Review of Economic Studies (2023) 90, 1724–1758 https://doi.org/10.1093/restud/rdac061

Advance access publication 8 September 2022

An Instrumental Variable Approach to Dynamic Models

STEVEN T. BERRY
Yale University, Cowles Foundation and NBER

and

GIOVANNI COMPIANI
University of Chicago, Booth School of Business

First version received August 2020*; Editorial decision February* 2022*; Accepted September* 2022 (*Eds.*)

We present a new class of methods for identification and inference in 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 methods to identify those dynamic policy functions that are consistent with instrumental variable (IV) restrictions. Extending popular "two-step" 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 sunk costs.

Key words: Dynamic models, Instrumental Variables, State Dependence, Initial Conditions

JEL Codes: L1, C36, C57

1. INTRODUCTION

We propose an instrumental variable (IV) approach to identification and inference for infinite-horizon 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 inference. As a result, much of the literature to date either assumes away serial correlation in the unobservables or else deals with such correlation in particularly simple fashions.

We mostly focus on 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. As an example, consider a simple model of entry and exit where profits depend on the number of firms in the market, x. Standard empirical approaches assume that any

The editor in charge of this paper was Francesca Molinari.

variables affecting firms' profits and not captured by the data—e.g. unobserved demand or supply shocks—are independent over time. This implies that "market structure" x is independent of the contemporaneous unobservables and thus the effect of x on, say, firm entry can be directly observed in the data. In contrast, when the unobservables are persistent over time, markets with large x are likely to be more systematically profitable in terms of unobservables. Thus, the observed entry probabilities reflect the correlation between unobservables and x, and we cannot directly observe the "causal effect" of x on entry. This is a classic endogeneity problem.

A natural and economically meaningful solution to this familiar IO problem of endogenous market structure is to use IV methods. More broadly, this article is a 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]

1.1. *Idea of the article*

Our article 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 policy function that results from the solution of that dynamic model evaluated at the true value of the structural parameters. It is this policy function that (according to the model) generates the data.

The task of identification and inference is made much easier by the 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 IVs. 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 affected past policy decisions that are correlated with present states. In a

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

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

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

Traditional IV and panel data methods face a difficult problem in our context: the policy function is derived from the "structural" dynamic model and this typically implies that the policy function is not additively separable in the serially correlated unobservable(s). The non-separability of the policy function in unobservables creates difficulties for both identification and inference. Luckily, there is a large recent literature on the non-parametric 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 article 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 *et al.* (2011) to characterize the sharp identified set.[3] This characterization will help us build intuition about how instruments serve to (set) identify policy functions.

The identifying power of these IV methods is increased by the presence of multiple periods of data. In particular, we note that even in the absence of any IVs, non-separable policy functions can be usefully restricted purely from the presence of multiple periods of data, as in the nonseparable error, non-parametric panel data papers of Altonji and Matzkin (2005) and Athey and Imbens (2006).[4]

We illustrate our approach in a simple single-agent entry and exit model. This minimal example allows us to build intuition about the sources of identification as well as to explore how the number of time periods, the presence of exogenous covariates and the strength of the instruments affect the identified set for the structural parameters. We then apply the method to data from the US ready-mix concrete industry and consider a counterfactual policy that increases the magnitude of the sunk costs of entry into the market. When we compare our approach to three different methods that assume away serial correlation in the unobservables, we find that the latter results are significantly different than ours in one of our model specifications. Moreover, the sign of the bias varies across the three methods. Two approaches tend to over-predict the responses to the policy in terms of both the number of firms and the fraction of new entrants. This stems from the fact that, in the counterfactual, the unobservables exhibit too much volatility over time when serial correlation is ruled out. On the other hand, a third approach estimates a very large sunk cost (as a way to match the persistence in the data without appealing to serially correlated unobservables)

3. These papers in turn build on advances in random set theory (Artstein, 1983). See the comprehensive treatment in Molchanov and Molinari (2018), as well as the discussions in Chesher and Rosen (2020) and Molinari (2020).

4. These papers do not explicitly consider the fully dynamic problems that we consider here but instead focus on non-parametric analogues of non-dynamic panel data-style arguments.

and thus predicts no response to the policy change. The difficulty in *a priori* signing the bias from standard, more restrictive methods further motivates our contribution.

1.2. *Some related papers*

The literature on the identification and estimation of dynamic problems is immense, and we can only highlight a set of related literatures here.

We are obviously not the first authors to consider the issue of serially correlated unobservables in dynamic models, including dynamic games. Outside of the two-step literature following on CCP methods, there is an important set of papers emphasizing computational approaches to estimation that allow for some form of persistent unobservables, sometimes in the form of a limited number of "discrete types" of agents. A classic single-agent example is Keane and Wolpin (1997). A classic oligopoly example is the full-solution 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. These computationally oriented papers do not typically discuss formal identification.

Work on the identification of mixture models, as in Kasahara and Shimotsu (2009), provides some formal results on identification of discrete dynamic policy functions with persistent unobservables. This work again emphasizes limited forms of discrete heterogeneity. In Berry and Compiani (2020a), we show that our framework includes the class of models they consider as a special case. 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 particular, the degree of point-identified heterogeneity is limited by the time-series dimension of the data. As a complementary result, our set-identification approach is applicable to settings with as few as two time periods irrespective of the dimension of the unobservable. Of course, if the data exhibit too little variation, our identified sets may be so large as to be of little use. In our empirical application, we obtain informative results with fewer than 500 cross-sectional observations. Hu and Shum (2012),[5] obtain point-identification results in a single-agent model featuring possibly continuous, time-varying scalar unobservables. As in Kasahara and Shimotsu (2009), a key source of identification is that the degree of persistence in the unobservables is assumed to be smaller relative to the variation in the data, together with a completeness condition in the case where the unobservables are continuous. Our approach does not require either, but in general yields partial identification and makes explicit use of instruments.

On the estimation side, Arcidiacono and Miller (2011) provide maximum likelihood computational methods for the structural parameters of dynamic models with discrete persistent heterogeneity. Norets (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.

Our article is also related to the literature on dynamic panel data models that are not derived from explicit dynamic optimization (e.g. Altonji and Matzkin, 2005; Athey and Imbens, 2006) and to the large literature on distinguishing between state dependence and unobserved heterogeneity (see e.g. Heckman and Singer, 1984; Israel, 2005; Dubé, Hitsch and Rossi, 2010; Torgovitsky, 2019). Similar to these papers, we face the challenge of disentangling the roles of past actions and persistent unobservables in driving current outcomes. Our IV intuition is very much in line with this literature on dynamic panels with state dependence. For

5. See also Hu, Shum, Tan and Xiao (2015).

example, Israel (2005) argues for the usefulness of past exogenous shocks that shift the current state but do not affect today's decision conditional on the current state. More recently, Heckman, Humphries and Veramendi (2016) study identification and estimation of dynamic treatment effects allowing for time-invariant unobserved heterogeneity.

In the dynamic panel context, Honoré and Tamer (2006) make a set of observations that are closely related to our motivations and our work. First, they note that the initial conditions problem in a dynamic setting leads to an "incomplete model" that may not be point identified. This calls into question approaches like Heckman and Singer (1984) that attempt to close the model with a pure functional form assumption. However, Honoré and Tamer (2006) note that the incomplete model still imposes bounds on the parameters that, in practice, may be quite informative. An implication is that it may be better to use these bounds instead of "completing" the model via an *ad hoc* assumption on initial conditions. The Honoré and Tamer approach has some strong similarities to our approach that places bounds on a dynamic policy function, but they do not explicitly discuss IVs and they do not consider models of optimal dynamic behaviour.

Even without serially correlated unobservables, there are typically no formal point-identification results for models with continuous actions,[6] which is one reason why Bajari *et al.* (2007) ("BBL") relies on set-identified inequality methods to recover the structural profit parameters, once given a "first-step" identified policy function. We propose much simpler second step methods, which aids in translating the policy functions (set) identified in our GIV first step into second-step profit parameters.

A recent paper by Kalouptsidi, Scott and Souza-Rodrigues (2021a) shares our IV intuition and is in many ways closest to our spirit. They show that in a class of dynamic discrete choice models with serially correlated market-level unobservables, one can obtain Euler equations that point-identify some firm-specific profit parameters. This approach leads to computationally light linear IV estimators that are robust to endogeneity problems caused by the market-level 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 distribution as an input. The paper provides some interesting examples of IV potential applications with serially correlated market-level states, including durable goods demand, land use, and dynamic labour supply. They discuss possible instruments in these settings. Their examples and instruments could be applicable to our methods as well. Blending our method with theirs might be a fruitful direction for applied work.

The rest of this article is organized as follows. Section 2 introduces the general notation and model. Section 3 discusses identification for single-agent problems. Section 4 illustrates the approach via numerical examples. Section 5 extends the analysis to the oligopoly setting. Section 6 contains the empirical application and Section 7 concludes.

2. MODEL

In this section, we describe the formal model. After introducing variables and notation, we focus on the single-agent case and present a simple monopoly entry example, which we will use throughout the rest of the article as an illustration.

6. Identification in the case of no serial correlation and discrete actions is considered in Magnac and Thesmar (2002) and related papers.

2.1. *Variables and notation*

We consider a model that generates data on a large set of markets, with one or more agents[7] 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 by a_{ijt}. These actions contribute over time to the firm's observed current state(s), denoted by x_{ijt}. The set of feasible actions for a firm with state x_{ijt} is denoted $\mathcal{A}(x_{ijt})$. As one example, in an entry model there might be a scalar action a_{ijt}, equal to 1 or 0, 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 will later prove useful as instruments, by r_i.

In addition, there are unobserved (to us) state(s) u_{ijt} that also evolve exogenously from the actions of firms. For example, in an entry model, u_{ijt} may represent the component of fixed costs not captured by the data. Within a 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 is private information to the firm.[8]

Suppose that there are a maximum of J firms within each market. We define

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

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 by $a_i=(a_{i1},\ldots,a_{iT})\in\mathbb{A}^T$, $x_i=(x_{i1},\ldots,x_{iT})\in\mathbb{X}^T$, $w_i=(w_{i1},\ldots,w_{iT})\in\mathbb{W}^T$, and $u_i=(u_{i1},\ldots,u_{iT})\in\mathbb{U}^T$.

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

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

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 structural

7. Since many IO dynamic models involve firms making decisions over time, we use the words "firms" and "agents" interchangeably throughout the article.

8. 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).

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	No. of stores	$\mathcal{I}^+$	$x_{it+1} = a_{it}$
Quality	R&D	$\mathbb{R}^+$	$x_{it+1} \sim f(x_{it}, a_{it})$

parameters, θ, then includes the unknown parameters of the single-period profit function and of the distribution of unobservables: $\theta = (\theta_\pi, \theta_u)$.[9]

2.2. *Single firm per market*

We begin with the single-agent case, returning to dynamic oligopoly in Section 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.

Similarly, we focus on the case where the exogenous states are first-order Markov, and let

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

and

$$\tilde{\Phi}(u_{it+1} | u_{it}; \theta_u) \tag{6}$$

be the transition probabilities for w_{it} and u_{it}, respectively. The first-order Markov assumptions are not needed for the identification argument; however, relative to the fully general model, they reduce the dimension of the state space and are thus helpful when conducting inference.[10] Our leading example here will be a model of first-order serial correlation where u_{it} is a scalar that obeys $u_{it} = \rho u_{it-1} + \nu_{it}\sqrt{1-\rho^2}$. In the simplest case, the period t innovation ν_{it} might be assumed to have a simple parameterized distribution. The parameter θ_u then includes those parameters plus the serial correlation parameter ρ.[11]

9. Following standard practice in the dynamics literature, we assume that the discount factor is known throughout.

10. Our framework also allows for the case where the transition processes for x_{it} and w_{it} in (4) and (5) depend on u_{it} (or components of it) provided that they can be identified from the data. Since we know of no empirical models featuring this dependence, we focus on the case where the transitions do not depend on u_{it} throughout the article. Similarly, the transitions for u_{it} could in principle depend on the observed states. None of the examples that we consider in this article has this feature and thus we maintain the assumption in (6) for ease of notation.

