Arellano–Bond GFIC Example

Francis J. DiTraglia

August 16, 2015

1 Arellano-Bond (1991) Simulation Setup

DGP – General Version Ten "burn-in" crossections discarded, $y_{i0} = 0$. Sample size $N = 100$ individuals, $T = 7$ time periods.

$$
\begin{aligned}
y_{it} &= \alpha y_{it-1} + \beta x_{it} + \eta_i + v_{it} \\
x_{it} &= \rho x_{it-1} + \epsilon_{it} \\
v_{it} &= \sigma_{it}(\xi_{it} + \phi \xi_{it-1}) \\
\sigma^2_{it} &= \theta_0 + \theta_1 x^2_{it} \\
\eta_i &\sim \text{iid N}(0, \sigma^2_\eta) \\
\xi_{it} &\sim \text{iid N}(0, 1) \\
\epsilon_{it} &\sim N(0, \sigma^2_\epsilon)
\end{aligned}
$$

DGP – Simplification (Table 1)

$$
\begin{aligned}
\theta_1 = \phi &= 0 \\
\theta_0 = \sigma^2 &= 1 \\
\sigma^2_\eta &= 1 \\
\beta &= 1 \\
\rho &= 0.8 \\
\sigma^2_\epsilon &= 0.9
\end{aligned}
$$

which gives

$$\begin{aligned}
y_{it} &= \alpha y_{it-1} + x_{it} + \eta_i + v_{it} \\
x_{it} &= 0.8 x_{it-1} + \epsilon_{it} \\
\eta_i &\sim \text{iid N}(0, \sigma^2_\eta) \\
v_{it} &\sim \text{iid N}(0, 1) \\
\epsilon_{it} &\sim N(0, 0.9)
\end{aligned}$$

Estimators Do they include constants? Does it make a difference?

2 My Simulation Setup

Data Generating Process

$$\begin{aligned}
y_{it} &= \alpha_1 y_{it-1} + \alpha_2 y_{it-2} + \beta x_{it} + \eta_i + v_{it} \\
x_{it} &= \theta \eta_i + \gamma v_{it-1} + \xi_{it} \\
\xi_{it} &= \rho \xi_{it-1} + \epsilon_{it}
\end{aligned}$$

Relationship to Arellano-Bond Simplified Simulation Set $\alpha_2 = \theta = \gamma = 0$, $\beta = 1$, and $\rho = 0.8$. Then we have:

$$\begin{aligned}
y_{it} &= \alpha y_{it-1} + x_{it} + \eta_i + v_{it} \\
x_{it} &= \xi_{it} \\
\xi_{it} &= 0.8 \xi_{it-1} + \epsilon_{it}
\end{aligned}$$

which is the same as the simulation setup in Table 1 of Arellano & Bond (1991). The only difference I can see is that in the paper they hold x_{it} fixed across replications whereas I do not.

Initial Conditions We set all pre-sample observations to zero and then generate $k + T$ time periods from the model. The first k are "burn-in" observations which are discarded. Specifically, we initialize as follows:

$$\xi_{i0} = v_{i0} = y_{i0} = y_{i,-1} = 0$$

$$\begin{aligned}
\xi_{i1} &= \rho\xi_{i0} + \epsilon_{i1} = \epsilon_{i1} \\
x_{i1} &= \theta\eta_i + \gamma v_{i0} + \xi_{i1} = \theta\eta_i + \xi_{i1} \\
y_{i1} &= \alpha_1 y_{i0} + \alpha_2 y_{i,-1} + \beta x_{i1} + \eta_i + v_{i1} = \beta x_{i1} + \eta_i + v_{i1}
\end{aligned}$$

$$\begin{aligned}
\xi_{i2} &= \rho\xi_{i1} + \epsilon_{i2} \\
x_{i2} &= \theta\eta_i + \gamma v_{i1} + \xi_{i2} \\
y_{i2} &= \alpha_1 y_{i1} + \alpha_2 y_{i0} + \beta x_{i2} + \eta_i + v_{i2} = \alpha_1 y_{i1} + \beta x_{i2} + \eta_i + v_{i2}
\end{aligned}$$

Error Terms The error terms ϵ_{it}, η_i and v_{it} are mutually independent, iid, mean zero and normally distributed with variances $\sigma^2_\epsilon, \sigma^2_\eta$ and σ^2_v, respectively. That is,

$$\begin{bmatrix} \boldsymbol{\epsilon}_i \\ \eta_i \\ \boldsymbol{v}_i \end{bmatrix} \overset{\text{iid}}{\sim} \text{N}\left(\begin{bmatrix} \mathbf{0}_{k+T} \\ 0 \\ \mathbf{0}_{k+T} \end{bmatrix}, \begin{bmatrix} \sigma^2_\epsilon \mathbf{I}_{k+T} & & 0 \\ & \sigma^2_\eta & \\ 0 & & \sigma^2_v \mathbf{I}_{k+T} \end{bmatrix} \right)$$

where $\mathbf{v}_i = (v_{i1}, \ldots, v_{i,k+T})'$, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{i,k+T})'$ and $\mathbf{I}_{k+T}$ is the $(k+T) \times (k+T)$ identity matrix.

