An Instrumental Variable Approach to Dynamic Models

work in progress, comments welcome

Steven Berry and Giovanni Compiani
Yale and UC Berkeley

Summer 2019
The **endogeneity of market structure** is a common theme in IO, but many recent approaches to firm and industry dynamics assume away serial correlation, resulting in econometrically exogenous market structure. We propose a **Generalized IV approach** to the identification/estimation of dynamic models with endogenous market structure.

### Some IO Examples

<table>
<thead>
<tr>
<th>State, $x_{it}$</th>
<th>Action, $a_{it}$</th>
<th>$A(x_{it})$</th>
<th>Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital</td>
<td>Investment</td>
<td>$\mathbb{R}^+$</td>
<td>$x_{it+1} = \lambda x_{it} + a_{it}$</td>
</tr>
<tr>
<td>Out/In</td>
<td>Entry/Exit</td>
<td>${0, 1}$</td>
<td>$x_{it+1} = a_{it}$</td>
</tr>
<tr>
<td># of Stores</td>
<td>+/- Stores</td>
<td>$\mathcal{I}^+$</td>
<td>$x_{it+1} = a_{it}$</td>
</tr>
<tr>
<td>Quality</td>
<td>R&amp;D</td>
<td>$\mathbb{R}^+$</td>
<td>$x_{it+1} \sim f(x_{it}, a_{it})$</td>
</tr>
</tbody>
</table>
Dynamics: Standard “CCP” Practice

- Much of the empirical IO literature models unobservables as
  - Private information
  - IID over time

- This simplifies estimation, as outlined in several influential papers
IID private shocks $\Rightarrow$ current state is *econometrically exogenous*.

Policy function:

$$a_{it} = \sigma(x_{it}, u_{it}), \quad x_{it} \perp u_{it}$$

State does not reflect any persistent unobserved factors, only accumulated past luck.
Standard Practice (cont’d)

\[ a_{it} = \sigma(x_{it}, u_{it}), \quad x_{it} \perp u_{it} \]

- This leads to two-step approaches à la Hotz-Miller
  1. obtain policy function via a (generalized) regression of \( a \) on \( x \)
  2. use this + Bellman’s equation to identify structural parameters
Problems with Exogenous States

- Unobservables (e.g. market profitability) are likely persistent
- This implies that the current state will likely be endogenous

\[ a_{it} = \sigma(x_{it}, u_{it}), \quad x_{it} \not\perp u_{it} \]

- Similar endogeneity problems have been tackled in the literature on static models using IVs
- Goal: extend this to dynamic models
Existing approaches to serial correlation

Full cites and discussion in the paper.

- **Discrete heterogeneity** (e.g. two “types”) in two-step methods. Helps particularly with the initial conditions problem and with computation.

- **Full solution methods** have to impose an initial condition model and have to pick an equilibrium in oligopoly models.

- Kalouptsidi, Scott, and Souza-Rodrigues (2018) have a similar IV logic in a special case model.

- Related large literatures include state dependence vs. persistent heterogeneity & initial conditions problems in panel data.

No existing systematic approach to identification, especially to flexible initial conditions, general models, multiple equilibria and general forms of serial correlation.
An IV approach

Essentially, we follow the classic two steps of the CCP literature:

1. Identify/Restrict policy
2. Use this + Bellman to identify/restrict structural parameters

The innovation is that in step 1, we use generalized IV methods, “GIV”, to allow for endogenous states. For identification, we rely on Chesher and Rosen (2017) and, by transitivity, on a much larger literature:

Accommodates models that

- incomplete
- partially identified

due to, e.g., discrete outcomes, initial conditions problem, multiple equilibria, “not strong enough” IVs.
IV Intuition

- Past exogenous variation will be correlated with current states
  - If Detroit was large and rich 50 years ago, it may have many Sears stores today
  - Past macro shocks may impact today’s market structure
  - Past regulatory regimes might have persistent effects

- In general, past exogenous cost and demand shifters are natural IVs
Generalized IV in a nutshell

