Linear Panels and Random Coefficients

Manuel Arellano

CEMFI

September 2013

Introduction

- Panel data models with fixed effects play an important role in applied econometrics.
- In the linear case several estimation methods are available (within groups, IV & GMM, likelihood methods...).
- Applications of these methods are widespread.
- The purpose of these lectures is to provide an overview of the literature on nonlinear panel data methods, including some emphasis on bias-reduction approaches.
- To set the stage, we begin with a review of some basic concepts of linear panels and random coefficients models.
- The focus is on microeconometrics: individuals, households, and firms, but also cross-country growth and development studies.
- Business cycle and financial volatility studies that relate to time series panels and factor models are out of scope here.

Linear panels and random coefficients

- Basic motivation in microeconometrics: Identifying models that cannot be identified on single outcome data. Two leading situations:
  - Fixed effects endogeneity (e.g. productivity analysis, price effects in demand models, wage effects in labor supply).
  - Error components, variance decomposition (e.g. inequality, mobility studies, quality-adjusted price indices).

GMM perspective

- The generalized method of moments has proved very useful for linear panel models as an organizing principle.

General idea:

- Start from a set of moment conditions suggested by the model.
- Use sample counterpart to get estimates of common parameters.
- Invoke a central limit theorem to approximate the distribution of standardized estimates by a normal distribution.
- If more moments than parameters are available, form linear combinations.

Leading example: within-groups

$$y_{it} = x_{it}'\theta_0 + \alpha_i + v_{it} \qquad E\left(v_{it} \mid x_{i1}, ..., x_{iT}, \alpha_i\right) = 0.$$

- In this model x_{it} may be correlated with α_i but not with v_{is} for all t, s. We say that x_{it} is endogenous wrt the fixed effect but strictly exogenous wrt the time-varying error.
- Letting $\widetilde{x}_{it} = x_{it} - \overline{x}_i$, the WG model implies the moment conditions

$$E\left[\sum_{t=1}^{T} \widetilde{x}_{it}\left(\widetilde{y}_{it} - \widetilde{x}_{it}'\theta_0\right)\right] = 0.$$

- The WG estimator $\widehat{\theta}_{WG}$ solves the sample moments

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{x}_{it}\left(\widetilde{y}_{it} - \widetilde{x}_{it}'\widehat{\theta}_{WG}\right) = 0.$$

Leading example: within-groups (continued)

- Inference can be based on the large N, fixed T approximation:

$$\widehat{V}^{-1/2}\left(\widehat{\theta}_{WG}-\theta_0\right)\approx\mathcal{N}(0,I)$$

where

$$\widehat{V}=H^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{v}_{it}\widehat{v}_{is}\widetilde{x}_{it}\widetilde{x}_{is}'\right)H^{-1},$$

$\widehat{v}_{it}=\widetilde{y}_{it}-\widetilde{x}_{it}'\widehat{\theta}_{WG}$, and $H=\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'$.

- The resulting "cluster-robust" standard errors are robust to heteroskedasticity and serial correlation but rely on cross-sectional independence.

Cluster-robust bootstrap standard errors

- A bootstrap approach is as follows. Let $W_i = \left(y_{i1}, x_{i1}', ..., y_{iT}, x_{iT}'\right)'$ and regard $W_1, ..., W_N$ as a multivariate random sample of size N according to some cdf F.
- The WG estimator is a function of the data $\widehat{\theta}_{WG} = h\left(W_1, ..., W_N\right)$ whose distribution we want to estimate

$$\Pr\left(\widehat{\theta}_{WG} \leq r\right) = \Pr_F\left[h\left(W_1, ..., W_N\right) \leq r\right].$$

- A simple candidate is the plug-in estimator. It replaces F by the empirical cdf $\widehat{F}_N$:

$$\widehat{F}_N\left(s\right) = \frac{1}{N}\sum_{i=1}^{N} 1\left(W_i \leq s\right),$$

  which assigns probability $1/N$ to each of the observed values $w_1, ..., w_N$ of $W_1, ..., W_N$
- Letting $W_1^*, ..., W_N^*$ denote a random sample from $\widehat{F}_N$, the resulting estimator is then

$$\Pr_{\widehat{F}_N}\left[h\left(W_1^*, ..., W_N^*\right) \leq r\right], \tag{1}$$

  which is conceptually simple but prohibitive to calculate.
- The bootrstap method evaluates (1) by simulation. M of samples $W_1^*, ..., W_N^*$ (the bootstrap samples) are drawn from $\widehat{F}_N$, and the frequency with which

$$h\left(W_1^*, ..., W_N^*\right) \leq r$$

  provides the desired approximation to the estimator (1).

Cluster-robust bootstrap standard errors (continued)

- As a result of resampling we have available M estimates from the artificial samples: $\widehat{\theta}_{WG}^{(1)}, ..., \widehat{\theta}_{WG}^{(M)}$.
- A bootstrap standard error is then obtained as

$$\left[\frac{1}{M-1} \sum_{m=1}^{M} \left(\widehat{\theta}_{WG}^{(m)} - \overline{\widehat{\theta}_{WG}} \right)^2 \right]^{1/2}$$

  where $\overline{\widehat{\theta}_{WG}} = \sum_{m=1}^{M} \widehat{\theta}_{WG}^{(m)} / M$.
- The bootstrap method is very flexible and applicable to many different situations such as the bias and variance of an estimator, the calculation of confidence intervals, etc.
- Under general regularity conditions, using the bootstrap standard error to construct test statistics has the same asymptotic justification as conventional asymptotic procedures.
- Sometimes a data producer will provide users with *replicate weights*, which enable the estimation of the sampling distribution of estimators from complex sample designs without disclosing confidential information.

