

[15] As noted, these two unobservables correspond to the unobservables labeled ν^1 and ν^2 in Pakes, Porter, Ho, and Ishii (2015).

[16] Note that in this model we are free to normalize the mean and variance of the unobservables given that we are specifying the deterministic profits in a fully nonparametric fashion.

$\tilde{\pi}(1,0) = -1$ and $\tilde{\pi}(1,1) = 0.5$, so that the sunk cost γ is 1.5. Further, the correlation for the unobservables is set to 0.75, i.e. there is persistent unobserved heterogeneity. We generate time-invariant excluded instruments z taking the values $\{0,1\}$ with equal probabilities and set $x_{i1} = z_i$ for a fraction of markets in the data equal to 0.50 or 0.75. We call this fraction "IV strength" and note that it is equal to the square root of the R^2 coefficient in the regression of the endogenous state x_{i1} on the IV (plus a constant).[17] Except when otherwise noted, the model does not have exogenous covariates w. In what follows, we show how the identified set for the structural parameters changes as we vary the sample size, the number of time periods in the data, and the variation in exogenous covariates. We also report the identified sets for counterfactuals outcomes. In each Monte Carlo design, we draw 50,000 cross-sectional markets. Setting the cross-sectional dimension of the panel to a large number allows us to abstract from sampling error in estimation of the identified set. In the application of Section 4, we show how recent developments in the literature on moment inequalities may be applied to obtain confidence sets for quantities of interest.

3.1 $T = 2$ case

With just two time periods, we are able to list all of the GIV restrictions implied by the model and thus to obtain sharpness of the identified set. For each value of IV strength, we compute the three-dimensional identified set for $(\tilde{\pi}(1,0), \tilde{\pi}(1,1), \rho)$ and we plot its projections onto the space of profits $(\tilde{\pi}(1,0), \tilde{\pi}(1,1))$ and the space of sunk cost and correlation parameters. As a comparison, we also compute the point-estimates given by the two-step HM procedure which assumes no serial correlation in the unobservables.

Figures 3 and 4 show that, when IV strength is low, the identified set is quite large. In particular, while the GIV three-dimensional set does not contain the HM estimate, its projection onto the space of profits does, i.e. the GIV restrictions do not rule out the profit parameters implied by the HM procedure. In contrast, when IV strength is high, the GIV set shrinks considerably and both projections do not include the HM estimate, as shown in Figures 5 and 6. In particular, the HM procedure tends to underestimate both profit parameters (besides, of course, not capturing the serial correlation in the unobservables).

[17] Specifically, IV strength = 0.50 corresponds to $R^2 = 0.25$ and IV strength = 0.75 corresponds to $R^2 = 0.56$.

Figure 3: Low IV strength

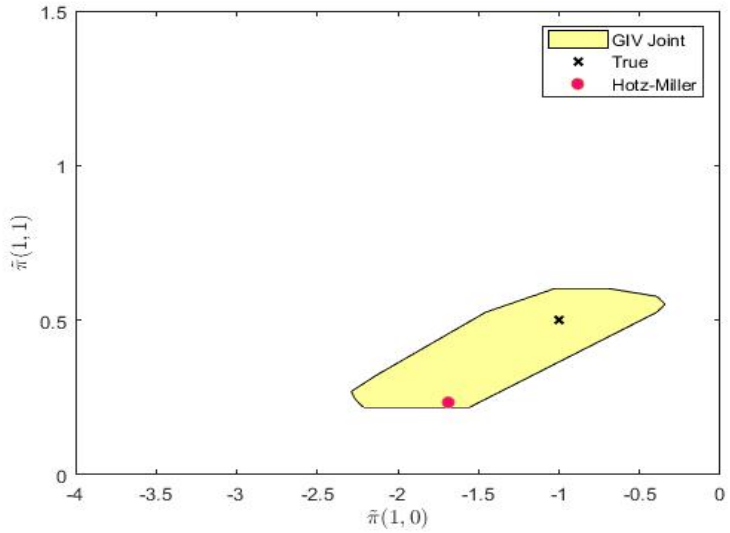


Figure 4: Low IV strength

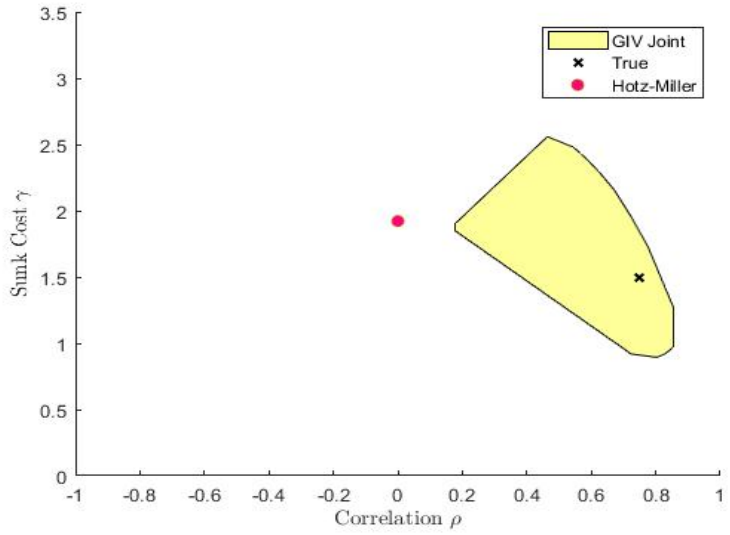


Figure 5: High IV strength

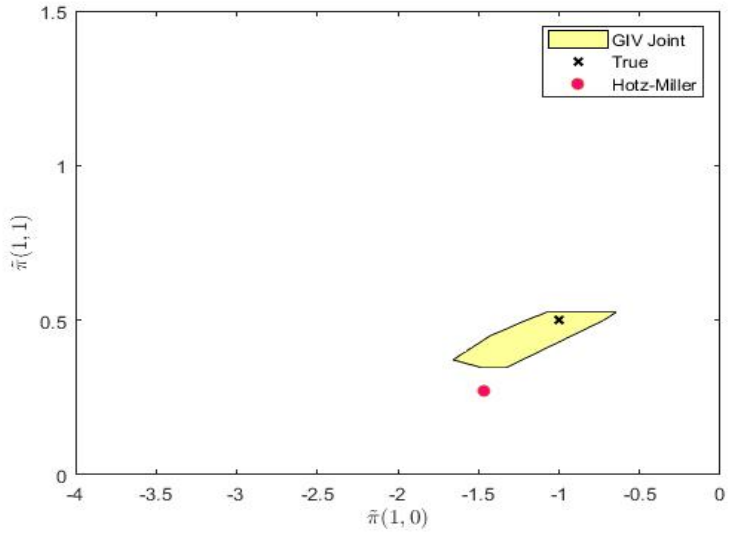
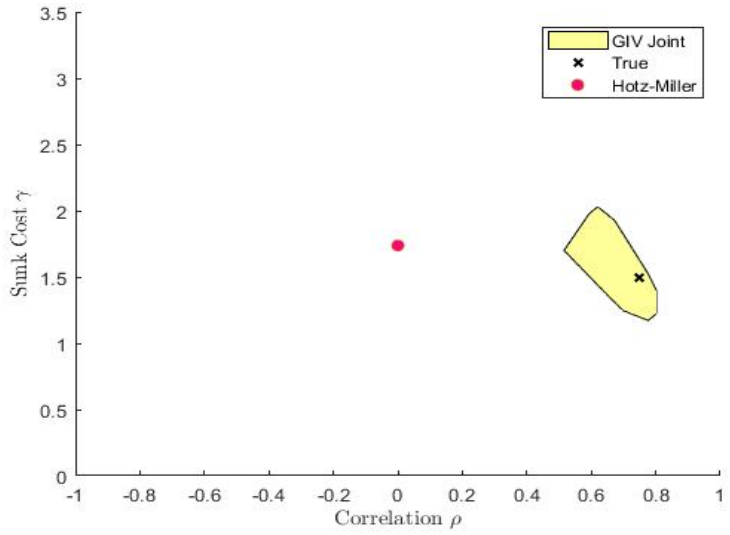