Properties of x_{it} and ξ_{it}

$$\begin{aligned}
\xi_{it} &= \rho\xi_{it-1} + \epsilon_{it} \\
&= \rho(\rho\xi_{it-2} + \epsilon_{it-1}) + \epsilon_{it} \\
&\vdots \\
&= \epsilon_{it} + \rho\epsilon_{it-1} + \rho^2\epsilon_{it-2} + \ldots + \rho^{t-2}\epsilon_{i2} + \rho^{t-1}\epsilon_{i1}
\end{aligned}$$

$$\begin{aligned}
E[\xi_{it}\xi_{is}] &= E\left[\left(\epsilon_{it} + \ldots + \rho^{t-1}\epsilon_{i1}\right)\left(\epsilon_{is} + \ldots + \rho^{s-1}\epsilon_{i1}\right)\right] \\
&= \sigma^2_\epsilon \sum_{j=0}^{\min\{s,t\}-1} \rho^{|t-s|+2j} = \sigma^2_\epsilon \rho^{|t-s|} \sum_{j=0}^{\min\{s,t\}-1} \rho^{2j} \\
&= \sigma^2_\epsilon \rho^{|t-s|} \left(\frac{1-\rho^{\min\{s,t\}}}{1-\rho^2} \right)
\end{aligned}$$

$$E[x_{it}\eta_i] = E[(\theta\eta_i + \gamma v_{it-1} + \xi_{it})\eta_i] = \theta\sigma_\eta^2$$

$$\begin{aligned} E[x_{it}v_{is}] &= E[(\theta\eta_i + \gamma v_{it-1} + \xi_{it})v_{is}] = \gamma E[v_{it-1}v_{is}] \\ &= \begin{cases} \gamma\sigma_v^2 \text{ for } s = t-1 \\ 0 \text{ otherwise} \end{cases} \end{aligned}$$

$$\begin{aligned} E[x_{it}x_{is}] &= E[(\theta\eta_i + \gamma v_{it-1} + \xi_{it})(\theta\eta_i + \gamma v_{is-1} + \xi_{is})] \\ &= \theta^2\sigma_\eta^2 + \gamma E[v_{it-1}v_{is-1}] + E[\xi_{it}\xi_{is}] \end{aligned}$$

3 Estimators and Selection

Recall that the DGP is

$$y_{it} = \alpha_1 y_{it-1} + \alpha_2 y_{it-2} + \beta x_{it} + \eta_i + v_{it}$$

3.1 Target Parameters and Parameter Restrictions

To take a simplified example, suppose that y_{it} is log employment and x_{it} is log wage. Then β is the *short-run* wage elasticity and $\beta/(1-\alpha_1-\alpha_2)$ is the *long-run* wage elasticity.[1] The two elasticities coincide when there are no dynamics: $\alpha_1 = \alpha_2 = 0$. We consider two cases: one in which the short-run elasticity is the target parameter and one in which the long-run elasticity is the target parameter.

Should mention empirical examples in which long-run elasticities are important and some others where short-run elasticities are more interesting for a particular policy question.

Long-Run Effect In this case, our target parameter is

$$\mu = \frac{\beta}{1-\alpha_1-\alpha_2}$$

If we're interested in a long-run effect, this more or less assumes that we think there are dynamics. Hence, it doesn't make sense for us to consider restricting both α_1 and

[1] See Hamilton 1.1.16 and Proposition 1.3. For "long-run" it is equivalent to consider *either* the cumulative effect of a one-period increase in x_{it} *or* the long-run effect of a *permanent* increase in x_{it}.