Generalizations

Improved GMM under heteroskedasticity and autocorrelation of unknown form

- Improved GMM based on the larger set of moments $E\left[x_i\left(\widetilde{y}_{it}-\widetilde{x}_{it}'\theta_0\right)\right]=0$, $(t=1,...,T)$ or

$$E\left[x_i\left(\Delta y_{it}-\Delta x_{it}'\theta_0\right)\right]=0,(t=2,...,T)$$

  where x_i stacks $x_{i1},...,x_{iT}$.

Instrumental variable fixed effects models

- IV versions where the starting assumption is

$$E\left(v_{it}\mid z_{i1},...,z_{iT},\alpha_i\right)=0$$

  for some strictly exogenous instrument z (e.g. tax component of price variation).
- The moments become

$$E\left[z_i\left(\widetilde{y}_{it}-\widetilde{x}_{it}'\theta_0\right)\right]=0.$$

- In this case x is treated as a strictly endogenous variable.

Generalizations (continued)

Testing for correlated effects

- If x is uncorrelated with α, valid moments are $E\left[x_i\left(y_{it}-x_{it}'\theta_0\right)\right]=0$, $(t=1,...,T)$, which include $E\left[x_i\left(\Delta y_{it}-\Delta x_{it}'\theta_0\right)\right]=0$, $(t=2,...,T)$ as a subset.
- Thus, an incremental Sargan test can be used for testing the null of fixed-effects exogeneity (Hausman type testing).

Models with both time-invariant and time-varying variables

- A model with a FE-exogenous time-invariant regressor w satisfies the moments:

$$\begin{aligned} E\left[x_i\left(\widetilde{y}_{it}-\widetilde{x}_{it}'\theta_0\right)\right] &= 0 \\ E\left[w_i\left(\overline{y}_i-\overline{x}_i'\theta_0-w_i\delta_0\right)\right] &= 0. \end{aligned}$$

- In an IV version the second moment would specify the orthogonality between the average error and an external time-invariant instrument.

Error in variables

- In a measurement error version of the WG model where x is measured with an iid error, valid moments are

$$E\left[\left(x_{i1},...,x_{i(t-2)},x_{i(t+1)},...,x_{iT}\right)\left(\Delta y_{it}-\Delta x_{it}'\theta_0\right)\right]=0 \qquad (t=2,...,T).$$

- Instruments are relevant as long as there is persistence in latent x's.
- If ignored first differencing may exacerbate measurement error bias as illustrated next.

- In a linear regression $y=\beta x^*+u$ with classical measurement error $x=x^*+\varepsilon$ where u, x^*, ε are mutually independent, the OLS parameter satisfies

$$\frac{Cov\left(y,x\right)}{Var\left(x\right)}=\frac{Cov\left(y,x^*\right)}{Var\left(x^*\right)+Var\left(\varepsilon\right)}=\frac{\beta}{1+\lambda}$$

  where $\lambda = Var\left(\varepsilon\right)/Var\left(x^*\right)$.
- Similarly, letting $\lambda_\Delta = Var\left(\Delta\varepsilon\right)/Var\left(\Delta x^*\right)$, the OLS parameter of the regression in differences satisfies

$$\frac{Cov\left(\Delta y,\Delta x\right)}{Var\left(\Delta x\right)}=\frac{\beta}{1+\lambda_\Delta}.$$

- If $Cov\left(\varepsilon_t,\varepsilon_{t-1}\right)=0$ but $Cov\left(x_t^*,x_{t-1}^*\right)>0$ then $\lambda_\Delta>\lambda$. Under these conditions, which are relevant in applications, differencing magnifies measurement error bias.

Illustration: measuring economies of scale in firm money demand

- Bover and Watson (2005) estimate firm-level money demand equations of the form

$$\log m_{it} = c(t) \log s_{it} + b(t) + \eta_i + v_{it}.$$

  where m is demand for cash and s denotes output (or sales).
- The economies of scale coefficient $c(t)$ is specified as a polynomial in t to allow for changes over the sample period.
- The year dummies $b(t)$ capture changes in relative interest rates together with other aggregate effects.
- The individual effect is meant to represent permanent differences across firms in the production of transaction services (so that η varies inversely with the firm's financial sophistication), and v contains measurement errors in cash holdings and sales.
- We would expect $Cov(\log s, \eta) \leq 0$ and a downward unobserved heterogeneity bias in economies of scale.
- We also expect measurement error to account for a larger share of variation in sales growth than in the level of sales.

Firm money demand estimates
Sample period 1986–1996

	OLS Levels	OLS WG	OLS 1st-diff.	GMM 1st-diff.	GMM 1st-diff. m. error	GMM Levels m. error
Log sales	.72	.56	.45	.49	.99	.75
	(30.)	(16.)	(12.)	(16.)	(7.5)	(35.)
Log sales	−.02	−.03	−.03	−.03	−.03	−.03
×trend	(3.2)	(9.7)	(4.9)	(5.3)	(5.0)	(4.0)
Log sales	.001	.002	.001	.001	.001	.001
$\times$trend^2	(1.2)	(6.6)	(1.9)	(2.0)	(2.3)	(1.4)
Sargan				.12	.39	.00
(p-value)						

All estimates include year dummies, and those in levels also include industry dummies. t-ratios in brackets robust to heteroskedasticity & serial correlation. N=5649. Source: Bover and Watson (2005).

All estimates in the table are obtained from an unbalanced panel of 5649 Spanish firms with at least four consecutive annual observations during the period 1986−1996.

