"Feedback in Panel Data Models" by Gary Chamberlain: A Discussion

Manuel Arellano
CEMFI

Chamberlain Seminar

March 21, 2025

- Thank you to Dalia and Jon for organizing this session and the opportunity to present.

- I first met Gary Chamberlain in 1990 at the Canadian Econometric Study Group. It was the first time I crossed the Atlantic.
- We were in the same session. He presented "Efficiency Bounds for Semiparametric Regression" and I presented my "Another Look" IV panel paper with Olympia.
- Before that, he had already been an important influence on my work.
- As a PhD student in the early 1980s, interested in panel data methods, Gary's handbook chapter was a never-ending source of inspiration.

Introduction

- "Feedback in panel data models" is an important paper, which has had a substantial influence on the literature.
- It has been widely cited and its discussion of identification difficulties in models with sequential exogeneity has become part of the econometric folklore.
- It has the unusual virtue of remaining relevant and interesting more than 30 years after it was written. Even more so considering that the paper remained unpublished during the author's lifetime and had a limited circulation.

- The paper is concerned with small-T panel models with an error term that is mean-independent of current and past instruments but not to future ones.
- The model may be nonlinear in parameters and observables, but is linear in an unobserved individual effect.

Introduction (continued)

- A simple example is:

$$y_{it} = \beta x_{it} + \alpha_i + v_{it} \qquad E\left(v_{it} \mid x_i^t\right) = 0$$

where x_{it} could be $y_{i,t-1}$, an external variable, or both.

- An important example is a one-factor random-coefficient model:

$$y_{it} = \left(\beta + \gamma\alpha_i\right) x_{it} + \alpha_i + v_{it} \qquad E\left(v_{it} \mid x_i^t\right) = 0.$$

- Other examples are a model with interacted time and individual effects:

$$y_{it} = \beta x_{it} + \delta_t \alpha_i + v_{it} \qquad E\left(v_{it} \mid x_i^t\right) = 0,$$

and an exponential regression:

$$y_{it} = \exp\left(\beta x_{it} + \alpha_i\right) + v_{it} \qquad E\left(v_{it} \mid x_i^t\right) = 0.$$

- Gary's **general setup** is:

$$d_t\left(z_i, \theta\right) = g_t\left(w_i^t, \theta\right) \alpha_i + v_{it} \qquad E\left(v_{it} \mid w_i^t\right) = 0 \tag{1}$$

where $\{z_i\}_{i=1}^n$ is an iid sequence and $w_i^T = \left(w_{i1}, ..., w_{iT}\right)$ is a component of z_i.

- The paper is in **two parts**. The first obtains the efficiency bound for this model.
- The second shows that it is very difficult to achieve point identification when the individual effect has more than one component.

I. Semiparametric efficiency

- The paper first notes that model (1) is equivalent to the semiparametric model

$$E\left[d_t\left(z_i,\theta\right)-g_t\left(w_i^t,\theta\right)h\left(w_i^T\right)\mid w_i^t\right]=0\qquad\left(t=1,...,T\right)\tag{2}$$

  where $h\left(w_i^T\right)=E\left(\alpha_i\mid w_i^T\right)$ is an unknown function. Interest is in θ and $\phi=E\left(\alpha_i\right)$.

- As in Chamberlain (1992), the paper obtains the information matrix for (θ,ϕ) for the multinomial case, then shows that the same bound applies to a general distribution.

- Next, for $d_{\theta it}=d_t\left(z_i,\theta\right)$, $g_{\theta it}=g_t\left(w_i^t,\theta\right)$, use the **transformation**: subtract the t equation from the $(t+1)$ equation multiplied by $g_{\theta it}g_{\theta i,t+1}^{-1}$, so that α_i is eliminated:

$$\lambda_{\theta it}\equiv d_{\theta it}-g_{\theta it}g_{\theta i,t+1}^{-1}d_{\theta i,t+1}=v_{it}^{*}\qquad\left(t=1,...,T-1\right)\tag{3}$$

  where the transformed error $v_{it}^{*}=v_{it}-g_{\theta it}g_{\theta i,t+1}^{-1}v_{i,t+1}$ still satisfies $E\left(v_{it}^{*}\mid w_i^t\right)=0$.

- Then it is shown that the bound for θ based on (3) is identical to the bound from (2).

- The information on ϕ is in the last period: multiplying the T equation in (2) by $g_{\theta iT}^{-1}$:

$$E\left(g_{\theta iT}^{-1}d_{\theta iT}-\phi\mid w_i^T\right)=0,$$

  which suggests a consistent estimator of ϕ and can be used to obtain its bound.