11. Berry and Compiani (2020a) consider cases where the unobservable consists of both a time-invariant discrete component and a serially uncorrelated shock in the spirit of Heckman and Singer (1984), Keane and Wolpin (1997), Kasahara and Shimotsu (2009), Arcidiacono and Miller (2011), and related literature.

The firm's dynamic problem is given by the classic Bellman equation:

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

where δ denotes the discount factor and V the value function. Note that we assume a stationary environment and thus let the function V be time invariant. The expected value function in this expression is

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

Equation (8) embeds the restriction that the conditional distribution of $\left(x_{it+1}, w_{it+1}, u_{it+1}|a_{it}, x_{it}, w_{it}, u_{it}\right)$ factors into $\Gamma(x_{it+1}|a_{it}, x_{it}, w_{it})H(w_{it+1}|w_{it})\tilde{\Phi}(u_{it+1}|u_{it}; \theta_u)$. This has been a standard assumption in the literature since Rust (1987). However, relative to the existing models, we relax the assumption that u_{it} be independent over time, which is why 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 standard conditions such that there is a unique policy function consistent with that value function.[12] We let σ denote the policy function that generates our data, so that the observed actions are

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

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 possible parameter vector $\theta = (\theta_\pi, \theta_u)$. In the single-firm case, these policy functions, generated by the model via the unique solution to Bellman's equations at different θ, are denoted by $a_{it} = \sigma_\theta(x_{it}, w_{it}, u_{it})$. Note that σ_θ is generated purely by Bellman's equation with no reference to the data.

The policy functions σ and σ_θ are connected because, according to our model, $\sigma(x_{it}, w_{it}, u_{it})$ is generated by Bellman's equation applied at the true value of the parameters. In the argument below, the distinction between σ and σ_θ is important because it allows us to separately ask [1] whether a given $\sigma \in \mathcal{F}$ is consistent with the data and IV conditions and [2] whether such a function is additionally consistent with some vector θ, in the sense that it matches the policy function associated with θ via the full dynamic model (i.e. the Bellman equation). As in Hotz and Miller (1993), this distinction will form the basis of our two-step identification procedure.

We now introduce a minimal single-agent model that we will use as an illustration throughout the article.

12. For the technical conditions guaranteeing a unique policy function in the case of a continuous policy, see Theorem 9.8 in Stokey, Lucas and Prescott (1989). Rust (1987) and many other papers provide examples of unique policy functions in the discrete case.

Example 1. *Consider a monopolist deciding whether to be in or out of the market in each of multiple time periods. In this example, the state is whether a firm is "Out" of or "In" 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 assume that exit is reversible and that there are no exogenous profit shifters w (for now). A firm that is already in the market ($x_{it} = 1$) and decides to stay in ($a_{it} = 1$) earns a single-period profit equal to $\bar{\pi} - \epsilon_{it}$ where $\bar{\pi}$ is the deterministic part of variable profits. A firm that is out of the market ($x_{it} = 0$) and decides to enter ($a_{it} = 1$) earns the same single-period profit minus a sunk cost γ. Whenever a firm decides to be inactive ($a_{it} = 0$), it earns zero profits. We interpret ϵ_{it} as a shock reflecting variation in per-period fixed costs and assume that it follows a first-order autocorrelation process,*

$$\epsilon_{it} = \rho \epsilon_{i,t-1} + \nu_{it} \sqrt{1-\rho^2}, \tag{10}$$

where ν_{it} is distributed standard normal.[13] *The resulting model then has three structural parameters: $\bar{\pi}, \gamma$, and ρ. The policy function that generates the data is*

$$a_{it} = \sigma(x_{it}, u_{it}), \tag{11}$$

where $u_{it} \sim \mathrm{Unif}(0,1)$ can be normalized to be the quantile of ϵ_{it}.[14] *We assume that the dynamic model generates the natural monotonicity results that σ is weakly increasing in x_{it} and weakly decreasing in u_{it}.*

3. IDENTIFICATION IN THE SINGLE-AGENT MODEL

Given the model described in the previous section, we now discuss identification. Again, we first focus on the single-agent case and refer the reader to Section 5 for the identification of the oligopoly model. At the end of this section, we briefly discuss inference in light of these identification results.

For purposes of identification, we assume that we observe the true distribution from which the data is drawn and denote it by $P(a_i, x_i, w_i, r_i)$, where again r_i denotes excluded exogenous variables. This is equivalent to seeing a T-period panel on a very large (in fact, infinite) cross-section of firms or agents. We wish to identify (possibly set-identify) the parameters θ. When (a_i, x_i, w_i) are discrete and the unobservables enter profits additively, the single-period profit function may be fully flexibly characterized by a finite number of parameters, one for each combination of (a_i, x_i, w_i) values. On the other hand, in the continuous case, flexibly modelling the profit functions generally leads to infinite-dimensional parameters θ. While this is allowed by our identification argument, the presence of an infinite-dimensional θ would complicate computation and inference, and we leave adapting the implementation of our approach to this setting for future work.[15]

The potential instruments in the model consist of the exogenous variables $z_i = (r_i, w_i)$. The critical assumption that allows for our IV approach is independence of the instrument and the unobservables:[16]

$$z_i \perp u_i.$$

13. Given the parameterization of the profit function, fixing the mean and variance of ν_{it} is without loss.

14. To see this, one can write $a_{it} = \overline{\sigma}\left(x_{it}, F_\epsilon^{-1}(F_\epsilon(\epsilon_{it}))\right)$, where $\overline{\sigma}$ is a non-parametric function and F_ϵ is the cdf of ϵ_{it}. Equation (11) then follows by defining $u_{it} = F_\epsilon(\epsilon_{it})$ and $\sigma(x_{it}, \cdot) = \overline{\sigma}\left(x_{it}, F_\epsilon^{-1}(\cdot)\right)$.

15. Compiani (2019) proposes a sieve estimator for a nonparametric regression model with shape restrictions that could be applied to the policy function (9) under point-identification.

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

TABLE 2
Examples of possible instruments r_i

State	Example instruments
Capital	Past investment cost
Out/in of market	Past market population, past regulation
No.# of stores	Distance from headquarters, interacted with time
Quality	Past R&D shocks, age of firm

Note that the assumption that w_i be exogenous is standard in the existing literature. In addition, we require excluded instruments r_i to deal with the endogeneity of the dynamic states x_i. Table 2 gives some ideas of possible instruments r_i 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.

In studying identification of the model, we follow the classic "two-step" approach. First, we discuss the (set-)identification of the policy function and serial correlation parameters, using GIV techniques. Given the results of the first step, we then discuss the identification of the structural parameters of the profit function using a broad generalization of existing approaches.

3.1. *First step: identification of the policy*

The broad idea is to (set-)identify the policy function from classic IVs conditions, extended to cases where the policy function is highly nonlinear in the states. The GIV framework achieves this, and it 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 (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 (Chesher, 2010); and (3) lack of point-identification of the parameters, even in the absence of Problems 1 and 2, e.g. due to instruments that are not strong enough.

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 sharp 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, i.e. the fact that the joint distribution of (x_{i1}, u_{i1}) is not known.[17] Traditionally, solutions to the initial conditions problem include either (1) parameterizing the initial joint distribution of states and unobservables or (2) specifying some process for the past history of the firm that uses the model parameters to construct that same initial joint distribution.[18] We argue that if the parameterization in method (1) is so flexible as to not impact the resulting identified set, then we might just as well look for the sharp identified set that does not restrict the initial distribution.

As in Tamer (2003), any given action a_{it} naturally leads to conditions on sets of unobservables. In particular, following CR and using similar notation, if the sequence (a_i, x_i, w_i) occurs, then u_i

17. 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.

18. 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.

must be in the inverse image set $\mathcal{U}(a_i, x_i, w_i, \sigma) \equiv \{u_i : \sigma(x_{it}, w_{it}, u_{it}) = a_{it}, \forall t\}$. The condition

$$\{u_i \in \mathcal{U}(a_i, x_i, w_i, \sigma)\} \tag{12}$$

is then a necessary condition for the observed event (a_i, x_i, w_i). If the model is incomplete, however, that condition is not sufficient for the event: when the exogenous variables (w_i, u_i) do not uniquely pin down the endogenous variables (x_i, a_i), it can happen that (12) is satisfied but the event (a_i, x_i, w_i) does not occur. This gives rise to the following characterization of the identified set for the policy and parameters for the unobservables based on inequalities (Artstein, 1983): a pair (σ, θ_u) is in the identified set if and only if for all closed sets $\mathcal{S} \subset \mathbb{U}^T$ and for all z

$$\Pr\big(\mathcal{U}(a_i, x_i, w_i, \sigma) \subseteq \mathcal{S} \mid z\big) \leq \Pr(u_i \in \mathcal{S} | z; \theta_u). \tag{13}$$

which, under the assumption of independence between z_i and u_i, simplifies to

$$\Pr\big(\mathcal{U}(a_i, x_i, w_i, \sigma) \subseteq \mathcal{S} \mid z\big) \leq \Phi(\mathcal{S}; \theta_u). \tag{14}$$

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 is the probability of an event whose occurrence can be observed in the data. The right-hand side is the probability of that necessary condition with respect to the distribution of u_i, which by assumption does not depend on z. Further, given θ_u, this term is known and can be computed in closed form or via simulation.[19]

In the case with discrete a_{it}, x_{it}, w_{it}, CR shows that to obtain the sharp identified set for θ, one only needs to check sets $\mathcal{S}$ (labelled "core-determining") that belong to a collection $Q(\sigma, z_i)$.[20] This collection includes the "elemental" sets, $\mathcal{U}(a_i, x_i, w_i, \sigma)$, associated with individual realizations of the observables, as well as appropriately chosen unions of sets of that form.[21] We illustrate these sets in Example 1 below and Supplementary Appendix A.

The CR approach operates in the space of the unobservables and it builds on earlier results that apply to the space of observables (Beresteanu *et al.*, 2011; Galichon and Henry, 2011). For us, the result is useful because it completely characterizes the inequality restrictions that define the sharp identified set of policy functions. We will illustrate this in the simple monopoly entry model of Example 1. However, the number of these restrictions can grow quite large in realistic problems. In these cases, one may not be able to list all the CR inequalities needed to obtain the sharp identified set. However, the CR characterization is still helpful to build intuition for selecting which inequalities to impose. We do this in the empirical application of Section 6 and show that, while we do not get sharp identification, the results are still informative.

When the necessary conditions (14) 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 (since e.g. the instruments might not be strong enough) and so, in the absence of a proof of point-identification, we might still want to consider set identification.

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

20. The notion of core-determining class was first introduced by Galichon and Henry (2006).

21. Note that in any particular example $Q(\sigma, z_i)$ need not be "irreducible", i.e. it need not be the smallest possible collection of sets that have to be checked.

FIGURE 1
Policy cutoffs in the one-period case

TABLE 3
Inverse image sets and & inequalities for the one-period example

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

To formalize the argument above, we 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

$$\Sigma^{IV}(\theta_u)\equiv\{\sigma:\text{condition (14) holds }\forall\mathcal{S}\in Q(\sigma,z)\text{ and }\forall z\}\cap\mathcal{F}. \tag{15}$$

We emphasize that $\Sigma^{IV}(\theta_u)$ is a subset of the space of admissible policies $\mathcal{F}$ and, as such, incorporates all the natural economic restrictions—e.g. monotonicity—that one may be willing to impose on $\mathcal{F}$. Definition (15) immediately gives the following characterization of the sharp identified set for (σ,θ_u) based on the GIV restrictions.

Result 1. *Given the GIV restrictions in* (15)*, the sharp identified set for* (σ,θ_u) *is given by* $\big\{(\sigma,\theta_u):\sigma\in\Sigma^{IV}(\theta_u),\,\Sigma^{IV}(\theta_u)\neq\emptyset\big\}$.

It should be noted that this set is sharp in the sense that it incorporates all the information contained in the GIV restrictions (as well as natural constraints on the policy functions, such as monotonicity). However, this set does not reflect the restrictions coming from the Bellman equation or the parametric specification of profits. This will be accomplished in the second step.