- The comparison between OLS-levels and WG (cols 1 & 2) is consistent with a positive fixed-effects bias (counter to expectation), but the smaller OLS-diff sales effect (col 3) suggests that measurement error bias may be important.
- Col 4 shows GMM estimates based on the moments $E\left(\log s_{it}\Delta v_{is}\right)=0$ for all t,s. Absent measurement error, we would expect them to be similar to WG and OLS-diff.
- Col 5 shows GMM estimates based on

$$E\left(\log s_{it}\Delta v_{is}\right)=0\ \left(t=1,...,s-2,s+1,..,T;s=1,...,T\right),$$

  thus allowing for both correlated firm effects and measurement error in sales.
- Interestingly, now the leading sales coefficient is much higher and close to unity, and the Sargan test has a p-value close to 40 per cent.
- Finally, col 6 shows GMM estimates based on

$$E\left(\log s_{it}v_{is}\right)=0\ \left(t=1,...,s-1,s+1,..,T;s=1,...,T\right),$$

  which allow for measurement error in sales but not for correlated effects. The leading sales effect in this case is close to OLS in levels, suggesting that in levels the measurement error bias is not as important as in differences.

Conclusion

- What is interesting about this example is that a comparison between estimates in levels and deviations without consideration of measurement error (e.g. restricted to compare cols 1 & 2, or 1 & 3, as in Hausman-type testing), would lead to the conclusion of correlated effects, but with biases going in entirely the wrong direction.

Predeterminedness and dynamics

Time patterns

- The previous examples include fixed effects but do not allow for time patterns in the dependence between x and time-varying errors.
- However, the time dimension makes it possible to go beyond the cross-sectional notions of strict exogeneity and strict endogeneity, whereby the time series of a regressor is either fully independent or fully dependent of the time series of errors.
- Thus, x may depend on past v's but not on future v's (predeterminedness), or on v's that are close in time but not on v's from distant periods.
- A linear model with general predetermined variables replaces the strict exogeneity assumption $E\left(v_{it} \mid x_{i1}, ..., x_{iT}, \alpha_i\right) = 0$ with the sequential conditioning assumption

$$E\left(v_{it} \mid x_{i1}, ..., x_{it}, \alpha_i\right) = 0.$$

  Letting $x_i^t = (x_{i1}, ..., x_{it})$, such model implies the moments:

$$E\left[x_i^{t-1}\left(\Delta y_{it} - \Delta x_{it}'\theta_0\right)\right] = 0.$$

- This notion can be generalized to external instruments and to alternative patterns of leads or lags.
- An example is the relationship between the presence of small children at home and female labor supply. Treating children as strictly exogenous in this context is a much more restrictive assumption than treating them as predetermined.

First-stage and second-stage regressions in panel GMM

- In Arellano-Bond GMM estimation there is a sequence of period-by-period first-stage regressions and a pooled second-stage regression.
- Letting for simplicity $T = 3$ and a single predetermined regressor, the period-by-period first-stage fitted values are

$$\begin{aligned} \widehat{\Delta x_{i2}} &= \widehat{\pi}_{21} x_{i1} \\ \widehat{\Delta x_{i3}} &= \widehat{\pi}_{31} x_{i1} + \widehat{\pi}_{32} x_{i2} \end{aligned}$$

  where $\widehat{\pi}_{21}$ is the cross-sectional OLS coefficient of Δx_{i2} on x_{i1}, etc. (in practice, orthogonal deviations are preferred to first-differences but the idea is the same).

- The second-stage is a pooled IV regression of $(\Delta y_{i2}, \Delta y_{i3})$ on $(\Delta x_{i2}, \Delta x_{i3})$ using $\left(\widehat{\Delta x_{i2}}, \widehat{\Delta x_{i3}}\right)$ as instruments.
- The latter is very different to the time-series perspective where instruments would come from a pooled first-stage regression:

$$\begin{pmatrix} \widetilde{\Delta x_{i2}} \\ \widetilde{\Delta x_{i3}} \end{pmatrix} = \widetilde{\pi} \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}$$

  where $\widetilde{\pi}$ is the pooled OLS coefficient of $(\Delta x_{i2}, \Delta x_{i3})$ on (x_{i1}, x_{i2}). The 2nd-stage would be pooled IV of $(\Delta y_{i2}, \Delta y_{i3})$ on $(\Delta x_{i2}, \Delta x_{i3})$ using $\left(\widetilde{\Delta x_{i2}}, \widetilde{\Delta x_{i3}}\right)$ as instruments.
- In a pooled first-stage regression one cannot easily project on different x's at different periods as one does using period-by-period first stage regressions.

Dynamic models

- Time patterns of dependence arise naturally in the context of dynamic models. These are models that consider the effects of lagged outcomes and/or lagged and current independent explanatory variables on current outcomes.
- The simplest example is an autoregressive model, which is a special case of the above with $x_{it} = y_{i(t-1)}$.
- The basic moments are:

$$E\left[y_i^{t-2}\left(\Delta y_{it} - \Delta y_{i(t-1)}\theta_0\right)\right] = 0,$$

- Under mean stationarity, the following moments for the errors in levels are also available:

$$E\left[\Delta y_{i(t-1)}\left(y_{it} - y_{i(t-1)}\theta_0\right)\right] = 0.$$

- Autoregressive models are the workhorse in the analysis of individual earnings and household income dynamics.

Permanent-transitory income models

- Permanent-transitory models are common in the literature that looks at the relationship between household income and consumption from a life-cycle perspective.
- Examples include Hall & Mishkin (1982) (HM), Blundell, Pistaferri & Preston (2008), and Kaplan & Violante (2010).
- HM used food consumption and labour income from a PSID sample of $N = 2309$ US households over $T = 7$ years to test the predictions of a permanent income model.
- We use HM as an illustration of permanent-transitory covariance structures.
- HM specified means of income and consumption changes as regressions on age, age^2, time, and changes in the number of children and adults in the household.
- They implicitly allowed for unobserved intercept heterogeneity in the levels of the variables, but only for observed heterogeneity in their changes.
- Deviations from the individual means of income and consumption, denoted $\overline{y}_{it}$ and $\overline{c}_{it}$ respectively, were specified as follows.