Optimal instruments

- Gary displayed the optimal instruments for this problem. If $T = 2$, there is only one transformed equation and the information bound for θ is
$$J_{T-1} = E\left(\frac{b_{i,T-1}b'_{i,T-1}}{\omega_{i,T-1}}\right)$$
with $b_{i,T-1} = E\left(\partial\lambda_{\theta i,T-1}/\partial\theta \mid w_i^{T-1}\right)$, $\omega_{i,T-1} = E\left(\lambda^2_{\theta i,T-1} \mid w_i^{T-1}\right)$.
- The optimal instrument is $m_{i,T-1} = b_{i,T-1}/\omega_{i,T-1}$ in the sense that the unfeasible IVE that solves $\sum_{i=1}^{n} m_{i,T-1}\lambda_{\theta i,T-1} = 0$ has asymptotic variance J_{T-1}^{-1}.
- When $T > 2$, the total information would be the sum of the bounds for each t if v_{it}^* were conditionally serially uncorrelated, which is not the case in general.
- To address this problem, Gary proposed a **forward filter** that removes serial correlation while preserving the sequential moment conditions.
- In terms of the filtered errors $\widetilde{v}_{it}$, the bound is $J = \sum_{t=1}^{T-1} E\left(\widetilde{b}_{it}\widetilde{b}'_{it}/\widetilde{\omega}_{it}\right)$ where $\widetilde{b}_{it} = E\left(\partial\widetilde{\lambda}_{\theta it}/\partial\theta \mid w_i^t\right)$, $\widetilde{\omega}_{it} = E\left(\widetilde{\lambda}^2_{\theta it} \mid w_i^t\right)$.
- Thus, the optimal instruments are $\widetilde{m}_{it} = \widetilde{b}_{it}/\widetilde{\omega}_{it}$ in the sense that the IVE that solves $\sum_{i=1}^{n}\sum_{t=1}^{T-1}\widetilde{m}_{it}\widetilde{\lambda}_{\theta it} = 0$ has asymptotic variance J^{-1}.

Why care?

- The optimal instruments involve various unknown conditional expectation functions, so they cannot be used in practice without further elaboration.
- Yet they can provide guidance on how to construct meaningful estimators.
- A simple example is:

$$y_{it} = \theta x_{it} + \alpha_i + v_{it} \qquad E\left(v_{it} \mid w_i^t\right) = 0$$

  where $d_{\theta it} = y_{it} - \theta x_{it}$, $g_{\theta it} = 1$ and $\lambda_{\theta it} = \left(y_{it} - y_{i,t+1}\right) - \theta\left(x_{it} - x_{i,t+1}\right)$.

- What auxiliary model for optimal instruments does Arellano-Bond GMM use?
- One in which the v_{it} are homoskedastic and serially uncorrelated, so that $\widetilde{v}_{it}$ and $\widetilde{x}_{it}$ boil down to forward orthogonal deviations, $\widetilde{b}_{it} = E\left(\widetilde{x}_{it} \mid w_i^t\right)$ and $\widetilde{\omega}_{it}$ is constant. In addition, GMM replaces $E\left(\widetilde{x}_{it} \mid w_i^t\right)$ with linear projections of $\widetilde{x}_{it}$ on w_i^t.
- However, the optimal instrument perspective suggests other possibilities.
- For example, use the implied $E\left(\widetilde{x}_{it} \mid w_i^t\right)$ by a first-stage VAR model, to avoid the proliferation of first-stage coefficients typical of panel GMM (Arellano, 2016).
- In a parametric approach to feasible IV estimation (such as 2SLS) there are two levels of assumptions: substantive restrictions used in estimation, and auxiliary (first-stage) restrictions used in estimating the optimal instruments.

II. Identification difficulties

- Gary points out that in feedback models with vector individual effects it is not possible, in general, to eliminate the effects by means of a transformation similar to the scalar case, which suggests identification failure.
- To illustrate the problem consider the bivariate effect model:

$$d_{\theta it} = g_{1,\theta it}\alpha_{1i} + g_{2,\theta it}\alpha_{2i} + v_{it} \qquad E\left(v_{it} \mid w_i^t\right) = 0. \tag{4}$$