Example 2. *(Continued). In our monopoly entry example, the first step consists in characterizing the identified set for* (σ,ρ)*. If we focus on only one period of data, the policy function in* (11) *is a non-parametric 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.*

As in Chesher (2010) and Chesher and Smolinski (2012), even one period of data will generate non-trivial bounds on the policy function. As a simple example of GIV restrictions, Table 3 illustrates the elemental 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 in the absence of IVs, but (2) instrumental variable variation is helpful to tighten 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.*

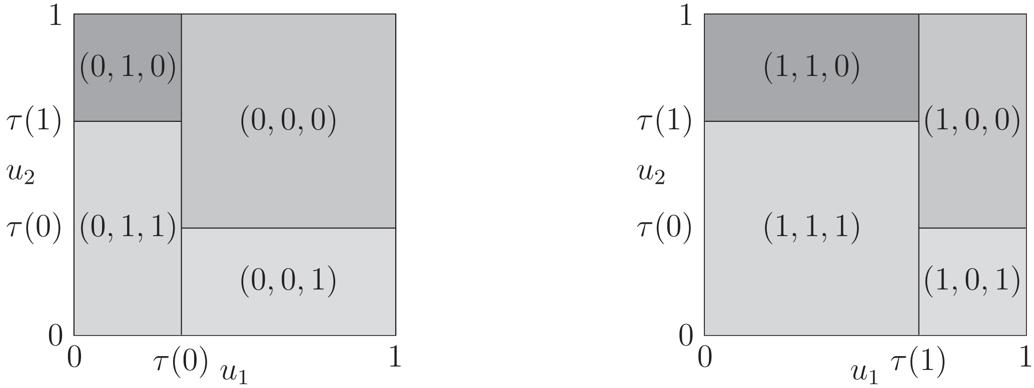

FIGURE 2
Elemental inverse image sets labelled as (x_{i1}, a_{i1}, a_{i2})

We illustrate this by considering two periods of data. Now there are eight elemental 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. The left panel gives the four elemental sets associated with the initial condition $x_{i1} = 0$, while the right panel gives the sets associated with $x_{i1} = 1$. For a given 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.

The sets in Figure 2 allow us to build some intuition about identification in this class of models.[22] *Recall that the probability of 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. 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 time periods, 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. If in large samples we observe the events $(0,1,0)$ or $(1,0,1)$ conditional on any value of the instrument, then we can reject $\rho = 1$ since the associated GIV inequalities of the form (14) have a positive left-hand side and the right-hand side equal to zero.*

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 MLE point-identifies the parameters $(\tau(0), \tau(1), \rho)$, then GIV identifies the same parameter values in this special case.

3.2. *Second step: identification of the profit parameters*

Given the identified set for (σ, θ_u) obtained in the first step, we now show how one can characterize the sharp identified set for the structural dynamic parameters entering the profit function. These

22. The core-determining collection of sets also includes unions of partially overlapping elemental sets and further unions of those sets, which we show in Supplementary Appendix A.

results can be viewed as a generalization of the approach in Hotz, Miller, Sanders and Smith (1994) (HMSS). As noted, for any $\theta = (\theta_\pi, \theta_u)$, we can use the Bellman equation to compute the implied policy σ_θ, defined as follows:

$$\sigma_\theta(x_{it}, w_{it}, u_{it}) \equiv$$
$$\underset{a_{it} \in \mathcal{A}(x_{it})}{\operatorname{argmax}} \left(\pi(a_{it}, x_{it}, w_{it}, u_{it}, \theta_\pi) + \delta E_{\theta_u}\left[V(x_{it+1}, w_{it+1}, u_{it+1}) | a_{it}, x_{it}, w_{it}, u_{it} \right] \right).$$

The sharp identified set of parameters is then given by all (θ_π, θ_u) pairs whose associated policy function is not rejected by the GIV restrictions. We formalize this in the following result.

Result 2. *The sharp identified set for the structural parameters* (θ_π, θ_u) *is given by* $\Theta_{ID} \equiv \{\theta = (\theta_\pi, \theta_u) \colon \sigma_\theta \in \Sigma^{IV}(\theta_u)\}$.

Note that Result 2 imposes the dynamic model—i.e. the restrictions from the Bellman equation and any parametric assumptions on profits—as well as the GIV restrictions. This is the sharp identified set because any θ in this set generates—via the dynamic model—a policy function that cannot be rejected by the data plus the sharp GIV conditions.

A natural question that arises is how to recover the profit parameters θ_π from any given (σ, θ_u) pair. This question has been investigated extensively for models without serial correlation in the unobservables. Here, we extend some of those methods to the case with serial correlation. In particular, we show that, under certain conditions, the parameter θ_π can be conveniently recovered by solving a system of linear equations. To this end, we first set up a system of equations with as many equations as unknowns; then, we use linearity of the profit function to show that the system is linear in the profit parameters.

3.2.1. Setting up a system of equations.

Define the action-specific value function as:[23]

$$v(a, x, w, u; \sigma, \theta) = \pi(a, x, w, u; \theta_\pi) + E_{\theta_u}\left[\sum_{t=1}^{\infty} \delta^t \pi(\sigma(x_t, w_t, u_t), x_t, w_t, u_t; \theta_\pi) \middle| a, x, w, u \right]. \tag{16}$$

Given that the true policy σ is optimal, it must be the case that

$$\sigma(x, w, u) = \underset{a}{\operatorname{argmax}}\, v(a, x, w, u; \sigma, \theta) \tag{17}$$

for every (x, w, u). Thus, if a value θ_π is in the identified set, it must be that it solves (17) for some θ_u and some $\sigma \in \Sigma^{IV}(\theta_u)$. Note that, given a pair (σ, θ_u) from the first stage, verifying this condition is a static optimization problem and is therefore much easier than solving the associated Bellman equation. Therefore, this static problem provides a general second-step method for finding the sharp identified sets, including the case of dynamic games as discussed in Section 5.

In order to illustrate an even simpler approach based on (17), we extend an argument made by HMSS for dynamic discrete choice models with i.i.d. unobservables. Our approach applies to a wide class of models with discrete actions and continuous, possibly serially correlated unobservables.[24] More specifically, we use indifference conditions implied by (17) to write a

23. Much of the literature refers to this as the "conditional" value function.

24. See also Pesendorfer and Schmidt-Dengler (2010). This working paper version of later published work emphasizes an indifference condition interpretation of policy functions in the context of dynamic discrete choice with independent errors.

system of equations in the profit parameters θ_π that has at least as many equations as unknowns. This is a minimal necessary condition for uniquely recovering θ_π.

To this end, fix a pair $\left(\tilde{\sigma},\tilde{\theta}_u\right)$ from the first step and let $\tilde{\theta}_\pi$ denote a value for the profit parameters that is consistent with $\left(\tilde{\sigma},\tilde{\theta}_u\right)$ given the model. For now, we do not assume that such a $\tilde{\theta}_\pi$ is unique; later, we will provide conditions that ensure it is unique. If there is no such $\tilde{\theta}_\pi$, then the model rejects the pair $\left(\tilde{\sigma},\tilde{\theta}_u\right)$. Suppose that, given (x,w) and a pair of actions, a and a', there exists a value of the unobservable, say $\tilde{u}(a,a',x,w)$, such that

$$v\left(a,x,w,\tilde{u}(a,a',x,w);\tilde{\sigma},\left(\tilde{\theta}_\pi,\tilde{\theta}_u\right)\right)=v\left(a',x,w,\tilde{u}(a,a',x,w);\tilde{\sigma},\left(\tilde{\theta}_\pi,\tilde{\theta}_u\right)\right). \tag{18}$$

We do not require that $\tilde{u}(a,a',x,w)$ be uniquely defined by (18), only that it exist.

Assumption 1.. *The variables (a,x,w) take discrete values and for each (a,x,w) there is an action $a'\neq a$ such that there is at least one $\tilde{u}(a,a',x,w)$ satisfying the indifference condition in* (18).

Assumption 1 is a high-level assumption. In Berry and Compiani (2020a), we show the indifference conditions in (18) are a natural extension of equations employed in the HMSS second-step for multinomial discrete choice with independent errors. Not all second-step CCP methods generalize easily (or at all) to the case of serial correlation, but this one does.[25] The next assumption provides more primitive sufficient conditions.

Assumption 2.. *(i) The variables (a,x,w) take discrete values; (ii) the support of u is connected; (iv) the action-specific value function v is continuous in u; (v) for each (x,w), $\tilde{\sigma}(x,w,\cdot)$ takes at least two distinct values.*

Assumption 2(ii) is a standard support restriction; Assumption 2(iii) is also standard and can be verified using results in Stokey *et al.* (1989) (see Theorems 9.8 and 9.11); Assumption 2(vi) can be directly verified by inspecting the $\tilde{\sigma}$ from the first step.

Lemma 1. *Assumption 2 implies Assumption 1.*

Proof. Fix any $(x,w)\in\mathbb{X}\times\mathbb{W}$. By Assumption 2(iv) and the definition of v, there are two actions, a',a'', and two values of the unobservable, u',u'', such that

$$v\left(a',x,w,u';\tilde{\sigma},\left(\tilde{\theta}_\pi,\tilde{\theta}_u\right)\right)\geq v\left(a'',x,w,u';\tilde{\sigma},\left(\tilde{\theta}_\pi,\tilde{\theta}_u\right)\right)$$
$$v\left(a'',x,w,u'';\tilde{\sigma},\left(\tilde{\theta}_\pi,\tilde{\theta}_u\right)\right)\geq v\left(a',x,w,u'';\tilde{\sigma},\left(\tilde{\theta}_\pi,\tilde{\theta}_u\right)\right).$$

Define

$$d(u)\equiv v\left(a',x,w,u;\tilde{\sigma},\left(\tilde{\theta}_\pi,\tilde{\theta}_u\right)\right)-v\left(a'',x,w,u;\tilde{\sigma},\left(\tilde{\theta}_\pi,\tilde{\theta}_u\right)\right)$$

25. HM and Arcidiacono and Miller (2011) employed related multi-period indifference conditions to simplify a second-step in problems with "finite dependence" and independent errors. These methods could be adapted to our case as well if the underlying model featured additive independent errors in addition to any serially correlated component. Kalouptsidi *et al.* (2021a) make use of related finite-dependence indifference conditions in their special case IV method. Further exploration of finite dependence in our context is an interesting future research agenda.

and note that $d\left(u'\right) \geq 0 \geq d\left(u''\right)$. Thus, by the Intermediate Value Theorem for general metric spaces,[26] there exists at least one u''' such that $d\left(u'''\right) = 0$, which proves the claim. ‖

We now show that these assumptions ensure that the system has at least as many equations as unknowns. We consider the case in which profits are parameterized in a flexible way with the elements of θ_π representing the (deterministic) single-period profits for each combination of (a,x,w). However, we maintain the economic restriction that profits are zero when $a=0$ for all (x,w).[27] Under more restrictive parameterization of profits, it may be possible to recover θ_π under weaker conditions.

Lemma 2. *Under Assumption 1, the parameters $\tilde{\theta}_\pi$ associated with the pair $\left(\tilde{\sigma}, \tilde{\theta}_u\right)$ from the first step satisfy a system of equations with at least as many equations as the cardinality of $\mathbb{A}$ minus one times the cardinality of $\mathbb{X} \times \mathbb{W}$.*

Proof. The result follows immediately from the fact that, given Assumption 1, one can write at least as many equation of the form (18) as the cardinality of $\mathbb{A}$ minus one times the cardinality of $\mathbb{X} \times \mathbb{W}$. ‖

3.2.2. Exploiting linearity. Next, we show that, when the single-period profits are linear in θ_π, the above yields a system of linear equations. The coefficients of this system are known given a candidate $\left(\tilde{\sigma}, \tilde{\theta}_u\right)$ from the first step and can be computed via forward-simulation as in BBL.

Assumption 3.. *The single-period profit function is linear in θ_π.*

Result 3. *Fix a pair $\left(\tilde{\sigma}, \tilde{\theta}_u\right)$ from the first step. Under Assumptions 1 and 3, the parameters $\tilde{\theta}_\pi$ associated with $\left(\tilde{\sigma}, \tilde{\theta}_u\right)$ satisfy a system of linear equations, with at least as many equations as the cardinality of $\mathbb{A}$ minus one times the cardinality of $\mathbb{X} \times \mathbb{W}$. Further, the coefficients of the system are known given the model, $\left(\tilde{\sigma}, \tilde{\theta}_u\right)$, and the transition functions in* (4)–(5).