Income process

- HM assumed that income errors $\overline{y}_{it}$ were the result of two different types of shocks, permanent and transitory:

$$\overline{y}_{it} = y_{it}^{L} + y_{it}^{S}.$$

- They also assumed that agents were able to distinguish one type of shock from the other and respond to them accordingly.
- The permanent component y_{it}^{L} was specified as a random walk

$$y_{it}^{L} = y_{i(t-1)}^{L} + \varepsilon_{it},$$

  and the transitory component y_{it}^{S} as a moving average process

$$y_{it}^{S} = \eta_{it} + \rho_1 \eta_{i(t-1)} + \rho_2 \eta_{i(t-2)}.$$

- A limitation was lack of measurement error in observed income (a component to which consumption does not respond). This is important since measurement error in PSID income is large, but identification requires cross-validation information.

Consumption process

- Mean deviations in consumption changes were specified to respond one-to-one to permanent income shocks and by a fraction β to transitory shocks.
- The magnitude of β depends on the persistence in transitory shocks (ρ_1 and ρ_2) and real interest rates. Dependence on age is ignored for simplicity.
- This model can be formally derived from an optimization problem with quadratic utility, and constant interest rates that are equal to the subjective discount factor.
- Since only food consumption is observed, an adjustment was made by assuming a constant marginal propensity to consume food α.
- With these assumptions we have

$$\Delta \overline{c}_{it} = \alpha \varepsilon_{it} + \alpha \beta \eta_{it}.$$

- HM also introduced a measurement error in the level of consumption (or transitory consumption that is independent of income shocks) with an MA(2) specification:

$$c_{it}^{S} = v_{it} + \lambda_1 v_{i(t-1)} + \lambda_2 v_{i(t-2)}.$$

Bivariate covariance structure

- The model that is taken to the data consists of a joint specification for mean deviations in consumption and income changes as follows:

$$\Delta\overline{c}_{it} = \alpha\varepsilon_{it} + \alpha\beta\eta_{it} + v_{it} - (1-\lambda_1)\, v_{i(t-1)} - (\lambda_1 - \lambda_2)\, v_{i(t-2)} - \lambda_2 v_{i(t-3)}$$

$$\Delta\overline{y}_{it} = \varepsilon_{it} + \eta_{it} - (1-\rho_1)\, \eta_{i(t-1)} - (\rho_1 - \rho_2)\, \eta_{i(t-2)} - \rho_2 \eta_{i(t-3)}.$$

- The three innovations are mutually independent with variances σ^2_ε, σ^2_η and σ^2_v. Thus, the model contains 9 coefficients:

$$\theta = \begin{pmatrix} \alpha & \beta & \lambda_1 & \lambda_2 & \rho_1 & \rho_2 & \sigma^2_\varepsilon & \sigma^2_\eta & \sigma^2_v \end{pmatrix}'.$$

- The model specifies a covariance structure for the 12×1 vector

$$w_i = \begin{pmatrix} \Delta\overline{c}_{i2} & \Delta\overline{c}_{i3} & \cdots & \Delta\overline{c}_{i7} & \Delta\overline{y}_{i2} & \Delta\overline{y}_{i3} & \cdots & \Delta\overline{y}_{i7} \end{pmatrix}'$$

$$E\left(w_i w_i'\right) = \Omega(\theta).$$

Bivariate covariance structure (continued)

- Let us look at the form of some elements of $\Omega(\theta)$.

$$Var(\Delta\overline{y}_{it}) = \sigma_\varepsilon^2 + 2\left(1 - \rho_1 - \rho_1\rho_2 + \rho_1^2 + \rho_2^2\right)\sigma_\eta^2 \qquad (t = 2, ..., 7)$$

$$Cov(\Delta\overline{y}_{it}, \Delta\overline{y}_{i(t-1)}) = -\left[(1 - \rho_1) - (1 - \rho_1 + \rho_2)(\rho_1 - \rho_2)\right]\sigma_\eta^2$$

and also

$$Cov(\Delta\overline{c}_{it}, \Delta\overline{y}_{it}) = \alpha\sigma_\varepsilon^2 + \alpha\beta\sigma_\eta^2 \qquad (t = 2, ..., 7) \tag{2}$$

$$Cov(\Delta\overline{c}_{it}, \Delta\overline{y}_{i(t-1)}) = 0 \tag{3}$$

$$Cov(\Delta\overline{c}_{i(t-1)}, \Delta\overline{y}_{it}) = -\alpha\beta(1 - \rho_1)\sigma_\eta^2. \tag{4}$$

- A fundamental restriction of the model is lack of correlation between current consumption changes and lagged income changes, as captured by (3).
- The model, nevertheless, predicts correlation between current consumption changes and current and future income changes, as seen from (2) and (4).

Empirical results

- HM estimated their model by Gaussian PML. They estimated $\widehat{\beta} = 0.3$, which given their estimates of ρ_1 and ρ_2 ($\widehat{\rho}_1 = 0.3$, $\widehat{\rho}_2 = 0.1$) turned out to be consistent with the model only for unrealistic values of real interest rates (above 30 percent).
- Moreover, they estimated the marginal propensity to consume food as $\widehat{\alpha} = 0.1$, and the moving average parameters for transitory consumption as $\widehat{\lambda}_1 = 0.2$ and $\widehat{\lambda}_2 = 0.1$.
- The variance of the permanent income shocks was twice as large as that of the transitory shocks: $\widehat{\sigma}^2_{\varepsilon} = 3.4$ and $\widehat{\sigma}^2_{\eta} = 1.5$.
- They tested the covariance structure focusing on the fundamental restriction of lack of correlation between current changes in consumption and lagged changes in income. They found a negative covariance which was significantly different from zero.
- As a result of this finding they considered an extended version of the model in which a fraction of consumers spent their current income.