- Subtracting the t equation from the $(t+1)$ equation multiplied by $g_{1,\theta it}g_{1,\theta i,t+1}^{-1}$, α_{1i} is eliminated:

$$d_{\theta it}^{*} = g_{2,\theta it}^{*}\alpha_{2i} + v_{it}^{*} \tag{5}$$

  with transformed error $v_{it}^{*} = v_{it} - g_{1,\theta it}g_{1,\theta i,t+1}^{-1}v_{i,t+1}$ such that $E\left(v_{it}^{*} \mid w_i^t\right) = 0$.
- We might hope to apply sequentially the same transformation to (5), but we cannot because now $g_{2,\theta it}^{*}$ is a function of w_i^{t+1}.
- In effect, in the twice transformed equation

$$d_{\theta it}^{*} - g_{2,\theta it}^{*}g_{2,\theta i,t+1}^{*-1}d_{\theta i,t+1}^{*} = v_{it}^{**}$$

  where $v_{it}^{**} = v_{it}^{*} - g_{2,\theta it}^{*}g_{2,\theta i,t+1}^{*-1}v_{i,t+1}^{*}$, $E\left(v_{it}^{**} \mid w_i^t\right) \neq 0$ in general. The difficulty is that $g_{2,\theta i,t+1}^{*}$ is a function of w_i^{t+2}.
- Contrary to the scalar case, it is no longer possible to construct unbiased estimators of $E\left(\alpha_{ji}\right)$ given the knowledge of θ.

Identification difficulties (continued)

- The paper then formally shows lack of identification of θ in the model:

$$y_{it} = \theta' r_{it} + \alpha_{1i} + \alpha_{2i} x_{it} + v_{it} \qquad E\left(v_{it} \mid w_i^t\right) = 0$$

  where $w_{it}' = (r_{it}', x_{it})$.

- As an illustration consider the case where x_{it} is a $0-1$ binary variable.
- Since $E\left(\alpha_{1i} \mid w_i^T\right)$ is unrestricted, the only moments that are relevant for the identification of θ are

$$E\left(\Delta y_{i,t+1} - \theta' \Delta r_{i,t+1} \mid w_i^t\right) = E\left(\alpha_{2i} \Delta x_{i,t+1} \mid w_i^t\right),$$

  which are equivalent to

$$\begin{aligned} E\left(\Delta y_{i,t+1} - \theta' \Delta r_{i,t+1} \mid w_i^{t-1}, r_{it}, x_{it} = 0\right) &= E\left(\alpha_{2i} \mid w_i^{t-1}, r_{it}, x_{it} = 0\right) \\ &\quad \times \Pr\left(x_{i,t+1} = 1 \mid w_i^{t-1}, r_{it}, x_{it} = 0\right) \\ E\left(\Delta y_{i,t+1} - \theta' \Delta r_{i,t+1} \mid w_i^{t-1}, r_{it}, x_{it} = 1\right) &= -E\left(\alpha_{2i} \mid w_i^{t-1}, r_{it}, x_{it} = 1\right) \\ &\quad \times \Pr\left(x_{i,t+1} = 0 \mid w_i^{t-1}, r_{it}, x_{it} = 1\right). \end{aligned}$$

- If $E\left(\alpha_{2i} \mid w_i^{t-1}, r_{it}, x_{it} = 0\right)$ and $E\left(\alpha_{2i} \mid w_i^{t-1}, r_{it}, x_{it} = 1\right)$ are unrestricted, θ cannot be identified from those equations.

Where to go from here?

- On the surface, the paper's identification results are negative, but its insights into the nature of underidentification open the way to exploring the consequences of different patterns of dynamics and heterogeneity with an economic motivation.

- Let me give some examples that are close to the linear context of Gary's paper. The important topic of feedback in nonlinear panels will be the subject of Kevin's talk.

Euler equation models

- A leading motivation for the sequential conditional mean models in Chamberlain's work was the martingale implications of models with time-additive expected utility.
- In those situations the equation of interest is an equilibrium condition.
- Rather than seeking to make inference about a causal effect, the goal is to infer characteristics of the outcome itself in the knowledge that it satisfies certain conditional mean restrictions.
- From Gary's paper we learned that a consequence of unrestricted feedback is that increasing T does not increase the effective size of the measurement system to identify more than one latent factor.
- However, economics tell us that interdependent choices rest on the same primitives.
- For example, a system of Euler equations for multiple goods or multiple asset returns will all depend on the same latent factors capturing heterogeneity in discounting and risk aversion.
- E.g. we could identify bivariate heterogeneity in the bivariate system:

$$y_{jit} = \theta_j' r_{it} + \delta_j \alpha_{1i} + \gamma_j \alpha_{2i} x_{it} + v_{jit} \qquad E\left(v_{jit} \mid w_i^t\right) = 0 \qquad (j = 1, 2).$$

Timing matters

- In Gary's analysis –and in dynamic economics– timing aspects matter a great deal.