α_2 to be zero. Instead we'll consider the restriction $\alpha_2 = 0$. This is the equivalent of $\gamma = \gamma_0$ from the theoretical derivations. In the notation of the theoretical portion of the paper, the derivatives that appear in the limiting distribution of $\widehat{\mu}$ (Corollary 3.2) are

$$\nabla\varphi_0 = \begin{bmatrix} \left.\dfrac{\partial\mu}{\partial\beta}\right|_{\alpha_2=0} \\ \\ \left.\dfrac{\partial\mu}{\partial\alpha_1}\right|_{\alpha_2=0} \\ \\ \left.\dfrac{\partial\mu}{\partial\alpha_2}\right|_{\alpha_2=0} \end{bmatrix} = \begin{bmatrix} 1/(1-\alpha_1) \\ \beta/(1-\alpha_1)^2 \\ \beta/(1-\alpha_1)^2 \end{bmatrix}$$

The element that corresponds to the γ block is $\beta/(1-\alpha_1)^2$. This is the component that multiples the bias parameter δ in the second term of Corollary 3.2.

Short-Run Effect In this case, our target parameter is simply β. If we're interested purely in the short-run effect, this doesn't commit us to any view about the importance of dynamics. We can consider restricting both α_1 and α_2 to be zero. In fact, in the simulation experiment it could make the most sense to only work with a model in which there is a single lag of y. In the specification for which α_1 and α_2 are assumed to be zero, we can simply use OLS! In this case, the target parameter doesn't depend on the nuisance parameters, so $\nabla\varphi_0 = (1,0,0)'$. The elements that correspond to the γ block are zeros.

3.2 Models

To remove the correlated individual effects, we estimate all models in first differences:

$$\Delta y_{it} = \alpha_1 \Delta y_{it-1} + \alpha_2 \Delta y_{it-2} + \beta \Delta x_{it} + \Delta v_{it}$$

As mentioned above, the question of which model specifications to consider depends on the target parameter we have in mind. But since both target parameters depend on β, the only models we consider correspond to restrictions on α_1 and α_2, namely: where L0 is shorthand for the model with zero lags of y, L1 for the model with one lag of y and L2 for the model with two lags of y. The corresponding model-implied residuals are $\Delta v_{it}^{(0)}$, $\Delta v_{it}^{(1)}$ and $\Delta v_{it}^{(2)}$.

Model	Restriction	Model-Implied Residual
L0	$\alpha_1 = \alpha_2 = 0$	$\Delta v_{it}^{(0)} = \Delta y_{it} - \beta \Delta x_{it}$
L1	$\alpha_2 = 0$	$\Delta v_{it}^{(1)} = \Delta y_{it} - \alpha_1 \Delta y_{it-1} - \beta \Delta x_{it}$
L2	None	$\Delta v_{it}^{(2)} = \Delta y_{it} - \alpha_1 \Delta y_{it-1} - \alpha_2 \Delta y_{it-2} - \beta \Delta x_{it}$

3.3 Moment Conditions

The regressor x_{it} is predetermined with respect to the error term v_{it}. Since $x_{it} = \theta\eta_i + \gamma v_{it-1} + \xi_{it}$, however, it is only strictly exogenous when $\gamma = 0$. The moment selection problem is whether or not to use the additional moment conditions that arise from the assumption that x is strictly exogenous. The choice of moment conditions only makes sense after we've specified a model, so we examine L2, L1 and L0 in turn. We will only consider linear GMM estimators in the style of Arellano and Bond (1991).

Moment Selection for L2 This is the "true" model in that it is correctly specified regardless of the values of α_1 and α_2. The model-implied residual is

$$\Delta v_{it}^{(2)} = \Delta y_{it} - \alpha_1 \Delta y_{it-1} - \alpha_2 \Delta y_{it-2} - \beta \Delta x_{it}$$

Now the question is which moment conditions to use in estimation. Since this model is correctly specified, the model-implied residual equals the true differenced error term: $\Delta v_{it}^{(2)} = \Delta v_{it} = v_{it} - v_{it-1}$. Hence $y_{i1}, \ldots, y_{it-2}$ are valid instruments for time period t. If x is predetermined, then $x_{i1}, \ldots, x_{it-1}$ are valid instruments for period t. Under the stronger assumption that x is strictly exogenous, which requires $\gamma = 0$, $x_{i1}, \ldots, x_{iT}$ are valid instruments for period t.

Now, if $t = 1, \ldots, T$ then we can form the residuals $\Delta v_{it}^{(2)}$ for periods $4, \ldots, T$. We lose one time period from taking first differences, another from estimating including the first lag of y, and a third period from including the second lag of y.

Use the optimal instruments or the Arellano-Bond style instruments? Possibility of weak instrument problems. Need second step anyway to get the J-test. Implement parts of the estimator as sums?