GMM estimation of covariance structures

- The previous model specifies a structure on a data covariance matrix. Abstracting from mean components, suppose the covariance matrix of a $p \times 1$ time series y_i is a function of a $k \times 1$ parameter vector θ given by

$$E(y_i y_i') = \Omega(\theta).$$

- If y_i is a scalar time series its dimension will be T, but in the HM context $p = 2T$.
- Vectorizing the expression and eliminating redundant elements (due to symmetry) we obtain a vector of moments of order $r = (p+1)p/2$:

$$vechE\left[y_i y_i' - \Omega(\theta)\right] = E\left[s_i - \omega(\theta)\right],$$

  where the *vech* operator stacks by rows the lower triangle of a square matrix.
- If $r > k$ and $H(\theta) = \partial\omega(\theta)/\partial\theta'$ has full column rank, the model is overidentified. In that case a standard optimal GMM estimator solves:

$$\widehat{\theta} = \arg\min_c \left[\overline{s} - \omega(c)\right]' \widehat{V}^{-1} \left[\overline{s} - \omega(c)\right]$$

  where $\overline{s}$ is the sample mean vector of s_i:

$$\overline{s} = \frac{1}{N}\sum_{i=1}^{N} s_i$$

  and $\widehat{V}$ is some consistent estimator of $V = Var(s_i)$. A natural choice is the sample covariance matrix of s_i:

$$\widehat{V} = \frac{1}{N}\sum_{i=1}^{N} s_i s_i' - \overline{s}\overline{s}'.$$

GMM estimation of covariance structures (continued)

- The first-order conditions from the optimization problem are

$$-H(c)'\widehat{V}^{-1}\left[\overline{s}-\omega(c)\right]=0.$$

- The two standard results for large sample inference are, firstly, asymptotic normality of the scaled estimation error

$$\left[\frac{1}{N}H(\widehat{\theta})'\widehat{V}^{-1}H(\widehat{\theta})\right]^{-1/2}\left(\widehat{\theta}-\theta\right)\xrightarrow{d}\mathcal{N}(0,I)$$

  and, secondly, the asymptotic chi-square distribution of the minimized estimation criterion (*test statistic of overidentifying restrictions*)

$$S=N\left[\overline{s}-\omega(\widehat{\theta})\right]'\widehat{V}^{-1}\left[\overline{s}-\omega(\widehat{\theta})\right]\xrightarrow{d}\chi^2_{r-k}.$$

Random coefficients

- Fixed effects methods are a standard way of controlling for endogeneity or unobserved heterogeneity in the estimation of common parameters.
- But sometimes we wish to treat a parameter as a heterogeneous quantity and therefore its mean and other characteristics of its distribution become central objects of interest.
- Examples are random trend earnings models, heterogeneous production functions, and heterogeneous treatment effects.
- The T equations of the random coefficients model in compact form can be written as

$$y_i = Z_i\delta_0 + X_i\gamma_i + v_i \qquad E\left(v_i \mid Z_i, X_i, \gamma_i\right) = 0.$$

- The WG model is a special case in which the only random coefficient is the intercept.
- We assume that $T > \dim\left(\gamma_i\right) = q$ and only consider the subpopulation with $\det\left(X_i'X_i\right) \neq 0$.
- The parameters of interest are δ_0 and characteristics of the distribution of γ_i, such as $\gamma_0 = E\left(\gamma_i\right)$ and $\Sigma_0 = Var\left(\gamma_i\right)$.
- Now instead of considering LS in deviations from means we consider LS of the residuals in individual-specific regressions of y and z on x ($\widetilde{x}_{it}$ is the residual of a regression of the i-th time series of x on an intercept).

Estimating common parameters and average effects

- The generalized WG operator $Q_i = I - X_i \left(X_i' X_i\right)^{-1} X_i$ leads to the transformed equation

$$Q_i y_i = Q_i Z_i \delta_0 + Q_i v_i$$

and the moments

$$E\left[Z_i' \left(Q_i y_i - Q_i Z_i \delta_0\right)\right] = 0.$$

- The WG estimator is

$$\widehat{\delta} = \left(\sum_{i=1}^{N} Z_i' Q_i Z_i\right)^{-1} \sum_{i=1}^{N} Z_i' Q_i y_i$$

- Pre-multiplying the model by the LS operator $H_i = \left(X_i' X_i\right)^{-1} X_i'$ we get

$$H_i \left(y_i - Z_i \delta_0\right) = \gamma_i + H_i v_i$$

so that γ_0 satisfies the moment

$$\gamma_0 = E\left[H_i \left(y_i - Z_i \delta_0\right)\right]$$

and a large-N consistent estimator is

$$\widehat{\gamma} = \frac{1}{N} \sum_{i=1}^{N} \left(X_i' X_i\right)^{-1} X_i' \left(y_i - Z_i \widehat{\delta}\right) \equiv \frac{1}{N} \sum_{i=1}^{N} \widehat{\gamma}_i.$$

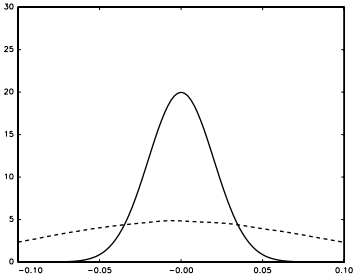
Is $\widehat{\gamma}_i$ informative about γ_i? An illustration