Figure 6: High IV strength

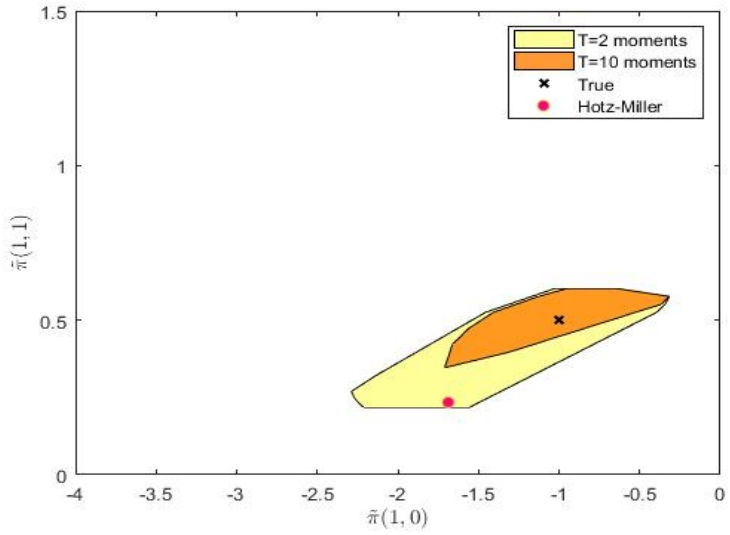


3.2 $T = 10$ case

We now consider the case where the data has $T = 10$ time periods. In addition to the inequalities corresponding to the sharp two-period GIV identified set, we impose inequalities

corresponding to several observable events over all of the 10 time periods.[18] The goal is to illustrate how various patterns in the data help bound the model parameters. Specifically, we use the events "at least one entry," "at least one exit," "at least one entry and one exit," and "the number of firms does not change for at least six consecutive periods." We would expect the latter event to help rule out values of ρ close to zero, since it yields inequalities where the sample probabilities on the left-hand side are large (because the data exhibits a lot of persistence) and the model necessary conditions on the right-hand side are relatively small (because when ρ is close to zero the model predicts little persistence in the observables). Indeed, Figures 7 to 10 show that imposing these additional restrictions substantially shrinks the GIV identified set and that this effect is particularly pronounced for relatively low values of ρ.

Figure 7: Low IV strength



[18]Since it is hard to obtain a closed form for the model probabilities associated with these events, we approximate the probabilities via simulation.

Figure 8: Low IV strength

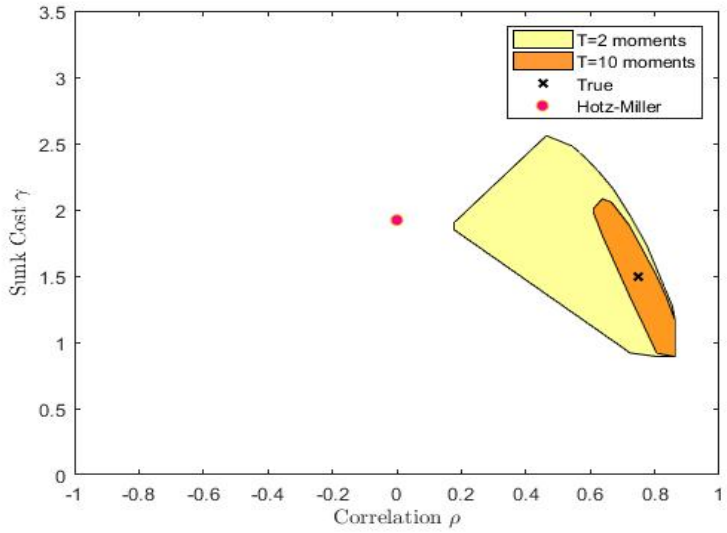


Figure 9: High IV strength

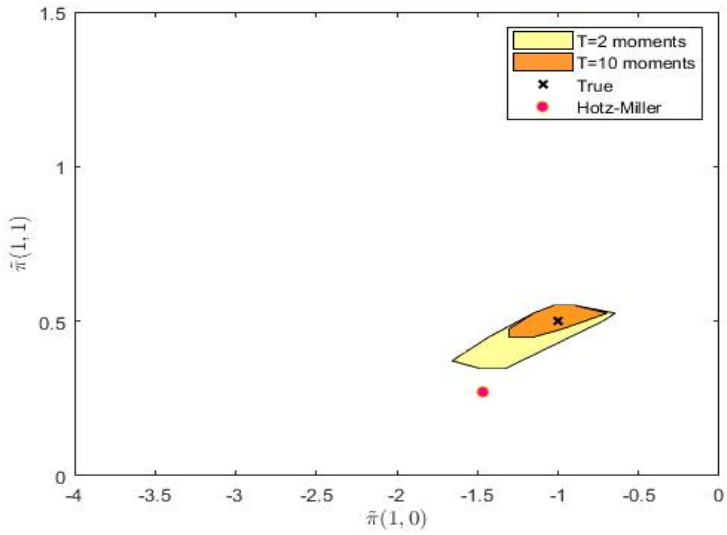
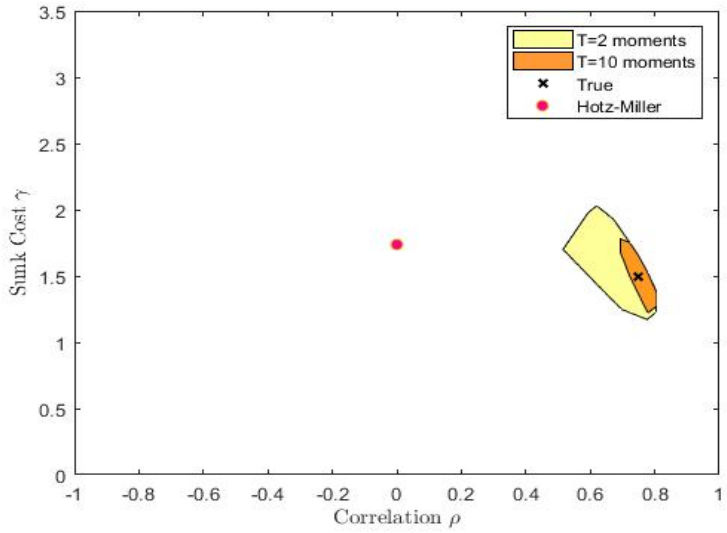


Figure 10: High IV strength



3.3 Marginal GIV Restrictions

In order to shed further light on what is driving (partial) identification in the model, we now go back to the case with two time periods and consider what happens if we only impose the marginal GIV restrictions, i.e. the restrictions implied by the independence between the marginal distributions of u_1 and u_2 and the distribution of the IV. This set of restrictions is weaker than those implied by independence between the joint distribution of (u_1, u_2) and that of the IV. Specifically, since the marginal restrictions do not contain any information on the serial correlation in the unobservables, the identified set for ρ is the entire $(-1, 1)$ interval and we do not plot it. On the other hand, the marginal restrictions do place bounds on the sunk cost, as shown in Figures 11 and 12. The bounds are wide and contain the HM estimates, even when IV strength is high, which confirms that the *joint* GIV constraints play a crucial role in obtaining informative results.

Figure 11: Low IV strength

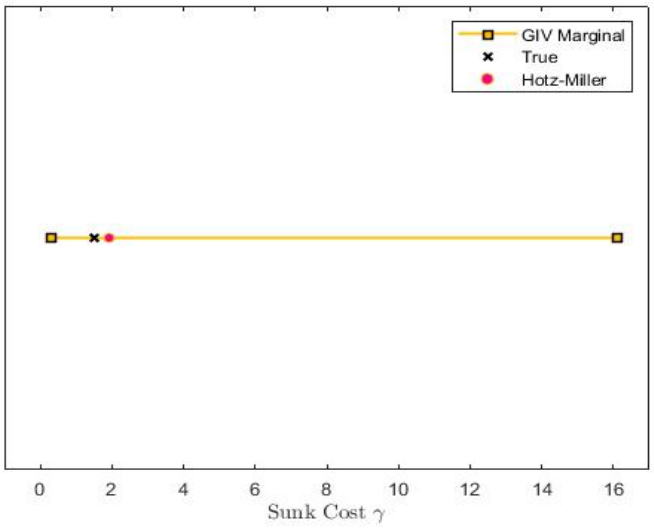
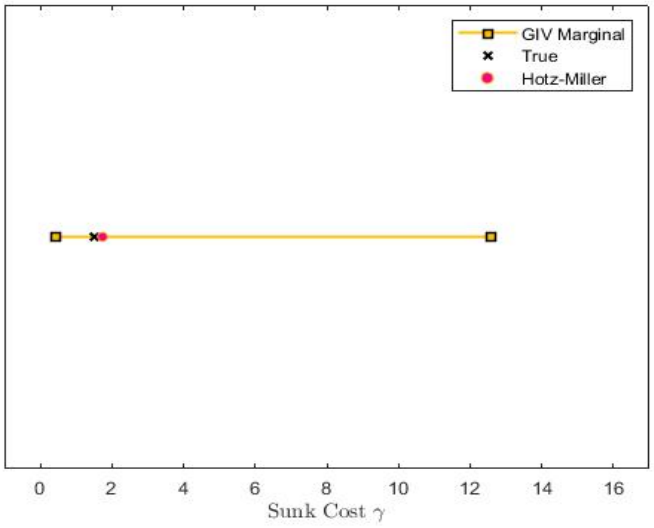