Proof. HMSS show that when the single-period profit is linear in θ_π, then the action-specific value function v is also linear in θ_π. This trivially extends to the case of serially correlated unobservables, so that we can write

$$v\left(a,x,w,u;\tilde{\sigma},\left(\tilde{\theta}_\pi,\tilde{\theta}_u\right)\right) = h_0(a,x,w,u,\tilde{\sigma},\tilde{\theta}_u) + h_1(a,x,w,u,\tilde{\sigma},\tilde{\theta}_u)\tilde{\theta}_\pi,$$

for some known functions h_0, h_1.

For each (a,x,w), Assumption 1 then guarantees that we can write

$$\begin{aligned}\Big[h_1(a,x,w,\tilde{u}(a,a',x,w),\tilde{\sigma},\tilde{\theta}_u) - h_1(a',x,w,\tilde{u}(a,a',x,w),\tilde{\sigma},\tilde{\theta}_u)\Big]\tilde{\theta}_\pi = \\ h_0(a',x,w,\tilde{u}(a,a',x,w),\tilde{\sigma},\tilde{\theta}_u) - h_0(a,x,w,\tilde{u}(a,a',x,w),\tilde{\sigma},\tilde{\theta}_u)\end{aligned} \tag{19}$$

26. See e.g. Theorem 4.22 in Rudin (1976).

27. In general, some restrictions on single-period profits are necessary for identification. While this is natural in our entry example, these restrictions are not normalizations without loss (see e.g. Magnac and Thesmar, 2002; Kalouptsidi *et al.*, 2021b).

for some $a' \neq a$. This establishes the result. ||

Given this linear system, we have two cases depending on whether a standard (and directly verifiable) rank condition is satisfied. If the rank condition holds, we can solve for the unique $\tilde{\theta}_\pi$ in closed form. Otherwise, we obtain multiple values of $\tilde{\theta}_\pi$ associated with the first-step parameters.

One might wonder whether the value(s) of θ_π obtained in the second step are always consistent with the Bellman equation. This, paired with sharpness of the identified set for (σ, θ_u) from the first step, would imply that the identified set for θ_π is also sharp. In the next section, we address this for our illustrative entry/exit example and show that indeed the two-step approach delivers sharp identification of the structural parameters.

We now go back to the monopoly entry model in Example 1 and we illustrate how Result 3 can be used to recover the structural parameters.

Example 3. *(Continued). The goal is to recover the parameters $(\bar{\pi}, \gamma)$ given a pair (σ, ρ) from the first step. For each value of $x \in \{0,1\}$, firms are indifferent between being in and out of the market when $u_{it} = \tau(x)$. This gives two indifference conditions involving action-specific value functions of the form (18).[28] Since the action-specific value functions do not have a closed-form expression, we show how to approximate them via forward simulation. For any a and x,*

$$v^s(a,x,\tau(x),\sigma,\theta) =$$

$$\bar{\pi}\frac{1}{S}\sum_{s=1}^{S}\sum_{t=0}^{\bar{T}}\delta^t a_t^s - \gamma\frac{1}{S}\sum_{s=1}^{S}\sum_{t=0}^{\bar{T}}\delta^t a_t^s 1\{x_t^s = 0\} - \sum_{s=1}^{S}\sum_{t=0}^{\bar{T}}\delta^t \epsilon_t^s a_t^s,$$

where: (i) ϵ_t^s is set to the $\tau(x)$th quantile of ϵ_{it} for $t=0$ and for $t \geq 1$ is drawn using (10) and the correlation ρ from step 1, (ii) x_0^s and a_0^s are set to x and a, respectively, (iii) a_t^s for $t \geq 1$ is determined by the policy σ from step 1, and (iv) $S, \bar{T}$ are large numbers. The two equalities $v^s(0,x,\tau(x),\sigma,\theta) = v^s(1,x,\tau(x),\sigma,\theta)$ for $x \in \{0,1\}$ then give a system of linear equations in $(\bar{\pi}, \gamma)$.

As mentioned above, one might wonder whether the sharpness of the identified set for the first-step parameters is inherited by the identified set for the structural profit parameters in the second step. We study this in the context of Example 1. Specifically, we show that any value $\bar{\pi}$ obtained in the second step is consistent with the model's Bellman equation. Thus, if the identified set for (σ, ρ) from the first step is sharp, the resulting identified set for $\bar{\pi}$ from the second step is also sharp.

Result 4. *In Example 1, let $(\tilde{\sigma}, \tilde{\rho})$ be any pair from the first step with $\tilde{\sigma}$ weakly decreasing in u and $\tilde{\rho} \geq 0$, and let $\tilde{\pi}$ be any value of $\bar{\pi}$ returned by the second step. Then,*

$$\tilde{\sigma}(x,u) = \arg\max_a v(a,x,u;\tilde{\sigma},\tilde{\pi},\tilde{\rho}) \tag{20}$$

for all x, u, i.e. the policy $\tilde{\sigma}$ from the first stage solves the Bellman equation associated with $\tilde{\pi}$ and $\tilde{\rho}$.

28. In order to fully map our entry example into Result 3, note that we implicitly have used the two additional restrictions that single-period profits are zero when firms decide to be out. Thus, we have in total four equations corresponding to the cardinality of $\mathbb{A} \times \mathbb{X}$.

Proof. See Supplementary Appendix E. ‖

So far, the discussion of the second step has focused on the case where the actions and states are discrete. However, a similar logic applies to models with continuous actions. We illustrate this via the following example.

Example ***4 (Continuous-choice stochastic accumulation)*** *Consider the "stochastic accumulation problem" from Pakes (1994) and Doraszelski and Pakes (2007). In this model, firm i chooses the level of investment a_{it}, a continuous variable, based on its current efficiency or quality x_{it} (often taken to be discrete). The distribution of efficiency or quality at $t+1$ is assumed to be stochastically increasing in a_{it}. In this case, the action-specific value function takes the form*

$$\pi(a,x,w,u,\theta_\pi)+\delta E_{\theta_u}\left[\sum_k V(x'=k,w',u')\Gamma(x'=k|x,a)\Big|w,u\right], \tag{21}$$

where both the single-period profit and the transition $\Gamma(x'=k|x,a)$ are differentiable in a.

The optimal action then satisfies

$$\frac{\partial\pi(a,x,w,u,\theta_\pi)}{\partial a}+\delta E_{\theta_u}\left[\sum_k V(x'=k,w',u')\frac{\partial\Gamma(x'=k|x,a)}{\partial a}\Big|w,u\right]=0. \tag{22}$$

Given a candidate $(\tilde{\sigma},\tilde{\theta}_u)$, then we can reject a candidate $\tilde{\theta}_\pi$ unless

$$\begin{aligned}\frac{\partial\pi(\tilde{\sigma}(x,w,u),x,w,u,\tilde{\theta}_\pi)}{\partial a}+\\ \delta E_{\tilde{\theta}_u}\left[\sum_k V(x'=k,w',u')\frac{\partial\Gamma(x'=k|x,\tilde{\sigma}(x,w,u))}{\partial a}\Big|w,u\right]=0.\end{aligned} \tag{23}$$

As in the discrete case, linearity of the profit function is inherited by the value function and forward simulation can be used to approximate the coefficients on $\tilde{\theta}_\pi$, which allows one to back out the profit parameters. Specifically, assume that for each (a,x,w) there is one u that satisfies the first-order condition (23). *Under this modification of Assumption 1 (as well as Assumption 3), one can write a continuum of linear equations in the profit parameters and back out the latter subject to a rank condition.*

We conclude this subsection by noting that our proposed second step does not require the use of any inequality conditions. This is in contrast to existing approaches, such as BBL, which require considering perturbations of the policy function from the first stage and imposing the implied inequalities even in the absence of serially correlated unobservables.

3.3. *Inference and implementation*

So far, we have focused on identification. We now briefly discuss how to obtain confidence regions for the structural parameters. We start by following the two steps outlined above and then discuss alternative approaches.

The sharp identified set for the first-step parameters (σ,θ_u) is characterized by the inequalities (14). Therefore, estimates of the identified set and confidence regions for (σ,θ_u) can be obtained by applying methods from the by now large literature on moment inequalities models (see e.g. Chernozhukov *et al.*, 2007; Andrews and Soares, 2010; Beresteanu *et al.*, 2011; Galichon and Henry, 2011; Andrews and Shi, 2013; Chernozhukov, Lee and Rosen, 2013). As pointed out by CR, one issue that often arises is that the number of inequalities characterizing the sharp identified set is large relative to the sample size, or even infinite. For example, in our empirical application, we consider the entry and exit patterns from a cross section of markets over 12 years. Since there can be zero, one or two firms in the market at any given point in time, the number of inequalities associated with just the "elemental" sets is $3^{12}=531,441$. Fortunately, recent results provide some guidance on how to deal with the "many inequalities" case. References include Menzel (2009), Chernozhukov, Chetverikov and Kato (2019), and Andrews and Shi (2017). In our application, we use one of the bootstrap procedures proposed by Chernozhukov *et al.* (2019) to obtain valid confidence regions for the parameters. Roughly speaking, their approach provides an econometrically disciplined way of determining the subset of moment inequalities that are most informative about the parameter values.

The second step in our approach maps the first-step parameters into the primitive single-period profit parameters. As shown in Result 3, this map only depends on the model and does not involve the data. In particular, all that is needed is knowledge of the action-specific value functions given the first-step parameters. Standard forward-simulation methods can be used for this purpose, as we have illustrated in the context of our simple monopoly entry model. With a large number of simulation draws, the error from the second stage will be negligible relative to sampling error. Alternatively, one could adjust the standard errors to account for any noise from the second step.

The two-step approach is not the only way to conduct inference. We can consider any approach that imposes both the GIV restrictions in (15) and the structure of the Bellman equation. For example, one option is to solve the model for each candidate value of the structural parameter θ and verify whether the implied policy function satisfies the GIV restrictions. This full-solution approach may yield computational gains relative to the two-step method if the first step of the latter involves searching over many policies that are not consistent with the model. For example, the model may assume linearity of the profit function, but since this is generally hard to translate into *a priori* restrictions on the policy (i.e. into the definition of the set $\mathcal{F}$), then the two-step approach might involve a costly search over regions of the policy function space that are not consistent with any linear profit specification. On the other hand, the two-step approach does not require ever solving the model and thus may be preferable in cases where computing an equilibrium is computationally expensive (as in e.g. many oligopoly models). A further consequence of this is that the method is more robust in that it does not require uniqueness of the equilibrium.

The full-solution approach allows one to easily exploit a tight parameterization for the single-period profits. The empirical literature sometimes goes even further and parameterizes both the policy function and the profit function. This has the advantage of reducing the computational burden of the search over policy functions in the two-step method. On the other hand, this "double parameterization" has the undesirable feature that the functional form used for the policy might be inconsistent with the profit parameterization in the sense that no choice of the structural parameters leads—via the model—to the chosen functional form for the policy. Thus, in the numerical illustration of Section 4, as well as in the empirical application (Section 6), we focus on the first two approaches: the two-step procedure and the full-solution method.

4. NUMERICAL ILLUSTRATION

We now compute the identified set for the structural parameters in an instance of our simple entry model. We will pay special attention to how the identified set changes with the number of time periods, the strength of the instruments, and the presence of exogenous profit shifters. In order to abstract from sampling error and focus on the shape of the identified set, we draw a large number of cross-sectional markets (50,000).

We set the deterministic profit parameter $\bar{\pi}$ to 0.5, the sunk cost γ to 1.5, and the correlation parameter ρ to 0.75, so that 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.[29] 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).[30] We compute the three-dimensional identified set for $(\bar{\pi},\gamma,\rho)$ and we plot its projection onto the space of sunk cost and correlation parameters. Note that, since we are treating both the policy and the deterministic profit function fully flexibly, the full-solution method and the two-step approach give the same identified set.

