Ec485 Lecture 3, WT2024

1 Extensions to Models with Observable Dynamics — Linear Models:

1.1 *single or more lagged DV*

Consider the *linear dynamic balanced panel data model*:

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + z_i'\gamma + \alpha_i + \nu_{it}, \qquad i = 1, \cdots, N, \quad t = 1, \cdots, T$$

1. with: k_x time-varying regressors, k_z time-invariant regressors, a_i an unobservable error i.i.d. over i, with unconditional zero mean and variance $\sigma_\alpha^2 < \infty$, ν_{it} an error independent of all as and i.i.d. over both i and t with unconditional mean zero and variance $\sigma_\nu^2 < \infty$.

1.2 *RE1&2: Problems with OLS/GLS

1.3 *FE-type:

1.3.1 *FE1: Problems with First Differencing

;

1.3.2 *FE2/Within: Problems with Within/Quasi-Differencing

1.4 Solutions to FE-Type Estimators:

Arellano-Bond: Delta/FD + IVE

Applying first differencing to the original model gives:

$$\Delta y_{it} = \delta \Delta y_{i,t-1} + \Delta x'_{it}\beta + \Delta \nu_{it}, \qquad i = 1, \cdots, N, \quad t = 2, \cdots, T$$

1. (a) Since the error term $\Delta \nu_{it}$ is a MA(1) with known parameter -1, valid instrumental variables for the lagged dependent variable term are: $\Delta y_{i,t-m}$ and $y_{i,t-m}$ for $m \geq 2$. Note that there is a triangular structure in the set of optimal instruments, since the further along one moves in time, the greater the number of valid instruments.

   NB: the regular $y_{it} - \bar{y}_{i.}$ transformation is not useful for this model since in that case, no valid instruments can be obtained by lagging the ys and Δys any number of times, because $y_{i,t-1} - \bar{y}_{i.}$ and the implied error $\nu_{i,t-1} - \bar{\nu}_{i.}$ are serially correlated with y_{is} and Δy_{is} for *every* s.

1.5 Solutions to RE-Type Estimators:

Bhargava-Sargan: System Estimation (2SLS, 3SLS, FIML)
URL: <https://eprints.lse.ac.uk/102843/> Section 2

The Bargava and Sargan approach:

1. (a) Step 1 – write out explicitly as a separate equation for each t, so a cross-section on a System of T equations for the T endogenous variables $y_1, y_2, \cdots, y_T$ and the exogenous variables $x_{iT}, x_{i,T-1}, \cdots, x_{i1}, z_i$

   Step 2 – write out a linear quasi-reduced form equation for y_{i1} in terms of full exogenous information available to the econometrician (but not actually available to the economic agents at time $t = 1$):

$$\begin{array}{rcl} y_{iT} & = & \delta y_{i,T-1} + x'_{iT}\beta + z'_i\gamma + \alpha_i + \nu_{iT} \\ y_{i,T-1} & = & \delta y_{i,T-2} + x'_{i,T-1}\beta + z'_i\gamma + \alpha_i + \nu_{i,T-1} \\ \vdots & \vdots & \vdots \\ y_{i2} & = & \delta y_{i1} + x'_{i2}\beta + z'_i\gamma + \alpha_i + \nu_{i2} \\ y_{i1} & = & x'_{i1}\theta_1 + x'_{i2}\theta_2 + \cdots + x'_{iT}\theta_T + z'_i\zeta + u_{i1} \end{array} \qquad i = 1, \cdots, N$$

   i. Optimal estimation assuming normality of the errors is achieved through *Full Information MLE* of $(\delta, \beta', \gamma', \theta'_1, \cdots, \theta'_T, \zeta')'$ and $(\sigma^2_\alpha, \sigma^2_\nu, \sigma^2_0)$ implied by the cross-equation restrictions of the above system and the variance-covariance restrictions of the structure:

$$\begin{pmatrix} \omega_T^2 & \omega_{T,T-1} & \omega_{T,T-2} & \cdots & \omega_{T2} & \omega_{T1} \\ \omega_{T-1,T} & \omega_{T-1}^2 & \omega_{T-1,T-2} & \cdots & \omega_{T-1,2} & \omega_{T-1,1} \\ \omega_{T-2,T} & \omega_{T-2,T-1} & \ddots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \omega_{32} & \omega_{31} \\ \omega_{2T} & \omega_{2,T-1} & \cdots & \omega_{23} & \omega_2^2 & \omega_{21} \\ \omega_{1T} & \omega_{1,T-1} & \cdots & \omega_{13} & \omega_{12} & \omega_1^2 \end{pmatrix} =$$

$$\begin{pmatrix} \sigma_\nu^2+\sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 & A \\ \sigma_\alpha^2 & \sigma_\nu^2+\sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 & A \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \sigma_\alpha^2 & A \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 & \sigma_\nu^2+\sigma_\alpha^2 & A \\ A & A & \cdots & A & A & B \end{pmatrix}$$

where $A = \frac{\sigma_\alpha^2}{1-\delta}$ and $B = \frac{\sigma_\alpha^2}{(1-\delta)^2} + \frac{\sigma_\nu^2}{1-\delta^2} + \sigma_0^2$.