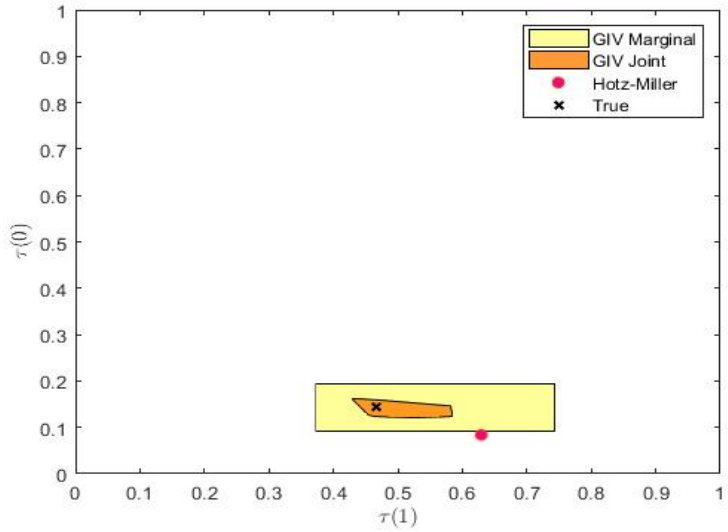
Figure 12: High IV strength



Next, we look at the identified sets for the policy function for the case with high IV strength. Since in this model the policy consists of a pair of thresholds $(\tau(0), \tau(1))$—one for each state—we are able to plot the identified sets in two dimensions. As shown in Figure 13, imposing the joint GIV restrictions substantially shrinks the identified set and again is key

to being able to rule out the HM estimate.

Figure 13: High IV strength



3.4 Adding exogenous covariates

We now explore the impact of adding an exogenous covariate, denoted w, to the model. Specifically, we specify the firm per-period profit as follows:

$$\pi_{it} = \begin{cases} \alpha w_{it} - \beta - \epsilon_{it} & \text{if } a_{it} = 1, x_{it} = 1 \\ \alpha w_{it} - \beta - \gamma - \epsilon_{it} & \text{if } a_{it} = 1, x_{it} = 0 \\ 0 & \text{if } a_{it} = 0 \end{cases}$$

so that w_{it} could be interpreted as a measure of market size such as population, αw_{it} represents variable profits, β is fixed costs, and γ is again the sunk cost of entry. Further, we let ϵ_{it} follow the AR(1) process in (15), with ν_{it} distributed $\mathcal{N}(0, \sigma^2)$. Following Pakes, Ostrovsky, and Berry (2007), we assume that the term $\alpha w_{it} - \beta$ has already been estimated outside the dynamic model and we focus instead on the parameters γ, ρ and σ.[19] When generating the data, we set $\alpha = 1.5$, $\beta = 1$, $\gamma = 1.5, \rho = 0.75$, $\sigma = 1$, and we focus on the case where IV strength is low. Regarding the distribution of the covariate, we let the initial w for each market take the values $\{0.15, 1.00, 1.65\}$ with equal probabilities and evolve according to the

[19]Note that, unlike the previous simulation designs, here we are not allowed to normalize σ since the non-stochastic part of the period-profits is modeled parametrically. Accordingly, for this design, we report the standardized sunk cost $\frac{\gamma}{\sigma}$ to ensure comparability with the plots for the case without w.

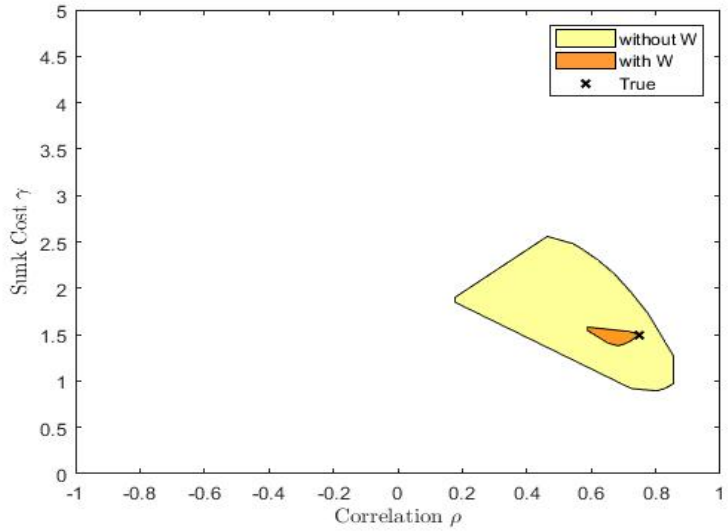
transition matrix

$$\begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}$$

Note that the parameter values are chosen in such a way that the per-period profit function when $w = 1$ is the same as in the model without w.

Because w is exogenous, it can be used as an additional conditioning variable in the GIV inequalities along with the excluded IV. Thus, adding w to the model increases the number of inequalities that each candidate parameter value must satisfy in order to be included in the identified set. We would then expect the identified set to be smaller in the model with the exogenous covariate.[20] As shown in Figure 14, this is indeed the case.

Figure 14: Effect of exogenous covariate



3.5 Counterfactuals

We now turn to counterfactual analysis and consider three scenarios: (i) an increase in the sunk cost of entry by 0.25—corresponding to 17% of its true value—which we call "the sunk cost counterfactual" for brevity; (ii) an increase in the fixed cost by 0.25—corresponding to 25% of its true value—and a simultaneous decrease in the sunk cost by 0.125—or 8% of

[20] However, note that, because w varies over time in the data, the policy function also changes relative to the "no w" case (even for the value of w that makes the per-period profit identical to that in the "no w" model). Thus, it need not be the case that the identified set in this section is a subset of that in the "no w" model.

its true value—which we call "the fixed cost counterfactual"; (iii) a 0.25 subsidy to entry—corresponding to 17% of the true value of the sunk cost—which we call "the subsidy counterfactual." The sunk cost counterfactual is meant to simulate a policy, such as environmental regulation, that only constrains new entrants. On the other hand, the fixed cost counterfactual corresponds to a policy restricting both incumbents and entrants. We simultaneously decrease the sunk cost in this counterfactual to reflect the fact that complying with the regulation may be easier for new entrants (e.g. retrofitting a plant to comply with environmental regulation may be costlier than doing so when starting from scratch). Finally, the subsidy counterfactual mimics a policy that encourages entry of new firms that might be using cleaner or otherwise better technology.

The procedure we employ to assess the impact of these shocks is as follows. For each market in the data, we draw many time series for the unobservables where the initial value is drawn from the conditional distribution associated with the initial number of firms and the initial value of the policy. After a large number of time periods,[21] we hit each market with the given counterfactual shock. For each of the three shocks, we report averages—across observations in the data and draws of the unobservables—of (a) the change in the number of firms relative to when the shock hit; (b) the number of "old" firms exiting (i.e. firms that were inside the market when the shock hit and exited after the shock); and (c) the number of "new" firms entering (i.e. firms that entered the market after the shock hit). We measure each of these quantities 5 and 10 years after the shock.

Tables 4 to 6 show the results for the simple model with $T = 2$ and no exogenous covariate w. Since in this case we are able to characterize the sharp identified set for the structural parameters, the bounds on the counterfactual outcomes presented here are also sharp.[22] As expected, the bounds become tighter when we move from the low IV strength to the high IV strength case. In contrast, the HM approach leads to consistently biased results. In particular, HM tends to overestimate the impact of each of the three shocks in absolute value, with the magnitude of the bias being especially large for the fixed cost and the subsidy counterfactuals. Intuitively, because HM assumes away serial correlation in the unobservables, the shocks drawn for the counterfactuals will exhibit much less persistence relative to their true distribution. Therefore, HM will predict an excessive amount of entry and exit in the counterfactual scenarios.

[21] We set this number to 250. Drawing many time periods prior to the sample periods is needed to make sure that the distribution of the unobservables converges to its stationary distribution.

[22] Based on the insight from Section 3.3, here we impose the joint GIV restrictions as opposed to the marginal ones.