- Consider the random trend model:

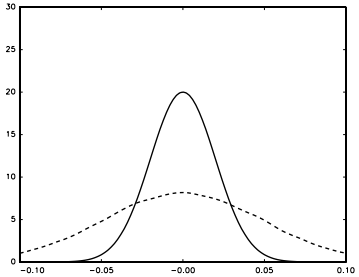
$$y_{it} = \alpha_i + \beta_i t + v_{it}$$

  where α_i and β_i are bivariate normal (or bimodal normal mixture), v_{it} is normal AR(1) with autoregressive coefficient ρ.
- Roughly calibrate the parameters to match Guvenen (2008): $\rho = .8$, $\mathrm{Var}(\alpha_i) = .02$, $\mathrm{Var}(\beta_i) = .0004$ (corr. $= -.2$), $\sigma_v^2 = .03$.
- Question: compare the density of $\widehat{\beta}_i$ (resp. $\widehat{\alpha}_i$) to that of β_i (α_i).

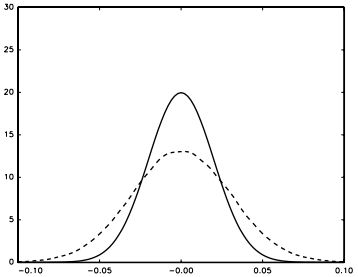
Densities: true β_i (solid) and fixed-effects estimates $\widehat{\beta}_i$ (dashed)



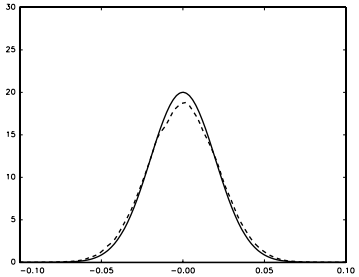
$T = 5$



$T = 10$

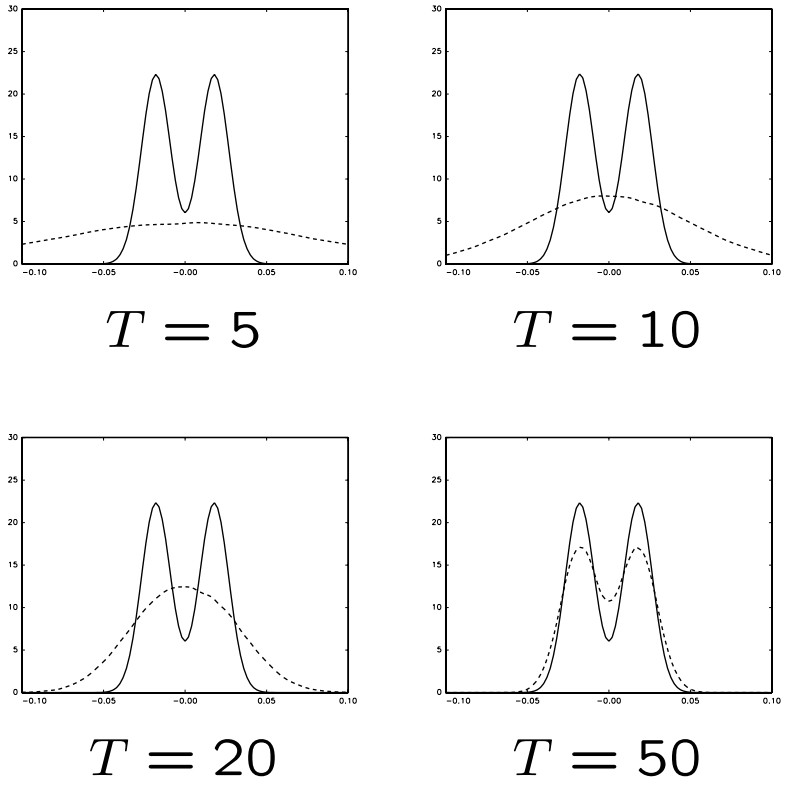


$T = 20$



$T = 50$

Densities: true β_i (solid) and fixed-effects estimates $\widehat{\beta}_i$ (dashed)



$\Rightarrow$ Must correct the densities of fixed-effects estimates for the sample noise (for fixed T).

Estimating variances of effects and distributions

- Without further restrictions Σ_0 is not identified. To see this let $\Omega_i = E\left(v_i v_i' \mid X_i\right)$ and note that only the variance of $Q_i v_i$ is identified, which is of reduced rank. In general

$$\Sigma_0 = Var\left[H_i\left(y_i - Z_i\delta_0\right)\right] - E\left(H_i \Omega_i H_i'\right).$$

- If $\Omega_i = \sigma^2 I_T$ then Σ_0 can be estimated as

$$\widehat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\gamma}_i - \widehat{\gamma}\right)\left(\widehat{\gamma}_i - \widehat{\gamma}\right)' - \widehat{\sigma}^2 \frac{1}{N}\sum_{i=1}^{N}\left(X_i' X_i\right)^{-1}$$

  where

$$\widehat{\sigma}^2 = \frac{1}{N\left(T-q\right)}\sum_{i=1}^{N}\left(y_i - Z_i\widehat{\delta}\right)' Q_i \left(y_i - Z_i\widehat{\delta}\right).$$

- Note that $E\left(Q_i v_i v_i' Q_i\right) = \sigma^2 E\left(Q_i\right)$ and $E\left(v_i' Q_i v_i\right) = \sigma^2\left(T-q\right)$.