Without assuming normality, the optimal linear system estimator is $3SLS$

ii. If one wants to test the one-factor analytic structure, one can carry out FIML $(\delta, \beta', \gamma', \theta_1', \cdots, \theta_T', \zeta')'$ with an unrestricted Ω cross-equation variance-covariance, and compare the results to those of the first FIML through, say, a Likelihood Ratio statistic.

iii. If one believes that the initial condition y_{i1} is exogenous, then one applies $FIML$ (under Normality) or $3SLS$ (without Normality) on the system with $T-1$ equations in the $T-1$ endogenous variables $y_2, \cdots, y_T$ and the exogenous variables $y_1, x_{iT}, x_{i,T-1}, \cdots, x_{i1}, z_i$. To test the exogeneity of y_{i1}, one would need to carry out a *non-nested* test, since the null hypothesis that y_{i1} is exogenous implies $A = B = 0$ *simultaneously.*

1.6 Various Extensions to PDMs: Static and Dynamic

1.6.1 General Endogeneity in Time-varying regressors w.r.t. disturbances

1. (a) i. One of the x_{it} regressors is correlated with ν_{it}.
       ii. All of the x_{it} regressors are correlated with α_i.

1.6.2 Regressor Engogeneity because of Measurement Errors

1. (a) i. One of the x_{it} regressors is measured with error, ξ_{it}.
       ii. One of the z_i regressors is measured with error, ζ_i.

1.6.3 Disturbances follow more complicated Autocorrelated Processes

Consider the dynamic linear regression model for balanced data:

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + z_i'\gamma + \epsilon_{it} \quad , \quad i = 1, \cdots, N \ , \quad t = 1, \cdots, T$$

where ϵ_{it} follows the one factor error components model: $\epsilon_{it} = \alpha_i + \nu_{it}$ with α_i modelling individual unobserved persistent heterogeneity.

1. (a) i. $\nu_{it} = \xi_{it} + \lambda\xi_{i,t-1}$ with $\xi_{it} \sim N(0, \sigma_\xi^2)$ i.i.d. over both i and t;
   Now valid instruments for AB are values of $y_{i,t-q}$, $q \geq 3$.

   ii. $\nu_{it} = \rho\nu_{i,t-1} + \xi_{it}$ with $|\rho| < 1$ and $\xi_{it} \sim N(0, \sigma_\xi^2)$ i.i.d. over both i and t.
   AB will not provide consistent estimates since no valid instruments exist in this case: The variables $y_{i,t-q}$, *are correlated with the regressor* $y_{i,t-1}$ *for *any** q.

   *In all three cases (i)-(iii), the B-S approach remains valid, since the system estimation approach used (3SLS or FIML) allows for *any* valid correlation structure among the equation errors,* $(u_{i1}, \epsilon_{i2}, \epsilon_{i3}, \cdots, \epsilon_{it}, \cdots, \epsilon_{iT})'$. *Hence in fact the particular error structures (i)-(iii) can be *tested* using classical tests (Wald, LR, LM), with the Restricted model imposing the particular correlation structure in the estimation vs. the Unrestricted model with allowing 3SLS or FIML to estimate the variance-cov structure of the errors..*

2 Final Extensions to PD Modelling: Nonlinearities

Nonlinear PDMs:
URL: <https://eprints.lse.ac.uk/102843/> Section 3

2.1 A. Additive Errors — Index Models and General Models

y(i,t)=f()+epsilon(i,t)
y(i,t)=g()+epsilon(i,t)
alternative models:

1. (a)

$$y_{it} = g(x_{it}, \beta, z_i, \gamma) + \delta y_{i,t-1} + \epsilon_{it} \qquad \text{(Model 1)}$$

   where the non-linear function $g(.)$ is known up to parameter vectors β and γ;