Table 4: Increase in sunk cost: identified sets

		after 5 years			after 10 years		
		True	GIV	HM	True	GIV	HM
Low IV strength	Change in # firms	0.01	(0.00,0.03)	0.03	0.01	(-0.00,0.03)	0.04
	Old firms exiting	0.18	(0.14,0.28)	0.19	0.32	(0.26,0.45)	0.34
	New firms entering	0.19	(0.15,0.30)	0.23	0.33	(0.30,0.47)	0.37
High IV strength	Change in # firms	0.01	(0.00,0.02)	0.04	0.01	(0.00,0.02)	0.04
	Old firms exiting	0.18	(0.17,0.24)	0.21	0.32	(0.30,0.41)	0.37
	New firms entering	0.19	(0.17,0.26)	0.25	0.33	(0.30,0.42)	0.40

Table 5: Increase in sunk cost and decrease in fixed cost: identified sets

		after 5 years			after 10 years		
		True	GIV	HM	True	GIV	HM
Low IV strength	Change in # firms	-0.12	(-0.29,-0.10)	-0.37	-0.13	(-0.32,-0.10)	-0.41
	Old firms exiting	0.31	(0.25,0.53)	0.59	0.48	(0.39,0.67)	0.75
	New firms entering	0.19	(0.14,0.29)	0.21	0.35	(0.28, 0.48)	0.34
High IV strength	Change in # firms	-0.12	(-0.17,-0.11)	-0.32	-0.13	(-0.18,-0.11)	-0.35
	Old firms exiting	0.31	(0.29,0.42)	0.58	0.48	(0.44,0.59)	0.75
	New firms entering	0.19	(0.17,0.25)	0.26	0.35	(0.32, 0.42)	0.41

Table 6: Subsidy to entry: identified sets

		after 5 years			after 10 years		
		True	GIV	HM	True	GIV	HM
Low IV strength	Change in # firms	-0.01	(-0.03,-0.00)	-0.04	-0.02	(-0.03,-0.00)	-0.04
	Old firms exiting	0.24	(0.17,0.36)	0.34	0.40	(0.30,0.54)	0.53
	New firms entering	0.23	(0.17,0.35)	0.30	0.38	(0.29, 0.52)	0.49
High IV strength	Change in # firms	-0.01	(-0.02,-0.00)	-0.04	-0.01	(-0.02,-0.00)	-0.04
	Old firms exiting	0.22	(0.20,0.30)	0.37	0.37	(0.34,0.48)	0.57
	New firms entering	0.21	(0.20,0.29)	0.33	0.36	(0.34, 0.47)	0.53

4 Empirical Application

In order to illustrate the approach, we apply it to the ready-mix concrete industry studied by Collard-Wexler (2014) (henceforth, CW). CW quantifies the magnitude of the sunk cost of entry in each of many isolated markets in the US and uses these estimates to assess how persistent the effects of a horizontal merger are in this industry. More specifically, CW first estimates the firms' policy functions based on data on the number of ready-mix concrete plants and demand shifters. Given the policies, CW then simulates the evolution of a market following a merger to monopoly and evaluates how long it takes for a second firm to enter.

We use the same data and modeling framework as CW, but estimate all of the structural parameters as opposed to just the policy functions.[23] This allows us to address a broader range of counterfactual questions. In particular, we consider the effects of a policy, similar to the environmental policies in the cement industry study of Ryan (2012), that alter sunk and/or fixed costs. Note that these are structural parameters of profit function that are not directly revealed by policy functions. In addition, since the GIV approach accommodates incomplete models, we are able to tackle the initial conditions problem in a more flexible way.[24]

Note that our application is intentionally simplified to serve as an example within a longer methodological paper.

Table 7 summarizes the variables we use. The data on number of plants and construction employment is the same as in CW and we refer the reader to that paper for more details. Briefly, the number of plants variable measures how many firms are active in each isolated town, while construction employment shifts ready-mix concrete demand. We follow CW and treat construction employment as exogenous, while the number of concrete plants is endogenous. In addition, we obtain data on past household income growth at the county

[23]Collard-Wexler (2013) estimates a full model of industry dynamics using firm-level data under the assumption that the dynamic states are econometrically exogenous. We use coarser market-level data, but address the endogeneity of market structure.

[24]CW addresses the initial conditions problem by assuming that the industry has been following the same set of policies for a long time and then using this assumption to simulate the probabilities of the initial states via a modification of the GHK algorithm.

level from the US Census website. We assume that past income growth is excluded from the current profits of concrete firms and that, conditional on current construction employment, past income growth is independent of within sample unobserved shocks to profitability. Past income growth therefore serves as an excluded instrument in our model.

Table 8 shows results for a "quasi first-stage regression". This table presents an ordered probit model with the number of firms as the dependent variable. We see that that the coefficients on the both exogenous variables are positive and precisely estimated. The result suggests that our excluded instrument is "relevant," even when conditioning on current demand. Of course, the true reduced form of the model is not an ordered probit and we present this merely as a descriptive result.

We conclude our brief descriptive analysis by reporting in Table 9 the transition probabilities for the number of plants, which shows that the outcome exhibits substantial persistence over time conditional on both values of the instrument.[25]

Table 7: Summary statistics. Fully balanced panel of 428 markets between 1994 and 2005

Variable	Mean	St.Dev.	Min	Max
Number of plants	0.97	0.93	0.00	6.00
Construction Employment	519	819	3	17,772
Household Income Growth 1969-1989	0.15	0.11	-0.16	0.69

[25]The results in Tables 8 and 9, as well as those from structural estimation, are based on a discretized version of the original data. Specifically, since more than 90% of the original observations at the market-year level have two or fewer plants, we censor the number of plants at two, which reduces the number of parameters to estimate in the structural model. Similarly, we discretize construction employment and household income growth. For each variable, we define a high and a low value depending on whether a given observation is above or below the median. Because construction employment essentially has no over-time variation after discretization, we take its value in the first year of the sample for each market and assume it is constant over time (and that firms know it). This means that, when we condition on construction employment to compute the moments implied by the model, the conditioning variable is scalar—as opposed to having dimension equal to the number of time periods—which leads to much more precisely estimated moments.

Table 8: Ordered probit results. Dependent variable is number of plants. ** denotes significance at the 95% level.

Log Construction Employment	0.14**
Income Growth 1969-1989	0.22**
Likelihood-Ratio Test p-value	0.00

Table 9: Transition probabilities conditional on instrument

	Low IV			High IV		
	Policy			Policy		
State	0	1	2	0	1	2
0	0.92	0.07	0.01	0.90	0.09	0.01
1	0.07	0.87	0.06	0.06	0.89	0.05
2	0.02	0.18	0.80	0.00	0.18	0.82

We now turn to structural estimation. As in CW, we estimate a version of the Last-In First-Out model developed by Abbring and Campbell (2010). Again, we refer the reader to CW for more details on what assumptions are imposed and why this is may be a suitable oligopoly model in the context of the ready-mix concrete industry. Importantly, the Abbring-Campbell assumptions motivate a model where we avoid the issues of multiple equilibria, which are discussed elsewhere in the paper. This allows for a relatively simple illustrative application.

We specify the flow profit function as follows:

$$\pi_{it} = \begin{cases} \alpha_{x_{it}} w_i - \beta + \epsilon_{it} & \text{if was in at } t-1, \text{ stays in at } t \\ \alpha_{x_{it}} w_i - \beta - \gamma + \epsilon_{it} & \text{if was out at } t-1, \text{ enters at } t \\ 0 & \text{if is out at } t \end{cases} \tag{25}$$

where w_i denotes construction employment (in thousands) in market i, $\alpha_{x_{it}}$ is a coefficient that depends on the number x_{it} of active firms in market i at time t, β represents fixed costs, γ is the sunk cost of entry, and ϵ_{it} denotes a potentially serially correlated unobservable shock to profitability. We assume that, conditional on ϵ_{it-1}, ϵ_{it} is equal to ϵ_{it-1} with probability ρ and is drawn uniformly from the $[-1, 1]$ interval with probability $1-\rho$.[26] Further, we impose the natural restriction $\alpha_2 \leq \frac{\alpha_1}{2}$, i.e. that per-firm variable profits (weakly) decrease with the number of competitors.

Because we have twelve years of data, the number of moment conditions defining the sharp identified set is extremely large and it is not practical to use all of them. However, the goal

[26]This parametric specification for the joint distribution of the unobservables is used by Abbring and Campbell (2010), who show that it satisfies their Assumption 3. This, along with other mild assumptions, ensures existence and uniqueness of a Markov-perfect equilibrium in Last-In First Out strategies.

of obtaining a small identified set suggests that we want to use a large number of them. This creates a potential problem, as many methods for obtaining confidence regions will not perform well when the number of moment inequalities is very large. This creates a tension between the desire for a small identified set (on one hand) and the desire for precise inference (on the other hand. To resolve this tension, we report confidence sets for the structural parameters via one of the approaches proposed by Chernozhukov, Chetverikov, and Kato (forthcoming) (henceforth, CCK). The method is intended for cases where the number of moment inequalities is very large, even larger than the dimension of the data. Speaking roughly, it provides an econometrically disciplined way of determining the subset of moment conditions that are most informative about the parameter values.

As in the Monte Carlo simulations from Section 3, we impose two sets of GIV moments: (i) the full list of moments characterizing the sharp identified set for the case with two time periods;[27] and (ii) several inequalities corresponding to observable events over the entire twelve years in the data.[28]. In addition, we include the complement of each of these events and we make sure that the left hand side of each GIV inequality incorporates not only the observable event associated with that inequality, but also all observable events whose necessary conditions (in the space of unobservables) are subsets of the set whose probability is on the rhs of the inequality. We follow CCK in generating a large number of draws to approximate the integrals corresponding to these events and ignoring the corresponding simulation error. Simple diagnostics suggest that the simulation variance is indeed negligible relative to the sample variance.