First, we consider how the identified set varies with the IV strength as well as the number of time periods ($T=2$ and $T=10$). In the case with $T=2$, we are able to list all of the GIV restrictions implied by the model and thus obtain sharpness of the identified set. On the other hand, with $T=10$, the number of inequalities in the GIV core-determining class becomes very large. So, instead of listing all the inequalities, we use those corresponding to the sharp two-period GIV identified set as well as those associated with several observable events over the ten time periods.[31] We pick events that intuitively should help us shrink the identified set. Specifically, we use the events "the firm enters at least once", "the firm exits at least once", "the firm enters at least once and exits at least once", and "the number of firms in the market, x_{it}, does not change for at least six consecutive periods". To build some intuition, consider the latter event. We would expect this 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 (the data exhibit a lot of persistence given that we set $\rho=0.75$) and the model necessary conditions on the right-hand side are relatively small (when ρ is close to zero the model predicts little persistence in the observables). A similar argument applies to the other events we include. Figure 3 shows that, when IV strength is low and $T=2$, the identified set is quite large. On the other hand, as expected, the set shrinks considerably as the number of time period grows or the IV becomes stronger (Figure 4).

As a comparison, we report estimates obtained via two standard methods (MLE and GMM) that assume away serial correlation in the unobservables, consistent with most of the existing literature. In the MLE approach, we pool all observations along the cross-section and time-series dimensions and maximize the resulting one-period likelihood. In contrast, in the GMM approach, we use moments based on the two-period and three-period transitions as well, while still restricting ρ to be zero. Both methods—and particularly GMM—tend to overestimate the sunk cost. Intuitively, a model that assumes no serial correlation in the unobservables will load all the persistence in the data onto the sunk cost, thus overestimating its magnitude. In addition, Figures 7 and 8 in Supplementary Appendix A.2 show that MLE and GMM with $\rho=0$ tend to underestimate the profit parameters.

29. For the remaining fraction, we let x_{i1} be drawn from the stationary distribution for the state generated by the model.

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

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

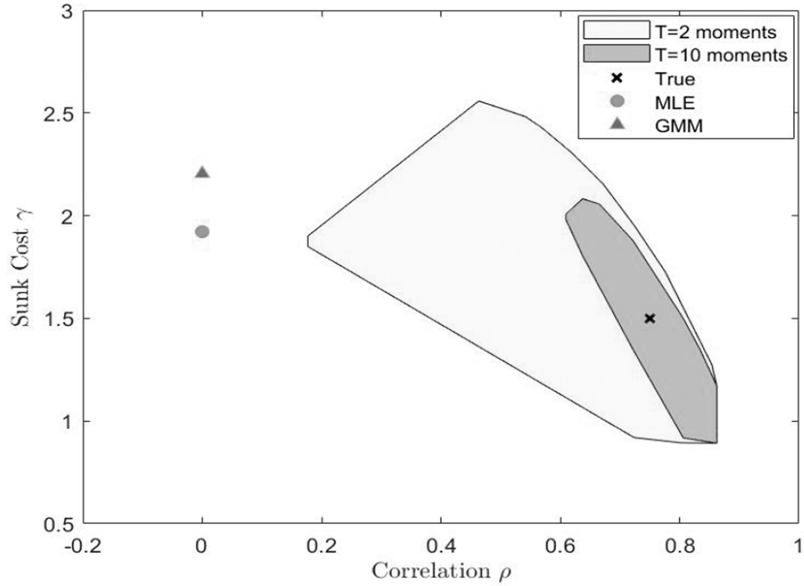

FIGURE 3
Low IV strength

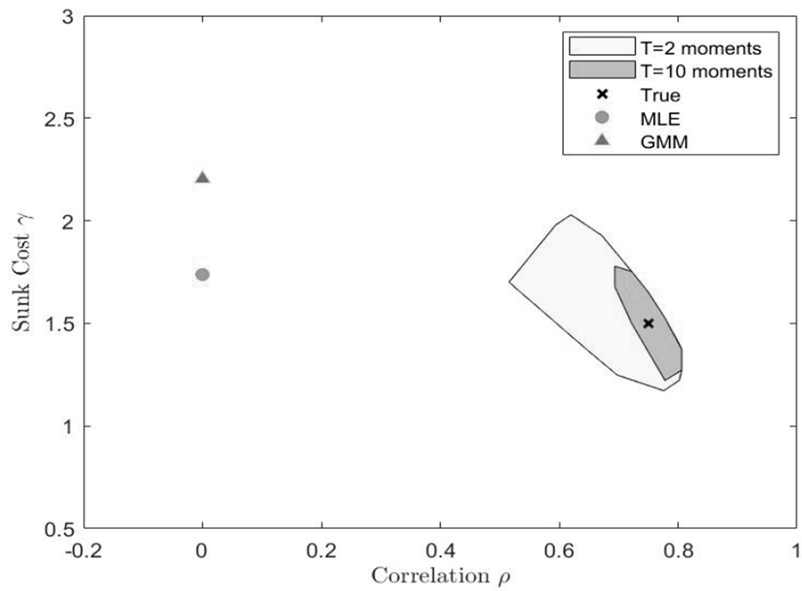

FIGURE 4
High IV strength

Next, we explore the impact of adding an exogenous covariate w_{it} to the model. We specify the firm per-period profit as follows:

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

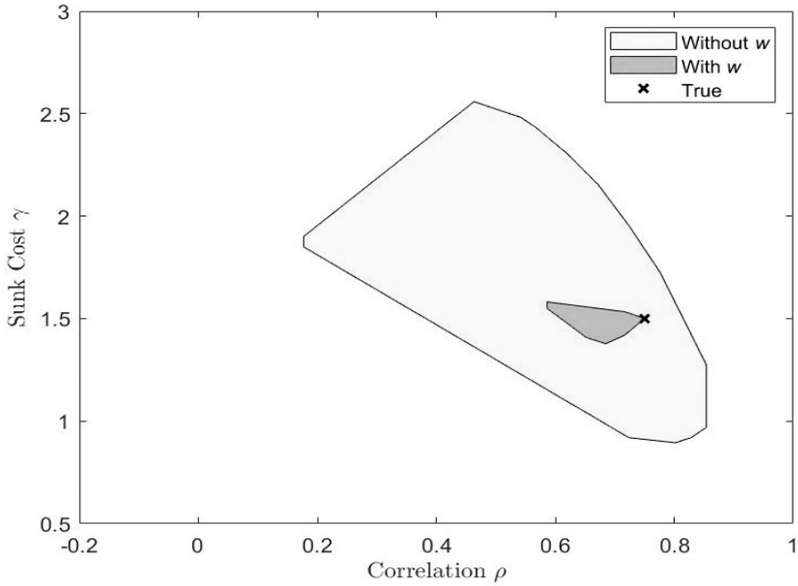

FIGURE 5

Effect of exogenous covariate

so that w_{it} could be interpreted as a measure of market size such as population, $\alpha w_{it}+\beta$ represents variable profits and γ is again the sunk cost of entry. Further, we let ϵ_{it} follow the AR(1) process in (10), with ν_{it} distributed $\mathcal{N}\left(0,\sigma_\nu^2\right)$. Following Pakes *et al.* (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_ν.[32, 33] When generating the data, we set $\alpha=1.5$, $\beta=-1$, $\gamma=1.5$, $\rho=0.75$, $\sigma_\nu=1$, and we focus on the case where IV strength is low. Regarding the distribution of the covariate, we let the initial w_{i0} for each market take the values $\{0.15, 1.00, 1.65\}$ with equal probabilities and evolve according to a transition matrix that has a value of 0.6 on the diagonal and of 0.2 in all the non-diagonal elements. Note that the parameter values are chosen in such a way that the single-period profit function when $w_{it}=1$ is the same as in the model without w_{it}.

Because w_{it} is exogenous, it can be used as an additional conditioning variable in the GIV inequalities along with the excluded IV. Thus, adding w_{it} to the model increases the number of inequalities that each candidate parameter value must satisfy in order to be included in the identified set. On the other hand, the model now features more parameters, so it is not clear *a priori* whether the identified set for, say, γ and ρ is going to be bigger or smaller relative to the case with no variation in w_{it}. As shown in Figure 5, in this case, adding variation in the exogenous covariate substantially shrinks the identified set for γ and ρ.

Finally, we illustrate how the identified sets translate into bounds on counterfactual quantities. For brevity, we focus on the case with $T=2$, strong IV and no exogenous covariate w_{it}. We consider

32. A large literature in IO discusses how to recover variable profits from data on equilibrium quantities and product attributes (see e.g. Bresnahan, 1981, 1982; Berry and Haile, 2014). Our example features sunk costs, but not fixed costs in order to keep the number of structural parameters small. The empirical application in Section 6 allows for both fixed and sunk costs.

33. Note that, unlike the case without w_{it}, here we are not allowed to normalize σ_ν since the non-stochastic part of the period-profits is modelled parametrically. Accordingly, for this design, we report the standardized sunk cost $\frac{\gamma}{\sigma_\nu}$ to ensure comparability with the plots for the case without w.

TABLE 4
Number of firms

	True	GIV	MLE $\rho=0$	GMM $\rho=0$
Baseline	0.659	(0.612, 0.701)	0.817	0.763
↑ Sunk cost (Δ)	0.059	(0.017, 0.123)	0.012	−0.154
Entry subsidy (Δ)	−0.023	(−0.052, −0.011)	−0.167	−0.140

Notes: The first row shows the average number of firms in the market in the absence of policy changes. Rows 2–3 show the change—relative to the first row—in the average number of firms in the market 10 years after the policy change (an increase in the sunk cost and a subsidy to entry, respectively). The dgp has $T=2$, strong IV and no w_{it} covariate.

TABLE 5
Fraction of new firms

	True	GIV	MLE $\rho=0$	GMM $\rho=0$
Baseline	0.591	(0.460, 0.781)	0.574	0.516
↑ Sunk cost (Δ)	−0.193	(−0.349, −0.098)	−0.291	−0.270
Entry subsidy (Δ)	0.124	(0.066, 0.209)	0.366	0.379

Notes: The first row shows the average fraction of new firms in the market in the absence of policy changes. Rows 2–3 show the change—relative to the first row—in the average fraction of new firms in the market 10 years after the policy change (an increase in the sunk cost and a subsidy to entry, respectively). The dgp has $T=2$, strong IV and no w_{it} covariate.

two counterfactual scenarios: (1) an increase in the sunk cost of entry by 4.00—corresponding to 267% of its true value—which we call "the sunk cost counterfactual" for brevity; and (2) a 1.25 subsidy to entry—corresponding to 83% 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 subsidy counterfactual mimics a policy encouraging 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 and look at how the number of firms as well as the fraction of new firms—defined as firms that enter after the policy change—evolve 10 years after each of the two counterfactual changes.[34] We then average across markets as well as realizations of the unobservables.

Tables 4 and 5 show the results for the GIV approach as well as the MLE and GMM models with $\rho=0$. One can see that assuming away serial correlation in the unobservables leads to bias in the estimated reaction to the policy changes, in terms of both the number of firms and the fraction of new firms. The effects of the policies tend to be overstated, in particular for the subsidy counterfactual. Intuitively, when $\rho=0$ the unobservables drawn in the counterfactuals will exhibit much less persistence relative to their true distribution. This, in turn, leads to an excessive amount of predicted entry and exit.

5. DYNAMIC OLIGOPOLY

We now extend the analysis to the setting where multiple firms interact in each market. In this section, we describe the model and study identification. Supplementary Appendix B presents a computed example.

34. This requires choosing an initial value for the unobservable. For each market, we compute the bounds on u_0 implied by (x_0, a_0) and the model, and repeat the exercise twice setting u_0 equal to the upper and lower bounds, respectively. We then take the convex hull of the two resulting outcomes.