Estimating variances of effects and distributions (continued)

- The previous situation can be generalized to less restrictive covariance patterns in Ω_i.
- In general

$$E\left[(y_i - Z_i\delta_0) \otimes (y_i - Z_i\delta_0) \mid Z_i, X_i\right] = (X_i \otimes X_i)\, E\left(\gamma_i \otimes \gamma_i \mid Z_i, X_i\right) + vec\left(\Omega_i\right).$$

- A WG operator $M_i = I - G_i \left(G_i' G_i\right)^{-1} G_i'$ for the cross-products $G_i = X_i \otimes X_i$ leads to

$$M_i E\left[(y_i - Z_i\delta_0) \otimes (y_i - Z_i\delta_0) \mid Z_i, X_i\right] = M_i\, vec\left(\Omega_i\right)$$

  but since M_i is singular, (moving-average) restrictions on Ω_i are needed:

$$vec\left(\Omega_i\right) = S_2\omega_i$$

  where S_2 is a known selection matrix and ω_i is a vector of unrestricted parameters.
- The rank condition for identification of Ω_i is

$$rank\left(M_i S_2\right) = \dim\left(\omega_i\right).$$

- The variance of γ_i is identified if Ω_i is known.

- Moreover, replacing mean independence by full independence assumptions a similar argument can be developed for distributions using second derivatives of log characteristic functions (Arellano and Bonhomme 2012).

Distributions

- Assume that γ_i and v_i are independent given $W_i = (Z_i, X_i)$.
- Statistical independence leads to functional restrictions on the second derivatives of log characteristic functions, which are formally analogous to the covariance restrictions.
- To derive the identification results, it is convenient to work with characteristic functions.

Properties of characteristic functions

- The conditional characteristic function of Y (of dimension L) given $X = x$ is defined as:

$$\Psi_{Y|X}(t|x) = E\left[\exp(jt'Y)|x\right], \quad t \in R^L$$

  where $j = \sqrt{-1}$.
- Inverse Fourier transform

$$f_{Y|X}(y|x) = \frac{1}{(2\pi)^L} \int \exp\left(-jt'y\right) \Psi_{Y|X}(t|x) dt.$$

- If Y_1 and Y_2 are independent given X then

$$\Psi_{Y_1+Y_2|X}(t|x) = \Psi_{Y_1|X}(t|x)\Psi_{Y_2|X}(t|x).$$

Distributions (continued)

- Independence implies that for all t we have:

$$\Psi_{y_i - Z_i\delta_0|W_i}(t|W_i) = \Psi_{\gamma_i|W_i}(X_i't|W_i)\Psi_{v_i|W_i}(t|W_i).$$

- Assuming that the characteristic functions $\Psi_{\gamma_i|W_i}$ and $\Psi_{v_i|W_i}$ are nonvanishing we can take logs:

$$\log\Psi_{y_i - Z_i\delta_0|W_i}(t|W_i) = \log\Psi_{\gamma_i|W_i}(X_i't|W_i) + \log\Psi_{v_i|W_i}(t|W_i).$$

- If $\Psi_{v_i|W_i}$ is identified, $\Psi_{\gamma_i|W_i}$ is also identified.
- Taking second derivatives:

$$\frac{\partial^2 \log\Psi_{y_i - Z_i\delta_0|W_i}(t|W_i)}{\partial t\partial t'} = X_i\left(\frac{\partial^2 \log\Psi_{\gamma_i|W_i}(X_i't|W_i)}{\partial t\partial t'}\right)X_i' + \frac{\partial^2 \log\Psi_{v_i|W_i}(t|W_i)}{\partial t\partial t'}.$$

- Evaluating this expression at $t = 0$ we are back at the variance case.

Distributions (continued)

- An independent moving-average model implies the following restrictions:

$$vec\left(\frac{\partial^2 \log \Psi_{v_i|W_i}(t|W_i)}{\partial t \partial t'}\right) = S_2 \omega_i(t), \qquad t \in R^T.$$

- So, if $M_i(X_i \otimes X_i) = 0$ then

$$M_i vec\left(\frac{\partial^2 \log \Psi_{y_i - Z_i\delta_0|W_i}(t|W_i)}{\partial t \partial t'}\right) = M_i S_2 \omega_i(t).$$

- The rank and order conditions for identification are the same as for variances.
- $\omega_i(t)$ identified for all t implies that $\Psi_{v_i|W_i}$ is identified, because the first derivative of $\log \Psi_{v_i|W_i}$ at $t = 0$ vanishes due to mean independence.

Illustration: the effect of smoking on children outcomes

- Arellano and Bonhomme (2012) apply this methodology to a matched panel dataset of mothers and births constructed in Abrevaya (2006).
- They find that the mean smoking effect on birthweight is significantly negative (-160 grams). Moreover, the effect shows substantial heterogeneity across mothers, the effect being very negative (-400 g) below the 20th percentile.

- The model is
$$y_{ij} = \mathbf{z}_{ij}'\delta + \alpha_i + \beta_i s_{ij} + v_{ij} \quad j = 1, 2, 3$$
i=mother, j=child. y_{ij}= weight at birth, $s_{ij} = 1$ if mother smoked during pregnancy of child j.
- v_{ij} are assumed i.i.d.
- Production function interpretation. The effect of smoking is mother-specific.
- Abrevaya (2006) estimates a restricted version, where β_i is homogeneous.

- The focus is on mothers with at least 3 children to be able to allow for two heterogeneous quantities.
- Also need x_{ij} to vary for every mother. So only 1445 mothers who changed smoking status between the three births are considered.
- Under predeterminedness of smoking behavior the moments of β_i are unidentified. However, several interesting average effects can still be identified and estimated when there are no time-varying regressors.