Since all our conditioning variables are discrete, we can easily turn the conditional moment inequalities into unconditional ones. We are then left with 614 inequalities. We choose to rely on the results in CCK because they allow for the number of moment inequalities to grow with the sample size. This seems to be relevant for our purposes given that we have more moments than observations in the data. For each candidate parameter value, we use as test statistic the maximum of the sample moments (at that parameter value) divided by their standard deviations and scaled by the square root of the sample size. In order to obtain valid critical values, we employ the one–step multiplier bootstrap method in CCK with 250

[27]We compute these moments by only using the first two years in the panel. In principle, one could obtain more moments by taking all subsets of the data with two consecutive years. Because we already obtain informative results when only using the first two years, we do not pursue this extension in the paper.

[28]These events are: "some entry and some exit occur," "there are always zero firms," "there is always one firm," "there are always two firms," "there is at least one period with zero firms," "there is at least one period with one firm," "there is at least one period with two firms," "the number of firms goes from zero to one," "the number of firms goes from one to two," "the number of firms goes from zero to two," "the number of firms goes from two to zero," "the number of firms changes at least once," "the number of firms changes exactly once," "the number of firms changes exactly twice," "the number of firms goes from one to zero," "the number of firms goes from two to one or zero," "the number of firms in the market is unchanged for at least five consecutive periods," "the number of firms in the market is unchanged for at least nine consecutive periods."

bootstrap draws.[29]

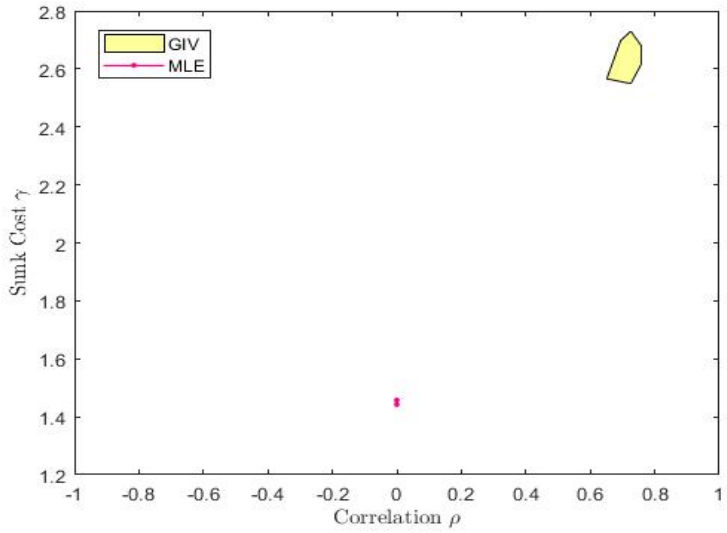
Table 10 displays confidence intervals for the structural parameters individually, while Figure 15 shows the projection of the confidence region onto the space of correlation parameter and sunk cost. Note that the correlation parameter is estimated to be positive and significantly different from zero. For comparison, we also consider a model that maintains the same specification as above, but assumes that the unobservables are i.i.d. over time. We estimate this model by maximum-likelihood and use the bootstrap to obtain standard errors. We can see that the MLE model with no serial correlation tends to under-estimate the fixed cost, the sunk cost, as well as the negative impact of competition on variable profits.

Table 10: 95% Confidence Intervals for the Structural Parameters

	after 5 years	
	GIV	MLE
α_1	(0.150,0.225)	(0.042,0.045)
α_2	(0.000,0.013)	(0.015,0.020)
β	(0.217,0.350)	(0.038,0.043)
γ	(2.549,2.730)	(1.442, 1.457)
ρ	(0.651,0.758)	—

[29]CCK propose several other testing procedures, including two– and three–step methods that have better power properties relative to one–step approaches. On the other hand, one–step methods have the advantage of not requiring one to choose any tuning parameters. Since the one–step multiplier bootstrap already delivers informative results in our data, that is our inferential procedure of choice.

Figure 15: Autocorrelation and sunk cost: Projection of 95% confidence sets



In order to investigate whether allowing for endogeneity of market structure makes a difference for policy-relevant questions, we turn to counterfactual analysis. We implement the same three counterfactuals as in the simulations from Section 3.5.[30] An increase in sunk and/or fixed costs can be thought of as a possible counterfactual environmental regulation, as in an investment of technology to reduce polluted water from running off concrete operations. If the proposed policy is not observed in the data, then the full structural model can still be used to predict the effect of the policy.

As Tables 11 to 13 show, for each of the counterfactual changes, the estimated impact on both the number and the composition of firms in the market varies substantially depending on whether we account for endogeneity of market structure. In particular, when the unobservables are assumed to be i.i.d., the model predicts much more counterfactual entry and exit, relative to the more flexible approach that allows for serial correlation. Intuitively, the i.i.d. assumption forces the unobservables to vary too much from one period to the next, which translates into excessive variation in the implied market outcomes relative to the model with serial correlation.

These policy counterfactual results show the large bias that may result from models that artificially set the serial correlation parameter to zero, as in common in much of the literature. Further, the relatively precise estimates from our procedure show that dropping restrictive initial conditions assumptions need not result in imprecise estimates. In part, this stems for our use of new econometric techniques for moment inequalities that allow us to employ very

[30] We choose the same values for the changes in sunk/fixed cost as in the Montecarlo simulations. If we take the midpoints of the GIV confidence intervals for the parameters as a reference, the sunk cost counterfactual increases the sunk cost by approximately 9% of its estimated level, the fixed cost counterfactual increases the fixed cost by approximately 88% of its estimated level, and the subsidy counterfactual decreases the sunk cost by around 9% of its estimated level.

large number of moment inequalities in an econometrically disciplined and correct way.

Table 11: Increase in sunk cost (95% confidence intervals)

	after 5 years	
	GIV	MLE
Change in total # of firms	(-0.09,-0.03)	(-0.18,-0.15)
Change in fraction of new firms	(-0.05,-0.01)	(-0.15,-0.14)

Table 12: Increase in fixed cost and decrease in sunk cost (95% confidence intervals)

	after 5 years	
	GIV	MLE
Change in total # of firms	(-0.08,-0.03)	(-0.11,-0.08)
Change in fraction of new firms	(0.01,0.06)	(0.07,0.10)

Table 13: Subsidy to entry (95% confidence intervals)

	after 5 years	
	GIV	MLE
Change in total # of firms	(0.05,0.10)	(0.34,0.38)
Change in fraction of new firms	(0.04,0.10)	(0.12,0.14)

5 Conclusion

In this paper, we have proposed an approach to identification and estimation of dynamic models with serially correlated unobservables. We tackle the resulting endogeneity of dynamic states by relying on the type of instrumental variables intuition that is commonly used in static models. In order to characterize the identified sets for quantities of interest and obtain confidence regions, we leverage recent results in the econometrics literature on partially identified models and the associated inference literature.

This paper opens several avenues for future research. First, it would be interesting to apply the proposed approach to a wider class of empirical settings and see how accounting for the

endogeneity of market structure affects the results. Second, while in the simple models we considered here we could obtain the identified sets and their confidence regions by grid search over the structural parameter space, this strategy becomes computationally prohibitive in more complicated models. In these cases, it might be possible to develop computational methods that avoid searching over the entire parameter space and instead focus directly on the final object of inference—often a scalar counterfactual quantity.

Appendix A: More General Forms of Unobserved Heterogeneity

A.1 Generalization to Discrete Types and Beyond

In this section, we consider an example of how to extend our method to a single-agent problem with persistent unobserved discrete types. As a particularly simple example, we again consider the monopoly entry problem but now consider a case of discrete firm heterogeneity, in the spirit of Heckman and Singer (1984), Keane and Wolpin (1997), Arcidiacono and Miller (2011) and related literature. In this case, the unobservable that enters the single-period profit function is sometimes modeled as

$$\epsilon_{it} = \lambda_{it} + \eta_{it}, \tag{26}$$

where η_{it} is continuously distributed and independent over t and i while λ_{it} is persistent and takes on a discrete set of values:

$$\lambda_{it} \in \{\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_K\}.$$

For example, the single-period return to operating in market t in a single-agent dynamic entry model could be specified as

$$\bar{\pi}\left(a_{it}, x_{it}, w_{it}, \theta_\pi\right) - (\lambda_{it} + \eta_{it}),$$

so that $(\lambda_{it} + \eta_{it})$ represents the fixed cost of operation, including a persistent and a transitory component.

A particularly simple example would be a pure "discrete firm-effects" model where λ_{it} does not vary over time, so that $\lambda_{it} \equiv \lambda_i$. More generally, λ_{it} might evolve across values according to a parameterized distribution

$$M(\lambda_{it}|\lambda_{it-1}; \theta_u). \tag{27}$$

We can think of this transition as being generated by a function

$$\lambda_{it} = h(\lambda_{it-1}, \nu_{it}, \theta_u) \tag{28}$$

where ν_{it} is a scalar or vector that is independent over time and markets, with distribution $\Phi(\nu_{it}; \theta_u)$.