- For example, in a classic paper, Stanley Fischer (1977) argued for the effectiveness of monetary policy in the presence of multiperiod wage contracts.

- In this environment a firm's response in t to a wage negotiated in $t-1$ takes the form

$$y_{it} = \theta r_{it} + \alpha_{1i} + \alpha_{2i} x_{i,t-1} + v_{it} \qquad E\left(v_{it} \mid w_i^t\right) = 0$$

  where $x_{i,t-1}$ denotes a negotiated wage based on $t-1$ information contained in w_i^{t-1}.

- In this model Gary's arguments tell us that sequential application of his transformation will secure identification with bivariate heterogeneity.

- The opposite case is one in which agents operate with advanced information.

- The following example, brings together three different situations:

$$y_{it} = \beta_\theta(\alpha_i) x_{it} + \gamma_\theta(\alpha_i) + v_{it} \qquad E\left(v_{it} \mid x_i^{t-j}\right) = 0$$

  where $\beta_\theta(.)$,$\gamma_\theta(.)$ are linear functions of α_i and $\phi = E(\alpha_i)$.

- If $j = 0$ or $j = -1$, x_{it} is sequentially exogenous.

- If $j = 1$, x_{it} is sequentially endogenous.

- If $j = 0$, Gary's analysis tells us that (θ, ϕ) are point identified if $\dim(\alpha_i) = 1$ but not if $\dim(\alpha_i) \geq 2$.

- If $j = -1$, we can have identification with $\dim(\alpha_i) = 2$.

- The case $j = 1$ is not covered by Gary's results and identification there is of a different level of difficulty.

- In optimizing models of economic behavior the distinction between sequential exogeneity and sequential endogeneity is a fine line that depends on what agents know and not know at each point in time.

Conditional heterogeneity

- Gary's analysis was based on the restriction that $E\left(v_{it} \mid w_i^t\right) = 0$.
- Different conclusions emerge if one uses the stronger restriction $E\left(v_{it} \mid w_i^t, \alpha_i\right) = 0$.
- After all, permanent unobserved heterogeneity is often conditioned upon in economic models of choice.
- Ahn and Schmidt (1995) showed that the restriction $E\left(v_{it} \mid w_i^t, \alpha_i\right) = 0$ led to additional moments in autoregressive models.
- Also conditioning on individual effects, Wooyong Lee (2022) characterized sharp identified sets of the coefficient distributions in an autoregressive model with individual-specific slope and intercept coefficients, which are empirically useful.

Markovian feedback

- Gary's paper also allowed for an unrestricted feedback process. The situation is different if we place constraints on the feedback process.

- For example, we may wish to consider the identifying content of limited feedback.

- The idea is that, after accounting for individual effects, we can expect a stronger association between states that are close to each other in time than between states that are far apart.

- In fact, many economic models rely on Markovian or hidden Markovian properties.

- Using nonlinear deconvolution ideas, Arellano and Bonhomme (2016) showed the nonparametric identification of a model with general predetermined variables and non-scalar conditional effects when the feedback process is hidden Markovian.

Partial identification

- In a panel model with feedback, some objects of interest may be identified while others are not.

- An example is a random coefficients model with a predetermined binary regressor in which average effects for movers are identified (Arellano and Bonhomme, 2012):

$$y_{it} = \beta_i d_{it} + \alpha_i + v_{it} \qquad E\left(v_{it} \mid d_{it}, d_{it-1}, ...\right) = 0 \qquad t = 1, 2$$

$$\begin{aligned} E\left(\Delta y_{i2} \mid d_{i1} = 0\right) &= E\left(\beta_i \mid d_{i1} = 0, d_{i2} = 1\right) \Pr\left(d_{i2} = 1 \mid d_{i1} = 0\right) \\ E\left(\Delta y_{i2} \mid d_{i1} = 1\right) &= -E\left(\beta_i \mid d_{i1} = 1, d_{i2} = 0\right) \Pr\left(d_{i2} = 0 \mid d_{i1} = 1\right) \end{aligned}$$

- In this model $E\left(\beta_i \mid d_{i1} = 0, d_{i2} = 1\right)$ and $E\left(\beta_i \mid d_{i1} = 1, d_{i2} = 0\right)$ are identified but not $E\left(\beta_i\right)$.