In 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 as well as the equilibrium actions of its rivals. Throughout this section, we assume complete information for the serially correlated components of unobservables. If the serially correlated unobservables are not common knowledge across all players, standard equilibrium concepts such as perfect Bayesian equilibrium often become intractable in that they imply that the entire history of play enters the current state. Fershtman and Pakes (2012) propose a tractable framework to deal with persistent sources of asymmetric information and our methods might be usefully merged with theirs.[35]

Since there are multiple firms per market, here, 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

$$\tilde{\Gamma}_j\left(x_{it+1}|a_{ijt},x_{it},w_{it},\sigma_{-j}(x_{it},w_{it},u_{it})\right). \tag{24}$$

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:

$$V_j\left(x_{it},w_{it},u_{it},\sigma_{-j}\right)= \tag{25}$$
$$\max_{a_{ijt}\in\mathcal{A}(x_{ijt})}\left(\pi_j(a_{it},x_{it},w_{it},u_{it};\theta_\pi)+\delta E_{\theta_u}\left[V_j\left(x_{it+1},w_{it+1},u_{it+1},\sigma_{-j}\right)|a_{it},x_{it},w_{it},u_{it}\right]\right).$$

The expected Bellman's equation is

$$E\left[V_j\left(x',w',u'\right)|a,x,w,u,\sigma_{-j}\right]=$$
$$\int\int\int V_j\left(x',w',u',\sigma_{-j}\right)d\tilde{\Gamma}_j(x'|a_j,x,w,\sigma_{-j}(x,w,u))dQ(w'|w)d\tilde{\Phi}(u'|u;\theta_u).$$

Associated with this dynamic program is a best response strategy for firm j, which we assume is unique,[36] 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 vector of equilibrium strategies, σ^*, must satisfy the fixed point

$$\sigma^*=\bar{\sigma}(\sigma^*,\theta). \tag{26}$$

We can then define the set of possible equilibrium policy function vectors as

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

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

35. In the working paper version of this article (Berry and Compiani, 2020b), we illustrate a computed oligopoly problem that features both (1) serially correlated errors that are observed by all the firms and (2) private information shocks that are independent over time. The combination of observed (by rivals) errors and private information errors is reminiscent of the discussion in Pakes *et al.* (2015).

36. Note that we assume uniqueness of the best reply (see footnote 12) but not of the overall equilibrium strategy profile.

the same state vectors.[37] The true policy function that generates the data is then an element of the set $\Sigma^{EQ}(\theta)$, where θ is the true parameter that generates our data.

The sharp identified set of parameters in the oligopoly case is the same as in the single-agent case, except with the further restriction that the policies associated with θ are a vector of equilibrium policies:

$$\Theta_{ID} \equiv \{\theta = (\theta_\pi, \theta_u) : \text{ there exists } \sigma^* \in \Sigma^{EQ}(\theta) \text{ such that } \sigma^* \in \Sigma^{IV}(\theta_u)\}. \tag{27}$$

That is, a parameter vector θ is in the identified set if there is a policy vector that both (1) is not rejected by the IV restrictions and the data (given θ_u) and (2) is a vector of equilibrium strategies given θ.

In practice, we may recover Θ_{ID} via a two-step procedure, just like in the single-agent case. First, note that the argument for the first step from Section 3.1 immediately extends to the oligopoly setting, since displays (12) to (15) continue to hold when the variables $a_{it}, x_{it}, w_{it}, u_{it}$ and the function σ are vector valued. Thus, we can characterize the identified set for the policy vector $\Sigma^{IV}(\theta_u)$ using the GIV restrictions. Second, one can easily extend the dynamic best reply condition (17) to the oligopoly setting as follows:

$$\sigma_j(x,w,u) = \underset{a_j}{\arg\max}\, v_j(a_j, \sigma_{-j}(x,w,u), x, w, u; \sigma, \theta), \tag{28}$$

where v_j is the oligopoly analogue of (16) for firm j. Note that, in defining v_j for the oligopoly case, we treat the actions of rival firms as being generated by σ_{-j} (in addition to generating the future actions of firm j based on σ_j). As in the single-agent case, the best reply in (28) is therefore a static optimization problem and does not require solving any Bellman equations. Furthermore, it is a general condition that can be used to characterize the identified set in any problem. Specifically, the sharp identified set for θ_π is characterized as the values of the profit parameters that solve (28) for some θ_u and some $\sigma \in \Sigma^{IV}(\theta_u)$.

As in the single-agent case, in many oligopoly cases one can find simple indifference conditions that are necessary for the best reply equation in (28). In that case, the second-step search for the identified set of single-period profit function parameters may again be characterized as the set of solutions to a system of linear equations. In the next section, we illustrate this via an oligopoly empirical application. The exact indifference conditions yielding the system of linear equations for our application is given in Supplementary Appendix D.

It may be useful to compare our oligopoly procedure to other two-step procedures in the literature. Our two-step procedure of the last paragraph is quite similar to Bajari *et al.* (2007), with two differences. First, the policy functions are identified via GIV conditions. Second, we obtain the sharp identified set via the static best-reply condition (perhaps through the implied linear indifference conditions), whereas BBL suggest the use of various inequality conditions that are motivated by the same best reply condition.

We can also compare our approach to "full-solution" approaches that search across the set of "structural" parameters (θ_π, θ_u), at each point solving for a (one hopes) unique equilibrium. We believe that our set-identification approach clarifies identification issues in a way that is often hard to do with full-solution approaches. There are also a number of more practical differences. Full-solution methods become particularly hard in the case of possible multiple equilibria.[38]

37. This approach was adopted approximately simultaneously in Bajari *et al.* (2007), Pakes *et al.* (2007), and Pesendorfer and Schmidt-Dengler (2008).

38. 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) provide a homotopy method for exploring a range of possible multiple equilibria.

Once an equilibrium is calculated for a given parameter value, a full computational method must then compare the model's predictions to data. If the model features a natural solution to the initial conditions problem, then the fit to data might be done via maximum likelihood (as in e.g. Rust, 1987; Igami, 2018) or else via a fit of data moments to moments predicted from the model (Pakes and McGuire, 2001). In contrast, our method naturally accounts for unrestricted initial conditions and never requires a computed solution to the equilibrium fixed-point problem, or even a solution to the single-agent contraction mapping. In our method, the "fit to data" is provided by the GIV method. In the absence of an initial conditions problem, the GIV approach often collapses into an MLE or method of moments approach, reducing the difference between the methods.

The computational trade-off is that our method requires us to find a confidence region for the identified set of policy functions. The nature of the trade-off here may vary with the fine details of the problem and (in the absence of multiple equilibria issues) might favour either method. Note that we could, if computationally advantageous, also employ a full solution method. In this case, the "fit to data" for a computed equilibrium would involve testing the GIV conditions for the policy function implied by the equilibrium behaviour. In our empirical example, we employ a full-solution GIV approach (assuming a unique equilibrium and allowing for an initial conditions problem) as well as the two-step approach that we have discussed in this section (which does not require a unique equilibrium).

6. EMPIRICAL APPLICATION TO ENVIRONMENTAL POLICY STYLE COUNTERFACTUALS

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, the article then simulates the evolution of a market following a merger to monopoly and evaluates how long it takes for a second firm to enter.[39] CW imposes an intuitively appealing parametric form for the policy functions, rather than deriving them from an underlying dynamic model. One of our approaches below will roughly mimic this approach.

We use the same data and modelling framework as CW, but estimate all of the structural parameters as opposed to just the policy functions and the serial correlation parameter.[40] This allows us to address the counterfactual effects of policies that affect the "structural" profit function. In particular, we consider policies—similar to the environmental policies in the cement industry study of Ryan (2012)—that alter sunk costs. In addition, since the GIV approach accommodates incomplete models, we are able to tackle the initial conditions problem in a flexible way. CW addresses the initial conditions problem by simulating the probabilities of the initial states via a modification of the GHK algorithm, which requires assuming that the industry has been following the same set of policies for a long time.

Note that our application is intentionally simplified to serve as an example within a longer methodological paper. In particular, for this worked empirical example, we want to avoid the large,

39. Lazarev, Benkard and Bodoh-Creed (2018) also consider counterfactual questions that only require knowledge of the dynamic policy functions, but not the structural single-period profit parameters.

40. 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.

TABLE 6
Summary statistics

Variable	Mean	St. Dev.	Min	Q25	Median	Q75	Max
Number of plants	0.97	0.93	0	0	1	1	6
Construction employment	519	819	3	166	316	592	17,772
% Household income growth 1969–89	0.15	0.11	−0.16	0.08	0.14	0.21	0.69

Notes: Fully balanced panel of 428 markets between 1994 and 2005.

TABLE 7
Ordered probit results

Log construction employment	0.14**
Income growth 1969–89	0.22**
Likelihood-ratio test p-value	0.00

Notes: Dependent variable is number of plants. ** denotes significance at the 95% level.

growing and important literature on methods that solve the computational challenge of methods that involve both (1) many parameters and (2) many inequality restrictions. Examples of this literature include Chen, Christensen and Tamer (2018) and Kaido, Molinari and Stoye (2019). To keep the number of parameters small, we use an intentionally simplified state space and, even then, we place further parametric restrictions. This gives us models with as few as five parameters so that we can easily apply a multidimensional grid search to compute the required confidence regions. Our restriction to problems amenable to a grid search is obviously strong given the state of the literature, but it removes complex computational choices from the list of problems we need to tackle.

Table 6 summarizes the variables we use. The data on the 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 captures demand for ready-mix concrete. 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 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 r_i in our model.

Table 7 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 the coefficients on 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.[41] We conclude our brief descriptive analysis by noting that the number of plants exhibits substantial persistence over time in the data, with x_{t+1} being equal to x_t with probability around 0.90.

41. The results in Table 7, 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.

We now turn to the structural analysis. 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 a suitable oligopoly model in the context of the ready-mix concrete industry. We specify the single-period profit function as follows:

$$\pi_{it} = \begin{cases} \alpha_{x_{it}} w_{it} - \beta + u_{it} & \text{if was in at } t-1\text{, stays in at } t \\ \alpha_{x_{it}} w_{it} - \beta - \gamma + u_{it} & \text{if was out at } t-1\text{, enters at } t \\ 0 & \text{if is out at } t \end{cases}, \tag{29}$$

where w_{it} denotes construction employment (in thousands) in market i in year t, $\alpha_{x_{it}}$ is a coefficient that depends on the number x_{it} of active firms in market i at time t, β represents the intercept of the variable profit function as well as any fixed costs, γ is the sunk cost of entry, and u_{it} denotes a potentially serially correlated unobservable shock to profitability. 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. Further, we assume that, conditional on u_{it-1}, u_{it} is equal to u_{it-1} with probability ρ and is drawn uniformly from the $[-1,1]$ interval with probability $1-\rho$. 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 the existence and uniqueness of a Markov-perfect equilibrium in Last-In First-Out strategies if the model features one scalar exogenous state variable. In our setting, this corresponds to the case where w does not vary over time, so that the only time-varying exogenous state is u. We estimate this model by fixing w at its initial value in each market and assuming firms take it to be constant over time (which is not too far from what is observed in the data). We also estimate a model in which w varies over time. In this case, the uniqueness result in Abbring and Campbell (2010) does not directly apply. We will show that our two-step approach (which does not require uniqueness of the equilibrium) delivers virtually the same results as the full-solution method (which does require uniqueness), indicating that multiplicity of equilibria does not play an important role in our context.

Given the relatively long time dimension of the panel (11 years), the number of moments that could characterize the sharp identified set is very large, implying a substantial computational burden.[42] In addition, the elemental sets are associated with events that have very low probability individually, so that the left-hand side of (14) is small and the moment inequalities tend to be underpowered. Thus, we focus on a subset of inequalities corresponding to observable events that intuitively should help identify the structural parameters and happen with relatively high probability in the data. A complete list of moments is provided in Supplementary Appendix C.

We consider three types of sets, largely inspired by the core determining sets in our simple two-period entry example. First, we take the elemental sets corresponding to the cases where the number of firms is constant over time; these are the analogs to the events (0,0,0) and (1,1,1) from Figure 2 in the two-period example. Second, we add inequalities corresponding to cases in which the number of firms changes. Instead of listing all possible elemental sets in this category, we focus on “coarser” events that have higher probability. For example, we consider the event “some entry and some exit occur”, which is the extension to more than two periods (and more than one firm) of events like $(0,1,0)$ from Figure 2. To provide some more intuition, consider the event “no

42. The number of elemental sets is the number of possible outcomes (3) to the power of the number of time periods (11) times the number of possible combinations of w and z (4 in our simplest case), giving more than 700,000 sets before we consider the unions of the elemental sets that may also contribute to the Chesher and Rosen (2017) core-determining sets. Note that additional unions of sets may also be useful in inference (because their probability is well-measured in the data) even if they are unnecessary for identification.

entry or exit occurs in the 11-year period". If this occurs often in the data, it might help rule out values of serial correlation close to zero. Conversely, if the event "the number of firms changes at least once" has large enough probability, the corresponding inequalities should rule out values of serial correlation very close to 1. Third, we include inequalities corresponding to the events "there is at least one period with n firms, for $n=0,1,2$". These events happen frequently in the data, leading to relatively large probabilities on the left-hand side of (14) and thus aiding identification.