This suggests a generalization to a broader set of models where we treat λ_{it} as the "persistent component" of the unobservables, ν_{it} as the innovation in that component and η_{it} as additional transient (independent over time) shocks to single-period profits. We then define u_{it} as the (possibly vector-valued) unobservable that enters the policy function. Note that in the discrete heterogeneity case $u_{it} = (\lambda_{it}, \eta_{it})$. As in (26), it might sometimes be helpful to distinguish ϵ_{it} as the particular combination of period t unobservables that affect single-period returns. However, as a more general case we can just let single-period profits depend on u_{it}.

The serial correlation case in the body of the text is particularly simple because the scalar u_{it} is simultaneously the persistent component (λ_{it}) that effects future expectations and is also the component (ϵ_{it}) that effects single period returns. That is, in the single-agent AR(1) serial correlation case, $u_{it} \equiv \lambda_{it}$ and $u_{it} \equiv \epsilon_{it}$. In the discrete heterogeneity case outlined above, those simple relationships do not hold. The term that enters single period profits, $\epsilon_{it} = \lambda_{it} + \eta_{it}$, is different from the term, λ_{it}, that conditions future expectations.

Having defined the firm-policy relevant vector of unobservables u_{it} in this way, the general discussion in the main body of identification via GIV in Section 2.3 goes through in its entirety. In the next subsection we further illustrate the details of the discrete heterogeneity case and tie our GIV treatment to the treatment of discrete heterogeneity in Kasahara and Shimotsu (2009) and related literature.

A.2 Details on the Discrete Heterogeneity Case

In the generalized single-agent case of the last section, the single period return is still $\pi(a_{it}, x_{it}, w_{it}, u_{it}, \theta_\pi)$, but now

$$u_{it} = (\lambda_{it}, \eta_{it}). \tag{29}$$

The Bellman equation is

$$V(x_{it}, w_{it}, u_{it}) = \max_{a_{it} \in \mathcal{A}(x_{it})} \left(\pi(a_{it}, x_{it}, w_{it}, u_{it}; \theta_\pi) + \delta E_{\theta_u} \left[V(x_{it+1}, w_{it+1}, u_{it+1}) \,|a_{it}, x_{it}, w_{it}, \lambda_{it} \right] \right). \tag{30}$$

This takes the same form as (7), except that the future expectation is conditioned on λ_{it}, rather than u_{it} and we again recall the new definition of u_{it} in (29).

The associated policy function is still $a_{it} = \sigma_{it}(x_{it}, w_{it}, u_{it})$. Consider the specialization to the case of Section 2.4 with binary actions and states, and with no role for w_{it}. That is, a_{it} and x_{it} both take on the values 0 or 1. Further, assume only two possible values for λ, fixed over time. If the policy function is monotonic in the independent shock η, then under appropriate conditions the dynamic firm policy function is described by four cutoff rules in η, one for each combination of x_{it} and λ_{it}. This is twice as many cutoff rules as existed in the serial correlation example, where there was a cutoff in u_{it} for each value of x_{it}.

To be more formal, consider an AR(1) example following on the simplest example in the main text, where the inverse image set of unobservables in market i is

$$\mathcal{U}(a_i, x_i, w_i, \sigma) = \{u_i \colon (-1 + 2a_{it}) u_{it} < (-1 + 2a_{it}) \tau(x_{it}, w_{it}), \forall t\}. \tag{31}$$

Note that the term $(-1 + 2a_{it})$ is just a device to switch the sign of the inequality depending on whether or not $a_{it} = 1$. In the discrete heterogeneity case, the corresponding inverse image set is

$$\mathcal{U}(a_i, x_i, w_i, \sigma) = \{(\eta_i, \lambda_i) \colon (-1 + 2a_{it}) \eta_{it} < (-1 + 2a_{it}) \tau(x_{it}, w_{it}, \lambda_{it}), \forall t\}. \tag{32}$$

It is clear that (31) involves fewer unknown policy function parameters τ than (32). This is one clear advantage of the AR(1) case.

However, Kasahara and Shimotsu (2009) exploit a clear advantage of the discrete heterogeneity case in handling initial conditions. If we are willing to assume that persistent heterogeneity takes only a small number of discrete values, then the initial conditions are similarly described by a small set of additional unknown parameters. In particular, the initial conditions of the model are completely characterized by the unknown terms

$$\Pr(x_{i1} | \lambda_{i1} = \bar{\lambda}_k, w_{i1}),$$

which is a relatively small number of unknowns when λ_{i1} takes on a small number of values. Further, the number of unknown initial condition parameters does not increase in T. This simple specification of the initial conditions then allows Kasahara and Shimotsu (2009) to complete the dynamic model, which in turn allows them to specify a set of strict equalities that are necessary to rationalize the observed data. In the fully discrete case, Kasahara and Shimotsu (2009) then discuss (point) identification of transition probabilities conditional on each possible value of λ. They provide sufficient conditions involving combinations of [a] a minumum number of time periods, [b] sufficient variation in observed (unconditional on λ) multiperiod transition probabilities in the data, [c] limits on the number of discrete points of support for λ and [d] limits on the variation in λ over time. These results have the further advantage of possibly identifying joint transitions for (x_{it}, w_{it}) that depend on λ, which is more general than the base specification in the main body of our text.

In the single-agent case, then, the chief advantage of the discrete heterogeneity model lies in a naturally low-dimensional parameterization of the initial conditions, following directly from the assumption that the persistent heterogeneity is, in a formal sense, of low dimension relative to the overall variation in the data over time. It is possible that the advantages of the discrete heterogeneity assumption are highest when the time dimension is of intermediate length. When the time dimension is too short, the large number of combined parameters of the policy functions and the initial conditions will overwhelm the data even given the advantages of equality restrictions. At long time periods, GIV inequalities applied to the AR(1) serial correlation case may do reasonably well since the issue of incomplete initial conditions may, in any case, be small in long panels, as highlighted by Honoré and Tamer (2006).

A.3 Oligopoly

Here we extend the discussion of the last subsection to oligopoly. As an example, again consider a simple "entry" model where the return to firm j of operating in market i at time t is

$$\bar{\pi}_j\left(x_{it}, w_{it}, \theta_\pi\right) - (\lambda_{ijt} + \eta_{ijt}).$$

A convenient set of assumptions is that λ_{ijt} is discrete, persistent and commonly known by all firms at time t while η_{ijt} is continuously distributed, i.i.d. over firms and time and is also private information to firm j. Note that the private continuously distributed η_{ijt} can help to ensure existence of equilibrium in an otherwise discrete dynamic game, as in Doraszelski and Satterthwaite (2010). The assumption that λ is commonly observed prevents issues of informational updating and strategic information revelation with respect to persistent firm types. This often then preserves the first-order Markov structure of the oligopoly strategies.

Given the model of the last paragraph, the unobservables that enter firm j's single-period profit function are then

$$u_{ijt} = (\lambda_{ijt}, \lambda_{i,-j,t}, \eta_{ijt}). \tag{33}$$

Assuming (as usual in the empirical literature) that firms all play the same strategy whenever they reach the same state (even when the underlying game admits multiple equilibria), the equilibrium policy functions once again take the form,

$$a_{ijt} = \sigma_j(x_{it}, w_{it}, u_{ijt}),$$

with u_{ijt} now defined by (33). Once again, in a discrete game the results by Kasahara and Shimotsu (2009) may or may not provide point identification of cutoff decision rules in the quantiles of η_{ijt} (as well as transition matrices for x_{it} and w_{it} that potentially depend on λ_{it}.) These transitions and cutoffs will in some commonly used cases point identify the structural parameters and in other cases they may not. Either way, the GIV framework generalizes the result.

In particular, the inverse image set in the simple oligopoly entry example is very close to the single agent version in (32)

$$\mathcal{U}(a_i, x_i, w_i, \sigma) = \{(\eta_{ijt}, \lambda_{it})\colon (-1 + 2a_{ijt})\eta_{ijt} < (-1 + 2a_{ijt})\tau(x_{it}, w_{it}, \lambda_{it}), \forall (j,t)\}. \tag{34}$$

Note that the cutoffs $\tau(x_{it}, w_{it}, \lambda_{it})$ take on a finite number of values as long as x_{it}, w_{it}, and λ_{it} all take on a finite number of values. One could allow for continuous λ, but then the function τ might easily take on a continuum of values.