   (b) *The first model is additive in the errors, so it can be analysed completely analogously by combining RE and FE or Δ transformations with NLLS instead of OLS, or GMM in place of IV as necesary. Key thing to remember: the FE and RE operators must be applied to the non-linear function* exp(.) *and *not* the non-linear function evaluated at the FE- or RE-transformed data.*
   I.e., using $g((x_{it} - \lambda_i \bar{x}_{i\cdot})'\beta + (1 - \lambda_i)z_i'\gamma)$ would be wrong for RE, while we should use instead: $g(x_{it}'\beta + z_i'\gamma) - \lambda_i g(x_{i,t-1}'\beta + z_i'\gamma)$ for the non-linear term.
   Since Model 1 contains the additive dynamic term $+\delta y_{i,t-1}$ it is not appropriate to combine the usual RE or FE transformations together with NLLS to account for the presence of the g(.) term, just like the linear case where OLS to the transformed models would lead to inconsistency because of the endogeneity of all transformations of the $+\delta y_{i,t-1}$ term. For example, applying first differencing to eliminate the alpha term, gives:

$$y_{it} - y_{i,t-1} = g(x_{it}, \beta, z_i, \gamma) - g(x_{i,t-1}, \beta, z_i, \gamma) + \delta(y_{i,t-1} - y_{i,t-2}) + \nu_{it} - \nu_{i,t-1}$$

Hence, one cannot apply NLLS to this model because of the MA(1) of the resulting error term. Instead, one should use NLIV/GMM based on $y_{i,t-2}, y_{i,t-3}, \ldots$ terms as valid instrumental variables.

1. (a) *In Model 2, there is a very significant additional complication: the non-linearity encompasses also the $y_{i,t-1}$ part. The presence of the lagged term under the non-linear function makes this model non-additive in the error term (at least with the α_i present in all periods). Hence RE- or FE- plus NLLS will *not* work for this model, but we need to use instead MLE that takes into account correctly the non-trivial Jacobian of the $y \longrightarrow error$ transformation.*

   *Another possibility for estimating this model consistently (though not efficiently, as is the case of MLE) would be as follows: assuming, as with MLE, that the regressors are *strongly* exogenous w.r.t. the error term, implies that lagged Xs are valid instruments for the (endogenous) lagged ys that appear as regressors. Hence NLIV/GMM could be used instead.*

2.2 B. Nonadditive Errors

y(i,t)=h()

(A): FE-Type Estimators

FE1: Delta/FD

FE2: Within/Quasi-Differencing

(A): RE-Type Estimators

RE1: OLS–>NLLS

RE2: GLS–>WNLL

1. (a)

$$y_{it} = h(x_{it}, \beta, z_i, \gamma, \delta y_{i,t-1}) + \epsilon_{it} \qquad \text{(Model 2)}$$

and where the non-linear function $h(.)$ is known up to parameter vectors β and γ and parameter δ.

...

1. (a)

Finally suppose that in part (b), the δ parameter equals 0. What happens to Models 1 and 2 in such case? Discuss estimation when (i) all regressors are measured without error; and (ii) when one or more regressor(s) contain(s) errors of measurement. In such case (ii), does it make a difference whether the mismeasured regressors are among the Xs or the Zs?

1. (a) *If δ is 0, then Models 1 and 2 have exactly the same structure. So let us focus on Model 1.*

   If all the regressors are measured correctly (case (i)), then this is the classic case (easy) with the error term appearing additively outside the nonlinear function g(.). Solution: FE or RE transformations plus NLLS.

   *In case (ii) however, the problem becomes essentially one with *non*additive error terms, since the measurement errors are inside the nonlinear function. It is equally hard to deal with this problem whether it is the Xs or Zs that are mismeasured — there is no simplification afforded by the within or first-differencing transformations, since the Zs are underneath the g(.) function, and hence such transformations will not eliminate the time-invariant Zs.*

   (b) ?

3 DYNAMICS and Nonlinearities:

Case 1: +delta*y(t-1): (A) or (B) depending on *where* lagged DV enters:
Case 2: ARMA in the errors Typically (B) because of Koyck transformations

4 Major Difficulties with B. Nonadditive Errors:

Difficulty 1: FD/Delta, Within differencing, GLS quasi-differencing transformations do not achieve anything special/useful

Difficulty 2: Fe-type alternative idea of introducing N intercepts/dummies leads to "Infinite Incidental Parameters" problem

Difficulty 3: The epsilon–>y transformation — Jacobian is not 1; is not constant; depends on data and unknown parameters

***Very interesting class of models with Nonadditive Nonlinearity is LDV class of models

Simple PD version:

Multiperiod Binary Probit Model (Heckman 1981)

Case 1: without lagged DV dynamics

Case 2: Lagged Limited DV vs. Lagged Latent DV —> State-Dependence *vs* Unobserved Persistent Heterogeneity

Difficulty 4: T*contemporaneous_correlated_dimension = M_i correlated dimensions per individual observation i —> typically *integrals* of order M_i for each likelihood contribution

—> Motivating Simulation-Based Inference