Estimates of common parameters δ

Generalized within-groups		
Variable	Estimate	Standard error
Male	130	22.8
Age	39.0	32.0
Age-sq	-.638	.577
Kessner=2	-82.0	52.7
Kessner=3	-159	81.9
No visit	-18.0	124
Visit=2	83.2	53.9
Visit=3	136	99.2

Regressions of α_i and β_i on mother-specific characteristics

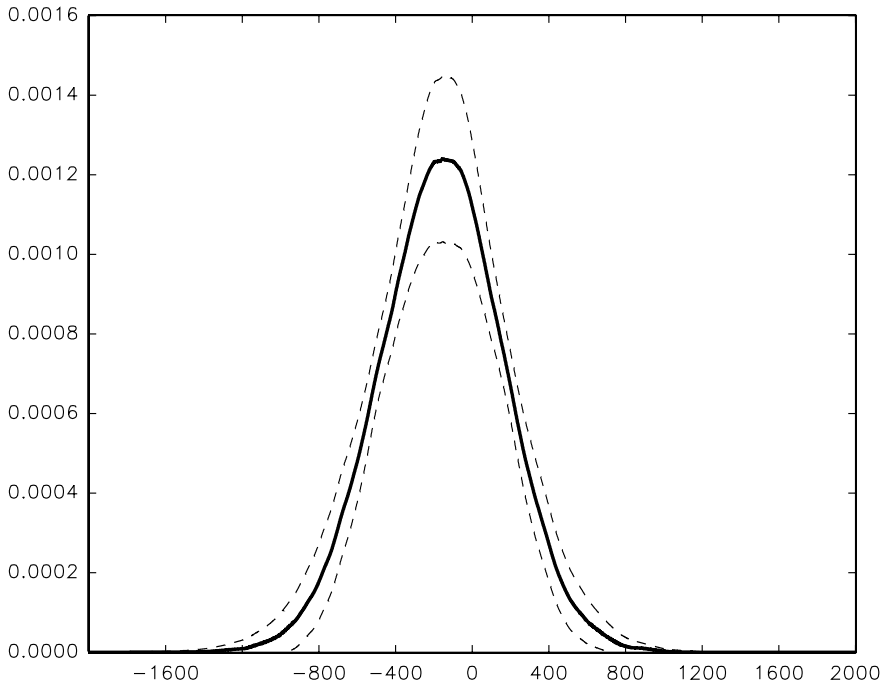
Variable	Estimate	Standard error
	α_i	
High-school	15.1	42.7
Some college	38.5	55.3
College graduate	58.7	72.1
Married	3.51	34.6
Black	-364	54.0
Mean smoking	-161	83.9
Constant	2879	419
corrected R^2= .113 (instead of .055, uncorrected)		
	β_i	
High-school	-15.9	42.8
Some college	-15.9	42.8
College graduate	64.5	63.8
Married	31.9	41.8
Black	132	60.6
Mean smoking	-49.8	101
Constant	-172	67.1
R^2= .021 (instead of .005)		

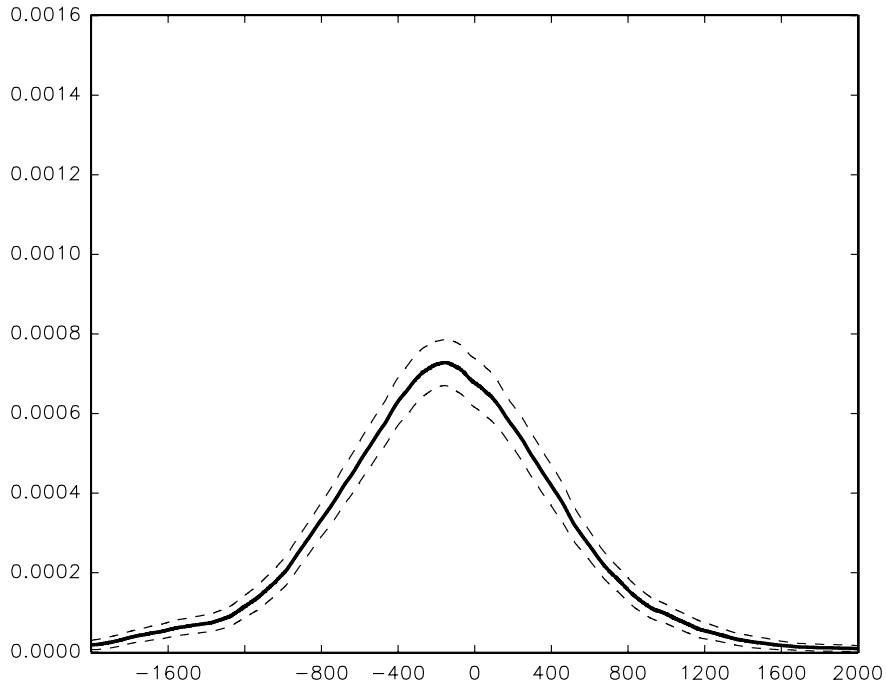
Moments of α_i and β_i

Moment	Estimate	Standard error
Mean α_i	2782	435
St. Dev. α_i	357	21.2
Skewness α_i	-1.67	.43
Kurtosis α_i	7.12	2.28
Mean β_i	-161	17.0
St. Dev. β_i	313	34.6
Skewness β_i	-1.29	.91
Kurtosis β_i	-.34	7.84
Correlation (α_i, β_i)	-.47	.07

- Mean effect of smoking is -161 grams, close to Abrevaya's FE estimate of -144 g.
- Density of β_i and $\widehat{\beta}_i$.
- Quantile function of β_i and $\widehat{\beta}_i$.

Density of β_i (left) and $\widehat{\beta}_i$ (right)





Quantile function of β_i (left) and $\widehat{\beta}_i$ (right)