Moment Selection for L1 This model is only correctly specified when the restriction $\alpha_2 = 0$ holds. The model-implied residual is

$$\Delta v_{it}^{(1)} = \Delta y_{it} - \alpha_1 \Delta y_{it-1} - \beta \Delta x_{it}$$

Now the question is which moment conditions to use in estimation. Suppose the restriction $\alpha_2 = 0$. Then the model-implied residual equals the true differenced error term: $\Delta v_{it}^{(1)} = \Delta v_{it} = v_{it} - v_{it-1}$.

The same points about instruments as discussed for L2 hold for L1 only we get more time periods.

If $t = 1, \ldots, T$ then we can form the residuals $\Delta v_{it}^{(1)}$ for periods $3, \ldots, T$. We lose one time period from taking first differences and another from including a lagged value of y. Hence, L1 gives us one more time period to use in estimation than L2.

Moment Selection for L0 This model is only correctly specified when the restriction $\alpha_1 = \alpha_2 = 0$ holds. The model-implied residual is

$$\Delta v_{it}^{(0)} = \Delta y_{it} - \beta \Delta x_{it}$$

Now the question is which moment conditions to use in estimation. Suppose that the restriction $\alpha_1 = \alpha_2 = 0$ holds. Then the model-implied residual equals the true, differenced error term: $\Delta v_{it}^{(0)} = \Delta v_{it} = v_{it} - v_{it-1}$.

For L0 we wouldn't use any of the y as instruments. I think in the case of strictly exogenous we can just do OLS? Doesn't the weight matrix cancel since this is just identified?

If $t = 1, \ldots, T$ then we can form the residuals $\Delta v_{it}^{(0)}$ for periods $2, \ldots, T$. We lose one time period from taking first differences. Hence, L0 gives us two more time periods than L2 and one more time period than L1.

OLD

The error terms are jointly normal with mean zero. Variances and covariances are given as follows:

$$
\begin{aligned}
E[\xi_{it}] &= E[v_{it}] = E[\eta_i] = 0 \\
Var(\epsilon) &= \sigma_\xi^2, \quad Var(\eta_i) = \sigma_\eta^2, \quad Var(v_{it}) = \sigma_v^2 \\
E[v_{it}\eta_i] &= 0 \\
E[v_{it}\xi_{is}] &= \begin{cases} \sigma_{v\xi} \text{ for } s = t+1 \\ 0 \text{ otherwise} \end{cases} \\
E[\xi_{it}\eta_i] &= \sigma_{\eta\xi}
\end{aligned}
$$

Let $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{i,k+T})'$ and $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{i,k+T})'$. Then, we have

$$
\begin{bmatrix} \boldsymbol{\xi}_i \\ \eta_i \\ \boldsymbol{v}_i \end{bmatrix} \overset{\text{iid}}{\sim} \mathrm{N}\left(\begin{bmatrix} \mathbf{0}_{k+T} \\ 0 \\ \mathbf{0}_{k+T} \end{bmatrix}, \begin{bmatrix} \sigma_\xi^2 \mathbf{I}_{k+T} & \sigma_{\eta\xi}\boldsymbol{\iota}_{k+T} & \sigma_{v\xi}\Gamma_{k+T} \\ \sigma_{\eta\xi}\boldsymbol{\iota}'_{k+T} & \sigma_\eta^2 & \mathbf{0}'_{k+T} \\ \sigma_{v\xi}\Gamma'_{k+T} & \mathbf{0}_{k+T} & \sigma_v^2 \mathbf{I}_{k+T} \end{bmatrix} \right)
$$

where $\mathbf{0}_m$ is an $m \times 1$ vector of zeros, $\boldsymbol{\iota}_m$ is an $m \times 1$ vector of ones, $\mathbf{I}_m$ is the $m \times m$ identity matrix and Γ_m is an $m \times m$ matrix with ones on the subdiagonal and zeros elsewhere, namely

$$
\Gamma_m = \begin{bmatrix} \mathbf{0}'_{m-1} & 0 \\ \mathbf{I}_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}
$$

Properties of Exogenous Regressor We want to parameterize everything in terms of x, so we first need to see how the underlying parameters of the DGP determine $E[x_{it}\eta_i]$ and $E[x_{it}v_{is}]$. Since x_{it} follows an AR(1) process with $x_{i0} = 0$,

$$
\begin{aligned}
x_{it} &= \rho x_{it-1} + \xi_{it} \\
&= \rho(\rho x_{it-2} + \xi_{it-1}) + \xi_{it} \\
&\vdots \\
&= \xi_{it} + \rho\xi_{it-1} + \ldots + \rho^{?}\xi_{i2} + \rho^{?}\xi_{i1}
\end{aligned}
$$