Each of the events described above leads to multiple inequalities, corresponding to the different values of the instruments (construction employment and past income growth) as well as x_0. Variation in the instruments is helpful for identification. If an event happens relatively often in the data conditional on a value of the instruments, it will tend to make the probability on the left-hand side of (14) large and reject parameter values that deliver smaller probabilities on the right-hand side. In particular, past income growth is our exogenous predictor of the initial conditions x_0 and therefore should help tackle the incompleteness of the model resulting from the endogeneity of x_0 (see Honoré and Tamer, 2006).

Even if we are only using inequalities that intuitively should aid identification, we are still left with a large number of them. In total, we obtain 266 (unconditional) inequalities, which is of the same order of magnitude as the number of markets in the data (428).[43] Thankfully, we can leverage recent developments in the econometrics literature to perform inference in this "many moment inequalities" setting. Specifically, we report (likely conservative)[44] projections of a confidence region for the multi-dimensional structural parameter obtained via the two-step multiplier bootstrap approach proposed by Chernozhukov *et al.* (2019) (henceforth, CCK). Note that the CCK procedure allows us to include a very large number of inequalities, which may alleviate some concerns about robustness to the inclusion or exclusion of particular inequalities. We also 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.

We estimate our GIV model in two ways. First, we use a traditional full-solution approach in which we solve the model for each candidate θ and verify whether the implied policy functions satisfy the GIV restrictions. This method requires a model with a unique equilibrium. However, as mentioned above, when we take w to be time-varying, our model does not satisfy the unique equilibrium conditions of Abbring and Campbell given that there are two exogenous state variables—w and u. Therefore, we also estimate a version of our two-step method, which does not require uniqueness of the equilibrium. Specifically, for each combination of policy thresholds and values of ρ in a grid, we check the GIV inequalities. If the candidate point passes the test, we then perform the second-step inversion to obtain the associated profit parameters, as described in Supplementary Appendix D. Finally, we check whether these unrestricted profit parameters satisfy the parametric restrictions in (29) (up to a tolerance). So, both the full-solution and the two-step method impose the same parametric restrictions on the profit function but leave the policy functions unrestricted. This implies that the two-step approach is strictly more flexible than the full-solution method in that it allows for multiplicity of equilibria under the same parametric assumptions. In terms of computation times, testing each candidate structural parameter using MATLAB on a 1.90 GHz processor takes 10.35 s in the full-solution approach. As a comparison, for each candidate parameter, the two-step method takes 8.12 s for the first step and 52.41 s for the

43. In the model where w is fixed, the sharp GIV inequalities corresponding to $T=2$ periods are easy to compute and we include those as well, yielding a total of 654 inequalities. Removing the $T=2$ inequalities does not substantially change the results.

44. The approach in Kaido *et al.* (2019) could be used to obtain confidence regions for the projections themselves. A recent paper by Bai, Santos and Shaikh (2022) provides another approach to testing many moment inequalities.

TABLE 8
Full-solution parameter estimates

		Full-solution GIV	MLE $\rho=0$	GMM $\rho=0$
w_{it} fixed	α_1	(0.16, 0.17)	(0.042, 0.045)	(0.10, 0.20)
	α_2	(0.001, 0.002)	(0.01, 0.02)	(−0.013, 0.015)
	β	(0.226, 0.232)	(0.038, 0.043)	(−0.04, −0.02)
	γ	(2.56, 2.57)	(1.44, 1.46)	(6.40, 10.56)
	ρ	(0.68, 0.70)	–	–
w_{it} varying	α_1	(0.05, 0.54)	(0.034, 0.055)	(−0.003, 0.071)
	α_2	(0, 0.21)	(0.004, 0.018)	(0.0008, 0.0259)
	β	(−0.23, 0.86)	(0.023, 0.063)	(−0.54, −0.10)
	γ	(0.43, 9.61)	(1.36, 1.50)	(4.21, 10.50)
	ρ	(0.34, 0.90)	–	–

Notes: The intervals are projections of a confidence region for the multi-dimensional structural parameter and have at least asymptotic 95% confidence level.

second step. In our model, the full-solution approach is relatively inexpensive since computing an equilibrium is fast. In other oligopoly models, this may not be true and the two-step method might yield substantial computational gains relative to the full-solution approach.

In addition, we estimate three models that set $\rho=0$ and thus assume away serial correlation. Two are full-solution methods corresponding to the MLE and GMM approaches used as benchmarks in the numerical illustration of Section 4. A third method mimics the standard two-step approaches with exogenous states. In particular, we use the same double parameterization as in the GIV two-step procedure described above, but we set $\rho=0$ and estimate the policy in the first step by MLE.

Tables 8 and 9 display confidence intervals for the structural parameters in specification (29) based on the full-solution and two-step approaches, respectively. First, both GIV approaches give estimates of ρ that are positive and significantly different than zero, indicating substantial persistence in the unobservables over time. Second, the full-solution GIV confidence intervals (which require uniqueness of the equilibrium) are virtually the same as the two-step GIV confidence intervals (which are robust to multiplicity). This is trivially the case when w is fixed over time since then the results in Abbring and Campbell (2010) guarantee uniqueness under mild conditions. The fact that the confidence intervals continue to be essentially the same when w varies over time suggests that multiplicity of equilibria is not a first-order concern in this setting.[45] Third, the approaches that rule out serial correlation in the unobservables give significantly different results in the case when w is fixed over time. In particular, while the MLE and two-step methods with $\rho=0$ are comparable, the GMM approach yields a much larger sunk cost γ and lower β. Under the null hypothesis that $\rho=0$, we would expect all three methods to give similar results and so we conclude that the data rejects the hypothesis in favour of serial correlation.

The model with w fixed yields very small confidence regions, which leads to possible concerns that the fixed w model is nearly rejected by the data. In contrast, in the model with w varying over time, the GIV confidence intervals are wider. This model is not close to being rejected by the data perhaps because it does a better job at capturing the true data generating process. Further, intuitively, when w varies over time, the conditioning variables in the GIV moments become "finer" and the moments tend to be estimated less precisely in our relatively small sample.

45. If we relaxed the tolerance used to impose the parametric restrictions in the second step, the two-step approach would yield larger confidence intervals. This points to a possible use of the two-step method as a way to assess robustness to the parametric assumptions of the model.

TABLE 9
Two-step parameter estimates

		Two-step GIV	Two-step $\rho=0$
w_{it} fixed	α_1	(0.16, 0.17)	(0.006, 0.010)
	α_2	(0.001, 0.002)	(−0.005, −0.002)
	β	(0.226, 0.232)	(−0.036, −0.030)
	γ	(2.56, 2.57)	(1.62, 1.68)
	ρ	(0.68, 0.70)	–
w_{it} varying	α_1	(0.05, 0.54)	(0.009, 0.192)
	α_2	(0, 0.21)	(−0.004, 0.009)
	β	(−0.23, 0.86)	(−0.032, −0.008)
	γ	(0.43, 9.61)	(1.49, 1.77)
	ρ	(0.34, 0.90)	–

Notes: The intervals are projections of a confidence region for the multi-dimensional structural parameter and have at least asymptotic 95% confidence level. The two-step GIV confidence intervals for the case with w_{it} fixed are the same as the full-solution intervals from Table 8 since the two models impose the same restrictions. The fact that the two-step GIV confidence intervals for the case with w_{it} varying are the same as the full-solution intervals from Table 8 suggests that multiplicity of equilibria does not play an important role in this application.

To investigate whether allowing for endogeneity of market structure makes a difference for policy-relevant questions, we turn to counterfactual analysis. As in the simulations of Section 4, we consider how the number and composition of firms in the market vary with two policy changes: (1) an increase in the sunk cost γ and (2) a subsidy to entry. The first can be thought of as arising from environmental regulation, such as a mandate for new firms to invest in technology to reduce polluted water from running off concrete operations, whereas the second could be motivated by the goal to incentivize entry of newer—and perhaps cleaner—firms. We assume that the policy change occurs at the end of the sample period and then compare the model predictions 5 years later to the model predictions in the absence of policy change. Repeating this for each value in the confidence region for the structural parameters yields confidence intervals for the counterfactual quantities of interest.[46]

Tables 10 and 11 show the counterfactual results based on the model with w fixed and w varying, respectively. In each case, we report only one set of GIV results given that the full-solution and two-step approaches yield virtually the same estimates and thus the same counterfactuals. In the case with w fixed (Table 10), all of the three methods that assume away serial correlation in the unobservables give results that are significantly different relative to the GIV method. In particular, maximum likelihood and two-step with $\rho=0$ tend to over-estimate the response to the counterfactual policy in terms of both the change in the number of firms and the decrease in the percentage of new firms after the policy change. 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. In contrast, the GMM approach with $\rho=0$ predicts much smaller changes due to the fact that it estimates a larger sunk cost to begin with.[47] Turning to the case with w varying (Table 11), we find broadly similar patterns. However, the GIV confidence intervals are now wider, consistent with the fact that the structural parameters are pinned down less precisely. This illustrates that the model specification matters.

46. For the first counterfactual, we increase the sunk cost by the same amount for each parameter value in the confidence set. We set this amount equal to 10% of the midpoint of the (projection of the) full-solution confidence region for γ.

47. The GMM confidence intervals for the baseline number of firms and fraction of new firms include negative values because the point estimates are close to zero and the standard errors are relatively large.

TABLE 10
Counterfactual outcomes with w fixed

	GIV	MLE $\rho=0$	GMM $\rho=0$	Two-step $\rho=0$
No. of firms				
Baseline	(0.65, 0.72)	(0.68, 0.77)	(−0.74, 1.52)	(0.88, 1.02)
↑ Sunk cost (Δ)	(−0.09, −0.05)	(−0.15, −0.14)	(−0.07, 0.07)	(−0.22, −0.15)
Entry subsidy (Δ)	(0.07, 0.08)	(0.49, 0.55)	(0.07, 0.29)	(0.53, 0.66)
% of new firms				
Baseline	(12.5, 14.5)	(31.3, 34.7)	(−3.1, 3.7)	(17.2, 22.2)
↑ Sunk cost (Δ)	(−3.5, −2.6)	(−14.6, −13.3)	(−2.1, 2.1)	(−7.1, −4.7)
Entry subsidy (Δ)	(10.2, 12.8)	(22.4, 24.8)	(−4.6, 4.6)	(22.2, 27.4)

Notes: For each outcome of interest (number of firms and percentage of new firms), the "baseline" numbers refer to the average outcomes in the absence of policy changes, whereas the next two rows report the change—relative to the baseline—in the counterfactual scenario where the sunk cost is higher and lower, respectively. All intervals have at least asymptotic 95% confidence level.

TABLE 11
Counterfactual outcomes with w varying

	GIV	MLE $\rho=0$	GMM $\rho=0$	Two-step $\rho=0$
No. of firms				
Baseline	(0.31, 1.56)	(0.73, 0.84)	(0.69, 3.06)	(0.87, 1.01)
↑ Sunk cost (Δ)	(−0.18, −0.003)	(−0.19, −0.16)	(−0.15, −0.03)	(−0.21, −0.14)
Entry subsidy (Δ)	(−0.05, 0.72)	(0.44, 0.55)	(−0.008, 0.22)	(0.52, 0.66)
% of new firms				
Baseline	(4.9, 46.9)	(29.3, 36.9)	(2.11, 9.25)	(17.9, 22.7)
↑ Sunk cost (Δ)	(−8.2, −0.1)	(−14.8, −10.8)	(−5.1, −0.2)	(−6.9, −5.0)
Entry subsidy (Δ)	(1.8, 37.6)	(20.5, 26.2)	(−1.7, 7.5)	(22.3, 26.9)

Notes: For each outcome of interest (number of firms and percentage of new firms), the "baseline" numbers refer to the average outcomes in the absence of policy changes, whereas the next two rows report the change—relative to the baseline—in the counterfactual scenario where the sunk cost is higher and lower, respectively. All intervals have at least asymptotic 95% confidence level.