For any $\sigma$, define

$$U(a, x, \sigma) \equiv \{u : a = \sigma(x, u)\}$$

i.e. the inverse image set according to $\sigma$, given $(a, x)$

Then, the identified set for the true $\sigma_0$ is given by all the functions $\sigma$ that, for some $\theta_u$, satisfy

$$\mathbb{P}\{U(a, x, \sigma) \subset S|z\} \leq \mathbb{P}\{U \in S; \theta_u\}$$

for all sets $S$ in an appropriate collection

The event on the rhs is a necessary condition for that on the lhs. In a complete discrete model, the $\leq$ become equalities that define MLE.
Monopolist entry example

- We start from a simple monopolist entry example
- We characterize the sharp ID set for the policy function via GIV when $T = 1, T = 2$
- This yields the ID set for the structural parameters
- We explore through simulations the effects of:
  - IV strength
  - Number of time periods in the data
  - Variation in covariates
Example: Single Firm “Entry / Exit”

A minimal model to think about dynamics. State is In/Out in the prior period

\[ a_{it} \in \{0, 1\} \quad x_{it} \in \{0, 1\} \]

Single-period profits \( \pi(a_{it}, x_{it}, w_{it}, \epsilon_{it}) \).

\[
\begin{align*}
\pi(0, x_{it}, w_{it}, \epsilon_{it}) &= 0 \\
\pi(1, 1, w_{it}, \epsilon_{it}) &= \bar{\pi}(w_{it}) - \epsilon_{it} \\
\pi(1, 0, w_{it}, \epsilon_{it}) &= \bar{\pi}(w_{it}) - \epsilon_{it} - \gamma
\end{align*}
\]

\[ \epsilon_{it} = \rho \epsilon_{i,t-1} + \nu_{it}, \quad \nu_{it} \sim N(0, 1). \]

Define \( u_{it} \) as the quantile of \( \epsilon_{it} \).

Without variation in \( w_{it} \), 3 structural parameters: \( \bar{\pi}, \gamma, \rho \).
Restrictions Using Only One Period

Chesher (2010), etc.

\[
\begin{array}{c|c|c}
\tau_0 & \tau_1 & u_{it} \\
\hline
0 & \tau_0 & \tau_1 & 1 \\
\end{array}
\]

<table>
<thead>
<tr>
<th>( a )</th>
<th>( x )</th>
<th>( \mathcal{U}(a, x, \sigma) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>( (0, \tau_1) )</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>( (0, \tau_0) )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>( (\tau_1, 1) )</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>( (\tau_0, 1) )</td>
</tr>
</tbody>
</table>

\[
\begin{array}{l|l|l}
S & \mathbb{P}(\mathcal{U}(a, x, \sigma) \subseteq S \mid z) & \leq \mathbb{P}(U \in S; \theta_u) \\
\hline
\mathcal{U}(1, 1) & \mathbb{P}((1, 1)\mid z) + \mathbb{P}((1, 0)\mid z) & \leq \tau_1 \\
\mathcal{U}(1, 0) & \mathbb{P}((1, 0)\mid z) & \leq \tau_0 \\
\mathcal{U}(0, 1) & \mathbb{P}((0, 1)\mid z) & \leq 1 - \tau_1 \\
\mathcal{U}(0, 0) & \mathbb{P}((0, 0)\mid z) + \mathbb{P}((0, 1)\mid z) & \leq 1 - \tau_0 \\
\end{array}
\]
Elemental Inverse Image Sets, $T = 2$

8 elemental sets $\mathcal{U}(a, x, \sigma)$, labeled $(x_1, a_1, a_2)$
Computing the Identified Set of Structural Parameters

True $\rho = 0.75$, varying IV strength, $T=2$ or $T=10$

(a) IV strength = 0.25  
(b) IV strength = 0.56

($\rho$ on the horizontal axis, sunk cost on the vertical)
We apply the literature on moment inequalities for inference. The literature on the “sharp identified set” is less useful for inference, as there are no strict rules for what moment conditions to use and there can be very many possibilities. Since as $T$ grows we have many inequalities relative to the sample size, we use Chernozhukov, Chetverikov, and Kato (2018). In Monte Carlos, we obtain confidence sets for structural parameters and counterfactual subsidy to entry. Inference method works well in the Monte Carlos.
Policies now depend on the serially correlated unobservables of all players.

With serial correlation, complete information is much easier than serially correlated private info, but we can add iid private information on top of serially correlated public (as in PPHI), which makes computation (and existence) easier.
Broad Idea of the Games Approach

In common with the full solution approach, looking for \( [a] \) policy functions and \( [b] \) dynamic parameters that satisfy the best response condition.