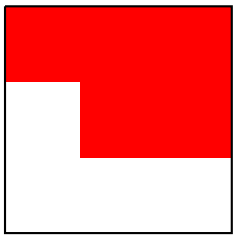
GIV inequality constraints associated with (34) can once again be transformed into equalities by introducing additional parameters for initial conditions. However, the oligopoly case naturally multiplies the number of these initial condition parameters. If each firm has a potentially different λ and if the number of potential values for λ is K, then we have K^J initial condition parameters. It is not immediately clear whether it is easier to include these parameters in the identification and/or estimation problem and take advantage of the resulting equality restriction, or whether it is better to use the GIV inequalities without the initial condition parameters. The answer is likely to be problem specific.

Another alternative to the discrete persistent heterogeneity model just described is a model that extends the AR(1) case from the main text to the multiple agents settings. In such a model, each agent has one continuously distributed persistent unobservable that is common knowledge. Continuity of the unobservable implies that we can establish existence of an equilibrium without the need for additional private information i.i.d. shocks. However, in general, the image sets $\mathcal{U}(a_i, x_i, w_i, \sigma)$ needed as an input to characterize the CR identified set for the policy functions will not be determined by a finite number of thresholds. Instead, they will typically be defined by curves in the space of unobservables, which makes implementation hard or outright infeasible. Therefore, we view the model described in this section as preferable for our purposes and we estimate a variant of it in Appendix C.

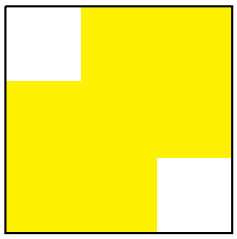
One could also consider examples of discrete/continuous games of the sort envisioned by Bajari, Benkard, and Levin (2007)'s discussion of possible set identification in a world without persistent heterogeneity. Perhaps some combination of discrete and continuous persistent λ variables will be appropriate in that case.

Appendix B: Non-Elemental Sets for Single-Agent Example

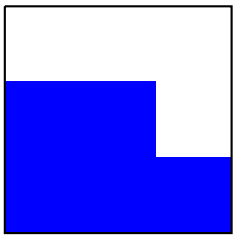
In this appendix, we present the non-elemental sets belonging to the collection of core determining sets for the example considered in Section 2.4. These are the sets given by the overlapping unions of elemental sets, excluding cases of strict subsets. Below each set we specify the events associated with the elemental sets that we are taking the union of (top three sets) or the non-elemental sets that we taking the union of (bottom two sets).



(a) (0,0,0)+(1,1,0)



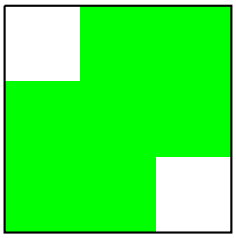
(b) (0,0,0)+(1,1,1)



(c) (1,1,1)+(0,0,1)



(d) (a)+(b)



(e) (b)+(c)

Appendix C: Computed Example of Dynamic Oligopoly

In this section, we illustrate how our approach applies to the dynamic oligopoly setting by considering a model with discrete persistent common knowledge unobservables and private information i.i.d. shocks, along the lines of the example in Appendix A.3.

Consider the problem faced by two firms choosing how many stores to open in each of several markets over time. We assume that the firms choose between having one or two stores open in each market at any given point in time. Given the number of stores in a market, the two firms engage in Bertrand competition and each of them charges the same price across its own stores. We let j index stores and $k(j)$ be the firm owning store j. Consumers view the products as horizontally differentiated across stores. Specifically, let

$$u_{ijt} = \overline{\delta} - \alpha p_{k(j)t} + \epsilon_{ijt}$$

be the utility that consumer i gets from buying from store j at time t. Assuming that ϵ_{ijt} is i.i.d. extreme value across consumers and stores, the market share of firm k at time t is given by

$$s_{kt}(p_t) = \frac{x_{kt} \exp^{\overline{\delta} - \alpha p_{kt}}}{1 + \sum_{r=1}^{2} x_{rt} \exp^{\overline{\delta} - \alpha p_{rt}}},$$

where $p_t = (p_{1t}, p_{2t})$ and x_{kt} denotes the number of stores that firm k has open at time t. The first order condition for firm k's static problem at time t is then

$$p_{kt} = mc_{kt} - \frac{s_{kt}(p_t)}{\frac{\partial}{\partial p_{kt}} s_{kt}(p)}$$

where mc_{kt} denotes the marginal cost of firm k at time t, which is assumed to be constant across firm k's stores. The system of first-order conditions implicitly determines the equilibrium prices in each market at any point in time and this in turn determines each firm's variable profits. Let $\pi_k^*(x_t, w_t)$ be firm k's variable profit at time t as a function of the endogenous states $x_t = (x_{1t}, x_{2t})$ and of market size w_t, which we assume to be exogenous. Further, each firm incurs a fixed cost for each open store at time t and a sunk cost if it decides to open a new store in the next period. The timing is such that at time t a firm chooses a_{kt}, i.e. the number of stores at time $t+1$, and incurs the associated sunk cost, if any, at time t. The final specification for flow profits is as follows:

$$\pi_{kt}(a_{kt}, x_t, w_t, \lambda_{kt}, \nu_{kt}) = \pi_k^*(x_t, w_t) - \beta x_{kt} \lambda_{kt} - \gamma(a_{kt} - x_{kt}) \mathbb{I}\{a_{kt} \geq x_{kt}\} \cdot \nu_{kt},$$

where β is the fixed cost, γ is the sunk cost and λ_{kt} and ν_{kt} are shocks that are unobserved to the econometrician. We assume that λ_{kt} is common knowledge and possibly correlated over time and across players. This is the structural shock that may be correlated with the state x_t, thus leading to an endogeneity problem. On the other hand, ν_{kt} is assumed to be i.i.d. over time and across players and to be observed by player k but not by the other firm. Adding this shock to the profit specification helps show existence of the dynamic equilibrium and compute said equilibrium in practice.

In the simulation, we set $\alpha = 0.5, \overline{\delta} = 2$ on the demand side and $mc_{kt} = 0.5$, $\beta = 0.7, \gamma = 1$ on the cost side. Moreover, we assume λ_{kt} can only take the two values 0.25 and 0.75, so that at any time t there are four possible combinations of $\lambda_t = (\lambda_{1t}, \lambda_{2t})$. The joint probability distribution for $(\lambda_t, \lambda_{t+1})$ is given by the following matrix

$$\begin{bmatrix} 0.175 & 0.025 & 0.025 & 0.025 \\ 0.025 & 0.175 & 0.025 & 0.025 \\ 0.025 & 0.025 & 0.175 & 0.025 \\ 0.025 & 0.025 & 0.025 & 0.175 \end{bmatrix}$$

In other words, the λ_t process is fairly persistent over time. Further note that the matrix above is characterized by just one parameter, which is convenient computationally. In particular, this correlation structure implies that, given any value of λ_t, λ_{t+1} will be equal to λ_t with probability $\rho = 0.7$ and will switch to any of the other three values with 0.1 probability each. Moreover, the shock ν_{kt} takes one of the two values 0.5 and 1 with equal probability and is i.i.d. over time and across players. The exogenous state w is either 0.8 or 2 and, for simplicity, is taken to be constant over the time span from $t = 1$ to $t = T$, where T is the number of observed time periods. In the simulation, we let $T = 2$.[31]

Finally, we use past values of w as instruments. For each firm, the instrument has a correlation of 0.81 and 0.56 with x from period 1 and 2, respectively.

The structural parameters are the fixed cost β, the sunk cost γ and the correlation parameter ρ for the λ_t process. To characterize the identified for $\theta = (\beta, \gamma, \rho)$, we proceed in two steps. First, we find all the policies that satisfy the CR conditions. Note that, in this simple example, there are only 64 possible states

[31] In order to focus on identification of the parameters and abstract from inference issues, we draw a very large sample of markets (50,000).

and therefore a policy is a vector of length 64 with elements taking two possible values (corresponding to choosing one or two stores). This makes it possible to enumerate all possible (monotonic) policies and check the GIV conditions for each of them. Only 5,335 policies survive at the end. This is the sharp GIV identified set for the policy.

In the second step, we go from the identified set for the policy to the identified set for the structural parameters. Specifically, for each candidate value θ in a grid, we check whether there exists a policy in the GIV identified set such that when the opponent plays according to that policy, it is optimal for a firm to play the same policy. Here we exploit the fact that the two firms face the same problem and thus there has to be a symmetric equilibrium. Further, note that this procedure does not rely on uniqueness of the equilibrium and thus can be applied without any changes even if a given value of θ is associated with multiple equilibria.

Figures 17 and 18 show projections of the identified set for the structural parameters.