In sum, these policy counterfactual results show that models which artificially set the serial correlation parameter to zero, as is common in much of the literature, may lead to substantial bias in counterfactual analyses.

7. CONCLUSION

In this article, we have proposed an approach to identification and inference in dynamic models with serially correlated unobservables. We tackle the resulting endogeneity of dynamic states by relying on the type of IVs 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. Our empirical application extends work by Collard-Wexler on dynamic entry models with serially correlated unobservables to consider policy counterfactuals that are motivated by classic questions in environmental economics. We find that approaches ignoring serial correlation can significantly misstate the effects of policies that affect the underlying profitability of an industry, such as an environmental regulation that affects the sunk cost of entry.

This article opens several avenues for future research. Most importantly, 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 additional counterfactual policy results.

Acknowledgments. Thanks to Allan Collard-Wexler for generously providing the data. For helpful comments we thank Xiaohong Chen, Liran Einav, Phil Haile, Francesca Molinari, Jesse Shapiro, Paulo Somaini, and the referees of this journal. 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. Mengsi Gao, Anshul Goel, and Jaehee Song provided excellent research assistance. All remaining errors are our own.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online. And the replication packages are available at https://dx.doi.org/10.5281/zenodo.6969758.

Data Availability Statement

The data and code underlying this article is available on Zenodo at https://dx.doi.org/10.5281/zenodo.6969758.

REFERENCES

ABBRING, J. H. and CAMPBELL, J. R. (2010), "Last-In First-Out Oligopoly Dynamics", *Econometrica*, **78**, 1491–1527.

ACKERBERG, D., BENKARD, L., BERRY, S. and PAKES, A. (2007), "Econometric Tools for Analyzing Market Outcomes", in Heckman, J. J. and Leamer, E. (eds) *Handbook of Econometrics*, vol. 6A, chap. 63 (Amsterdam: North-Holland).

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

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

ANDREWS, D. W. and SHI, X. (2013), "Inference Based on Conditional Moment Inequalities", *Econometrica*, **81**, 609–666.

——— (2017), "Inference Based on Many Conditional Moment Inequalities", *Journal of Econometrics*, **196**, 275–287.

ANDREWS, D. W. and SOARES, G. (2010), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection", *Econometrica*, **78**, 119–157.

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

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

ARTSTEIN, Z. (1983), "Distributions of Random Sets and Random Selections", *Israel Journal of Mathematics*, **46**, 313–324.

ATHEY, S. and IMBENS, G. (2006), "Identification and Inference in Nonlinear Difference-in-Difference Models", *Econometrica*, **74**, 431–497.

BAI, Y., SANTOS, A. and SHAIKH, A. M. (2022), "A Two-Step Method for Testing Many Moment Inequalities", *Journal of Business & Economic Statistics*, **40**, 1070–1080.

BAJARI, P., BENKARD, C. L. and LEVIN, J. (2007), "Estimating Dynamic Models of Imperfect Competition", *Econometrica*, **75**, 1331–1370.

BERESTEANU, A., MOLINARI, F. and MOLCHANOV, I. (2011), "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, **79**, 1785–1821.

BERRY, S. and COMPIANI, G. (2020a), "Empirical Models of Industry Dynamics with Endogenous Market Structure" (Technical Report, Yale, in preparation for the Annual Review of Economics).

——— (2020b), "An Instrumental Variable Approach to Dynamic Models" (Working Paper).

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

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

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

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

BLUNDELL, R. and BOND, S. (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, **87**, 115–143.

BORKOVSKY, R. N., DORASZELSKI, U. and KRYUKOV, Y. (2010), "A User's Guide to Solving Dynamic Stochastic Games Using the Homotopy Method", *Operations Research*, **58**, 1116–1132.

BRESNAHAN, T. (1981), "Departures from Marginal Cost Pricing in the American Automobile Industry", *Journal of Econometrics*, **17**, 201–227.

——— (1982), "The Oligopoly Solution Concept is Identified", *Economics Letters*, **10**, 82–97.

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

CHEN, X., CHRISTENSEN, T. M. and TAMER, E. (2018), "Monte Carlo Confidence Sets for Identified Sets", *Econometrica*, **86**, 1965–2018.
CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2019), "Inference on Causal and Structural Parameters using Many Moment Inequalities", *The Review of Economic Studies*, **86**, 1867–1900.
CHERNOZHUKOV, V., HONG, H. and TAMER, E. (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica*, **75**, 1243–1284.
CHERNOZHUKOV, V., Lee, S. and Rosen, A. (2013), "Intersection Bounds: Estimation and Inference", *Econometrica*, **81**, 667–737.
CHESHER, A. (2010), "Instrumental Variables Models for Discrete Outcomes", *Econometrica*, **78**, 575–601.
CHESHER, A. and ROSEN, A. (2017), "Generalized Instrumental Variable Models", *Econometrica*, **83**, 959–989.
CHESHER, A. and ROSEN, A. M. (2020), "Generalized Instrumental Variable Models, Methods, and Applications", in Durlauf, S. N., Hansen, L. P., Heckman, J. J. and Matzkin, R. L. (Eds) *Handbook of Econometrics*, vol. 7 (Amsterdam: Elsevier) 1–110.
CHESHER, A. and SMOLINSKI, K. (2012), "IV Models of Ordered Choice", *Journal of Econometrics*, **166**, 33–48.
CILIBERTO, F. and TAMER, E. (2009), "Market Structure and Multiple Equilibria in Airline Markets", *Econometrica*, **77**, 1791–1828.
COLLARD-WEXLER, A. (2013), "Demand Fluctuations in the Ready-Mix Concrete Industry", *Econometrica*, **81**, 1003–1037.
——— (2014), "Mergers and Sunk Costs: An Application to the Ready-Mix Concrete Industry", *American Economic Journal: Microeconomics*, **6**, 407–447.
COMPIANI, G. (2019), "Market Counterfactuals and the Specification of Multi-product Demand: A Nonparametric Approach" *Available at SSRN*.
DORASZELSKI, U. and PAKES, A. (2007), "A Framework for Applied Dynamic Analysis in I.O.", in Armstrong, M. and Porter, R. (Eds) *Handbook of Industrial Organization*, vol. 3 (Amsterdam: Elsevier) 1887–1966.
DORASZELSKI, U. and SATTERTHWAITE, M. (2010), "Computable Markov-Perfect Industry Dynamics", *The RAND Journal of Economics*, **41**, 215–243.
DUBÉ, J., HITSCH, G. and ROSSI, P. (2010), "State Dependence and Alternative Explanations for Consumer Inertia", *RAND Journal of Economics*, **41**, 417–445.
Ericson, R. and A. Pakes (1995), "Markov Perfect Industry Dynamics: A Framework for Empirical Work", *Review of Economic Studies*, **62**, 53–82.
FERSHTMAN, C. and PAKES, A. (2012), "Dynamic Games with Asymmetric Information: A Framework for Empirical Work", *The Quarterly Journal of Economics*, **127**, 1611–1661.
GALICHON, A. and HENRY, M. (2006), "Inference in Incomplete Models" *Available at SSRN 886907*.
——— (2011), "Set Identification in Models with Multiple Equilibria", *Review of Economic Studies*, **78**, 1264–1298.
HECKMAN, J. and SINGER, B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, **52**, 271–320.
HECKMAN, J. J., HUMPHRIES, J. E. and VERAMENDI, G. (2016), "Dynamic Treatment Effects", *Journal of Econometrics*, **191**, 276–292.
HONORÉ, B. and TAMER, E. (2006), "Bounds on Parameters in Panel Dynamic Discrete Choice Models", *Econometrica*, **3**, 611–629.
HOTZ, J. and MILLER, R. A. (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models", *Review of Economic Studies*, **60**, 497–529.
HOTZ, V. J., MILLER, R. A., SANDERS, S. and SMITH, J. (1994), "A Simulation Estimator for Dynamic Models of Discrete Choice", *The Review of Economic Studies*, **61**, 265–289.
HU, Y. and SHUM, M. (2012), "Nonparametric Identification of Dynamic Models with Unobserved State Variables", *Journal of Econometrics*, **171**, 32–44.
HU, Y., SHUM, M., TAN, W. and XIAO, R. (2015), "A Simple Estimator for Dynamic Models with Serially Correlated Unobservables" (Technical Report, Cal Tech, Working Paper).
IGAMI, M. (2018), "Industry Dynamics of Offshoring: The Case of Hard Disk Drives", *American Economic Journal: Microeconomics*, **10**, 67–101.
ISRAEL, M. (2005), "Tenure Dependence in Consumer-Firm Relationships: An Empirical Analysis of Consumer Departures from Automobile Insurance Firms", *RAND Journal of Economics*, **36**, 165–192.
KAIDO, H., MOLINARI, F. and STOYE, J. (2019), "Confidence Intervals for Projections of Partially Identified Parameters", *Econometrica*, **87**, 1397–1432.
KALOUPTSIDI, M., SCOTT, P. and SOUZA-RODRIGUES, E. (2021a), "Linear IV Regression Estimators for Structural Dynamic Discrete Choice Models", *Journal of Econometrics*, **222**, 778–804.
KALOUPTSIDI, M., SCOTT, P. T. and SOUZA-RODRIGUES, E. (2021b), "Identification of Counterfactuals in Dynamic Discrete Choice Models", *Quantitative Economics*, **12**, 351–403.
KASAHARA, H. and SHIMOTSU, K. (2009), "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices", *Econometrica*, **77**, 135–175.
KEANE, M. P. and WOLPIN, K. I. (1997), "The Career Decisions of Young Men", *Journal of Political Economy*, **105**, 473–522.
LAZAREV, J., BENKARD, L. and BODOH-CREED, A. (2018), "Simulating the Dynamic Effects of Horizontal Mergers: U.S. Airlines" (Working Paper, NYU).

MAGNAC, T. and THESMAR, D. (2002), "Identifying Dynamic Discrete Decision Processes", *Econometrica*, **70**, 801–816.

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

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

MENZEL, K. (2009), "Consistent Estimation With Many Moment Inequalities", *Journal of Econometrics*, **182**, 329–350.

MOLCHANOV, I. and MOLINARI, F. (2018), *Random Sets in Econometrics*, vol. 60 (Cambridge, UK: Cambridge University Press).

MOLINARI, F. (2020), "Microeconometrics with Partial Identification", *Handbook of econometrics*, **7**, 355–486.

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

PAKES, A. (1994), "Dynamic Structural Models, Problems and Prospects: Mixed Continuous Discrete Controls and Market Interactions", chap. 5 in Sims, C. (ed.) *Advances in Econometrics* (Cambridge: Cambridge University) 171–260.

PAKES, A. and MCGUIRE, P. (2001), "Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the "Curse" of Dimensionality", *Econometrica*, **69**, 1261–1281.

PAKES, A., OSTROVSKY, M. and BERRY, S. (2007), "Simple Estimators for the Parameters of Dynamic Games, with Entry/Exit Examples", *RAND Journal of Economics*, **38**, 373–399.

PAKES, A., PORTER, J., HO, K. and ISHII, J. (2015), "Moment Inequalities and Their Application", *Econometrica*, **83**, 315–334.

PESENDORFER, M. and SCHMIDT-DENGLER, P. (2008), "Asymptotic Least Squares Estimators for Dynamic Games", *Review of Economic Studies*, **75**, 901–928.

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

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

RUDIN, W. (1976), *Principles of Mathematical Analysis*, 3rd ed. (McGraw-Hill Education).

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

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

SOMAINI, P. (2015), "Identification in Auction Models with Interdependent Costs" (Technical Report, Stanford).

STOKEY, N. L., LUCAS, R. E. and PRESCOTT, E. C. (1989), *Recursive Methods in Economic Dynamics* (Cambridge MA: Harvard University).

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

TORGOVITSKY, A. (2019), "Nonparametric Inference on State Dependence in Unemployment", *Econometrica*, **87**, 1475–1505.