The key advance is that we can restrict the computation to those strategies that survive the GIV conditions. In a favorable case, this would be a small set.

A computed oligopoly example is in the appendix, more research is needed.
Illustrative Empirical Application

- Small-town ready-mix concrete example of Collard-Wexler (2014)
- That paper considers exog change in $N$, answered by the policy function itself, with an assumption on initial conditions that closes the model.
- We consider a policy, like the environmental question in the cement example of Ryan (2012), that changes sunk and/or fixed cost. Not answered by the observed policy function.
- We allow for an incomplete model, with no assumptions on initial conditions. Does it hurt us a lot?
Empirical Example

- The number of concrete firms is shifted by “construction demand,” proxied by local construction employment
- We use past income growth as IV
- We use the “Last-In-First-Out” model of Abbring and Campbell (2010)
  - Unique equilibrium (we hope)
  - Computationally equivalent to sequence of single-agent problems
Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>St.Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of plants</td>
<td>0.97</td>
<td>0.93</td>
<td>0.00</td>
<td>6.00</td>
</tr>
<tr>
<td>Construction Employment</td>
<td>519</td>
<td>819</td>
<td>3</td>
<td>17,772</td>
</tr>
<tr>
<td>Household Income Growth 1969-89</td>
<td>0.15</td>
<td>0.11</td>
<td>-0.16</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Quasi-First State Ordered Probit: $N$ of Plants

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Construction Employment</td>
<td>0.14**</td>
</tr>
<tr>
<td>Income Growth 1969-1989</td>
<td>0.22**</td>
</tr>
<tr>
<td>Likelihood-Ratio Test p-value</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Flow Profit Function
from an Abbring-Campbell example

\[
\pi_{it} = \begin{cases} 
\alpha x_{it} w_i - \beta + u_{it} & \text{if was in at } t - 1, \text{ stays in at } t \\
\alpha x_{it} w_i - \beta - \gamma + u_{it} & \text{if was out at } t - 1, \text{ enters at } t \\
0 & \text{if is out at } t 
\end{cases}
\]

(1)

It would be better if static profits were estimated from data on price & qty, leaving only sunk cost (\(\gamma\)), fixed cost (\(\beta\)), and the degree of serial correlation (\(\rho\)) for the dynamic estimation.
Summary statistics. 428 markets, 1994–2005

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>St.Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of plants</td>
<td>0.97</td>
<td>0.93</td>
<td>0.00</td>
<td>6.00</td>
</tr>
<tr>
<td>Construction Employment</td>
<td>519</td>
<td>819</td>
<td>3</td>
<td>17,772</td>
</tr>
<tr>
<td>Household Income Growth 1969-1989</td>
<td>0.15</td>
<td>0.11</td>
<td>-0.16</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Quasi-First Stage Ordered probit results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Construction Employment</td>
<td>0.14**</td>
<td></td>
</tr>
<tr>
<td>Income Growth 1969-1989</td>
<td>0.22**</td>
<td></td>
</tr>
<tr>
<td>Likelihood-Ratio Test p-value</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
GIV vs MLE with $\rho \equiv 0$

**Figure:** Autocorrelation (x-axis) and sunk cost (y-axis): Projection of 95% confidence sets
Counterfactual Change in Sunk Costs

after 5 years

<table>
<thead>
<tr>
<th></th>
<th>GIV</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in total # of firms</td>
<td>(-0.09, -0.03)</td>
<td>(-0.18, -0.15)</td>
</tr>
<tr>
<td>Change in fraction of new firms</td>
<td>(-0.05, -0.01)</td>
<td>(-0.15, -0.14)</td>
</tr>
</tbody>
</table>

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

In the results allowing serial correlation, few markets are near the margin of entry/exit and this is persistently true, so the policy has a relatively small effect on firm behavior. Incomplete model with GIV still leads to relatively precise estimates of counterfactual policy.
Conclusion

- Market structure in dynamic IO models should not be assumed to be exogenous
- GIV methods allow us to
  - deal with endogenous states
  - preserve the intuition of existing two-step methods
- IV Intuition: past exogenous demographics/regulation/shocks are correlated with today’s state


References II


References III