Figure 17: Projection of identified set for β, γ

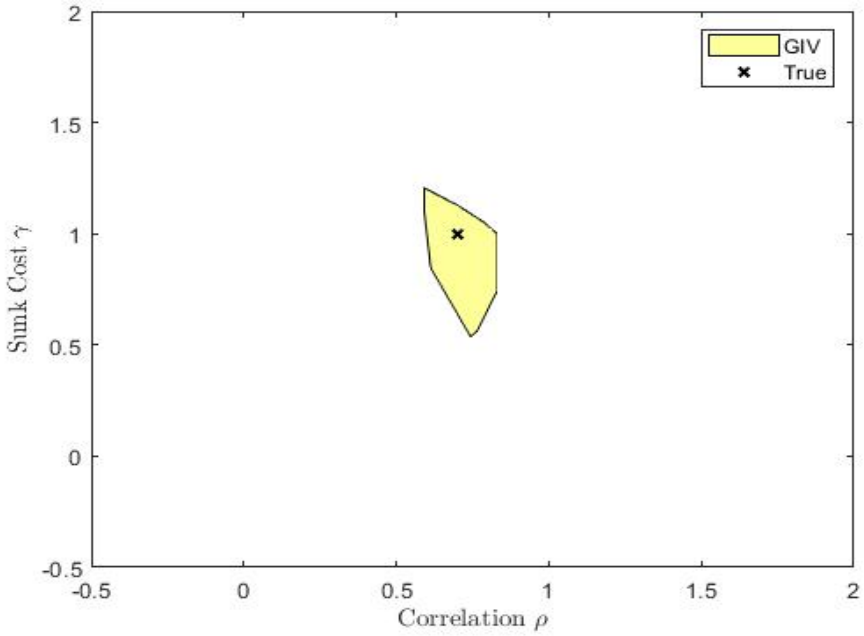
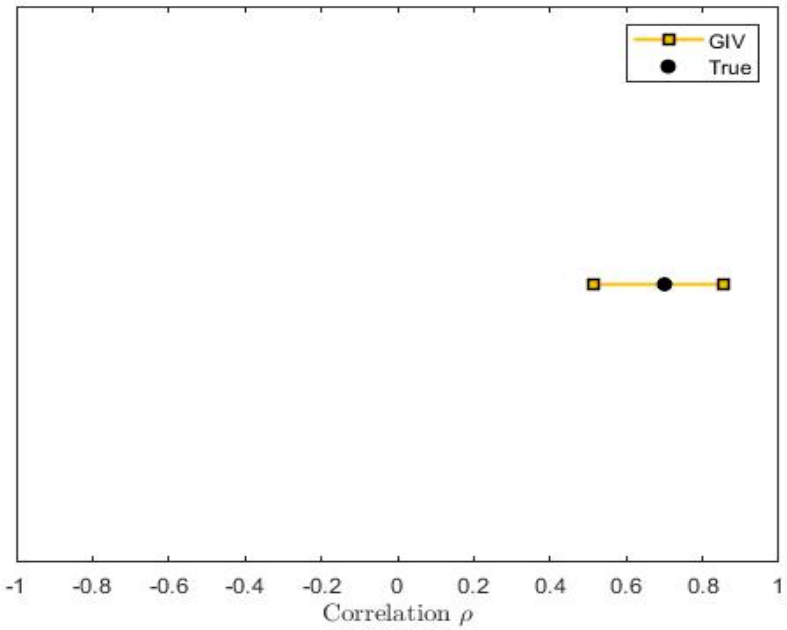


Figure 18: Projection of identified set for ρ



References

Abbring, J. H., and J. R. Campbell (2010): "Last-In First-Out Oligopoly Dynamics," *Econometrica*, 78(5), 1491–1527.

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcomes," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 6A, chap. 63. North-Holland.

Altonji, J., and R. L. Matzkin (2005): "Cross-Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053–1102.

Anderson, T., and C. Hsiao (1981): "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, 598–606.

Andrews, D. W., and X. Shi (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81(2), 609–666.

Arcidiacono, P., and R. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 7(6), 1823–1868.

Arellano, M., and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Athey, S., and G. Imbens (2006): "Identification and inference in nonlinear difference-in-difference models," *Econometrica*, 74(2), 431–497.

Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

Benkard, C. L., A. Bodoh-Creed, and J. Lazarev (2010): "Simulating the Dynamic Effects of Horizontal Mergers: U.S. Airlines," Discussion paper, Stanford GSB.

Beresteanu, A., F. Molinari, and I. Molchanov (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79(6), 1785–1821.

Berry, S., and E. Tamer (2007): "Identification in Models of Oligopoly Entry," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by W. N. R. Blundell, and T. Persson, vol. 2. Cambridge University Press.

Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60(4), 889–917.

Berry, S. T., and P. A. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82(5), 1749–1797.

Blevins, J. (2016): "Sequential Monte Carlo Methods for Estimating Dynamic Microeconomic Models," *Journal of Applied Econometrics*, 31(5), 773–804.

Blundell, R., and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87, 115–143.

Borkovsky, R. N., U. Doraszelski, and Y. Kryukov (2010): "A User's Guide to Solving Dynamic Stochastic Games Using the Homotopy Method," *Operations Research*, 58(4-part-2), 1116–1132.

Bresnahan, T. (1989): "Empirical Studies of Industries with Market Power," in *The Handbook of Industrial Organization*, ed. by R. Schamlensee, and R. Willig, no. 10 in Handbooks in Economics. North-Holland.

Chernozhukov, V., D. Chetverikov, and K. Kato (forthcoming): "Inference on causal and structural parameters using many moments inequalities," *Review of Economic Studies*.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models1," *Econometrica*, 75(5), 1243–1284.

Chesher, A. (2010): "Instrumental Variables Models for Discrete Outcomes," *Econometrica*, 78, 575–601.

Chesher, A., and A. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 83(3), 959–989.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

Collard-Wexler, A. (2013): "Demand Fluctuations in the Ready-Mix Concrete Industry," *Econometrica*, 81(3), 1003–1037.

——— (2014): "Mergers and Sunk Costs: An Application to the Ready-Mix Concrete Industry," *American Economic Journal: Microeconomics*, 6(4), 407–447.

Doraszelski, U., and M. Satterthwaite (2010): "Computable Markov-perfect industry dynamics," *The RAND Journal of Economics*, 41(2), 215–243.

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

Galichon, A., and M. Henry (2011): "Set identification in models with multiple equilibria," *Review of Economic Studies*, 78(4), 1264–1298.

Heckman, J., and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52(2), 271–320.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016): "Dynamic treatment effects," *Journal of Econometrics*, 191, 276–292.

Honoré, B., and E. Tamer (2006): "Bounds on parameters in panel dynamic discrete choice models," *Econometrica*, 3, 611–629.

Hotz, J., and R. A. Miller (1993): "Conditional Choice Probabilites and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

Hotz, V. J., R. A. Miller, S. Sanders, and J. Smith (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *The Review of Economic Studies*, 61(2), pp. 265–289.

Hu, Y., and M. Shum (2012): "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171, 32–44.

Hu, Y., M. Shum, W. Tan, and R. Xiao (2015): "A Simple Estimator for Dynamic Models with Serially Correlated Unobservables," Discussion paper, Working Paper.

Israel, M. (2005): "Tenure dependence in consumer-firm relationships: an empirical analysis of consumer departures from automobile insurance firms," *RAND Journal of Economics*, 36(1), 165–192.

Kalouptsidi, M., P. Scott, and E. Souza-Rodrigues (2018): "Linear IV Regression Estimators for Single-Agent Dynamic Discrete Choice Models," Discussion paper, Harvard.

Kasahara, H., and K. Shimotsu (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77(1), pp. 135–175.

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

Magnac, T., and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70(2), 801–816.

Manski, C. F. (2003): *Partial Identification of Probability Distributions*. Springer, New York.

Manski, C. F., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

Olley, S. G., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–97.

Pakes, A., and R. Ericson (1996): "Empirical Implications of Alternative Models of Firm Dynamics," *Journal of Economic Theory*.

Pakes, A., and P. McGuire (2001): "Stochastic Approximation for Dynamic Analysis: Markov Perfect Equilibrium and the 'Curse' of Dimensionality," *Econometrica*.

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple Estimators for the Parameters of Dynamic Games, with Entry/Exit Examples," *RAND Journal of Economics*, 38(2), 373–399.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and Their Application," *Econometrica*, 83(1), 315–334.

Pesendorfer, A. (2009): "Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables," *Econometrica*, 77(5), 1665–1682.

Pesendorfer, M., and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75(3), 901–928.

——— (2010): "Sequential Estimation of Dynamic Discrete Games: A Comment," *Econometrica*, 78(2), 833–842.

Reich, G. (2018): "Divide and Conquer: Recursive Likelihood Function Integration for Hidden Markov Models with Continuous Latent Variables," *Operations Research*, 66(6), 1457–1470.

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

Ryan, S. P. (2012): "The Costs of Environmental Regulation in a Concentrated Industry," *Econometrica*, 80(3), 1019–1061.

Somaini, P. (2015): "Identification in Auction Models with Interdependent Costs," Discussion paper, Stanford.

Stokey, N. L., R. E. Lucas, and E. C. Prescott (1989): *Recursive Methods in Economic Dynamics*. Harvard Univ., Cambridge MA.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70(1), pp. 147–165.