Dynamic panel data models

Přemysl Bejda

premyslbejda@gmail.com

2015

Contents

Contents

Contents

Our model

Definition 1 (Model with lagged dependent variable).

$$y_{i,t} = \delta y_{i,t-1} + x_{i,t}^{\top}\beta + u_{i,t},\ i = 1, \ldots, N;\ t = 1, \ldots, T, \quad (1)$$

where δ is a scalar and unknown parameter, $x_{i,t}^{\top}$ is $1 \times K$ and β is $K \times 1$ vector of unknown parameters. We will assume that the $u_{i,t}$ follow a one-way error component model

$$u_{i,t} = \mu_i + \nu_{i,t},$$

where μ_i are iid with mean zero and standard deviation σ_μ, $\nu_{i,t}$ are iid with mean zero and standard deviation σ_ν and μ_i and ν_{it} are independent of each other.

Utilization of dynamic panel data models

- Demand for natural gas
- Demand for an addictive commodity like cigarettes
- Model of employment
- Model of company investment
- Lifecycle labor supply model

Dynamic in our model

In the model there are two sources of persistence (dynamic):

1. Autocorrelation due to the presence of a lagged dependent variable among the regressors,
2. individual effects characterizing the heterogeneity among the individuals.

OLS - difficulties

- $y_{i,t}$ and $y_{i,t-1}$ are according to model functions of μ_i.
- Therefore $y_{i,t-1}$ a right hand regressor is correlated with the error term.
- From that we get that OLS estimator is not an unbiased estimator.

Within transformation - difficulties

$$\bar{y}_{i\cdot} = \delta \bar{y}_{i\cdot -1} + \bar{x}_{i\cdot}^{\top} \beta + \bar{u}_{i\cdot}, \tag{2}$$

where $\bar{y}_{i\cdot} = \frac{1}{T} \sum_{t=1}^{T} y_{i,t}$, $\bar{y}_{i\cdot -1} = \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1}$ etc.

- Let us subtract equation (2) from (1).
- This transformation wipes out μ_i, but $y_{i,t-1} - \bar{y}_{i\cdot -1}$ are still correlated with $\nu_{i\cdot -1} - \bar{\nu}_{i\cdot}$.
- The $y_{i,t-1}$ is correlated with $\bar{\nu}_{i\cdot}$, because $\bar{\nu}_{i\cdot}$ contains $\nu_{i,t-1}$ which is obviously correlated with $y_{i,t-1}$.

First difference

- An alternative transformation, that wipes out the individual effects is the first difference.

$$\Delta y_{i,t} = \delta \Delta y_{i,t-1} + \Delta x_{i,t}^{\top} \beta + \Delta \nu_{i,t}, \; i = 1, \ldots, N; t = 1, \ldots, T,$$

  where $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ etc.
- This does not suffice, because $\Delta y_{i,t-1}$ are still correlated with $\Delta \nu_{i,t}$.
- We can employ instrument $\Delta y_{i,t-2}$ or simply $y_{i,t-2}$ for $\Delta y_{i,t-1}$.
- Instrumental variable should be correlated with explanatory variable, but independent from residuals.
- This instrumental variable is already not correlated with $\Delta \nu_{i,t}$.

First difference - difficulties

- This means that we employ two stage least squares:
  1. We find an estimate $\hat{\zeta}$ in
     $$\Delta y_{i,t-1} = \zeta \Delta y_{i,t-2} + \varepsilon.$$
  2. In the second stage we estimate the rest of parameters in
     $$\Delta y_{i,t} = \delta \hat{\zeta} \Delta y_{i,t-2} + \Delta x_{i,t}^{\mathsf{T}} \beta + \Delta \nu_{i,t}, \; i = 1, \ldots, N; \, t = 1, \ldots, T.$$
- It does not take into account difference structure on the residual disturbances $\Delta \nu_{i,t}$.
- In the literature is shown:
  - This estimator is consistent but not the most efficient.
  - If we employ $\Delta y_{i,t-2}$ as an instrumental variable then an estimator has a singularity point (no derivative) and large variances over a significant range of parameter values.
  - In contrast $y_{i,t-2}$ has no singularities and much smaller variances. It is therefore much more recommended.

Outline

Model and requirement

Definition 2 (Simpler model).

Let us start with the simpler model

$$y_{i,t} = \delta y_{i,t-1} + u_{i,t},\ i = 1, \ldots, N;\ t = 1, \ldots, T, \tag{3}$$

where $u_{i,t}$ is defined as before.

We want to get a consistent estimate of δ as $N \to \infty$ with T fixed. In order to do that we difference (3) to eliminate the individual effects

$$\Delta y_{i,t} = \delta \Delta y_{i,t-1} + \Delta \nu_{i,t},\ i = 1, \ldots, N;\ t = 1, \ldots, T.$$

$\Delta \nu_{i,t} = \nu_{i,t} - \nu_{i,t-1}$ is MA(1) with unit root.

Instruments

- For $t = 3$, the first period we observe following relationship, we have

$$y_{i,3} - y_{i,2} = \delta(y_{i,2} - y_{i,1}) + (\nu_{i,3} - \nu_{i,2}).$$

- In this case, $y_{i,1}$ is a valid instrument, since it is highly correlated with $(y_{i,2} - y_{i,1})$ and not correlated with $(\nu_{i,3} - \nu_{i,1})$.
- For $t = 4$ we observe

$$y_{i,4} - y_{i,3} = \delta(y_{i,3} - y_{i,2}) + (\nu_{i,4} - \nu_{i,3}).$$

- In this case, $y_{i,1}$ as well as $y_{i,2}$ are valid instruments, because they are not correlated with $(\nu_{i,4} - \nu_{i,3})$.
- In this fashion we can continue to time T adding an extra valid instrument in each forward period. For period T, the set of valid instruments becomes $(y_{i,1}, y_{i,2}, \ldots, y_{i,T-2})$.

Covariance matrix

- This instrumental variable does not account for the differenced error term.
- In fact, $\mathrm{E}(\Delta\nu_i \Delta\nu_i^\top) = \sigma_\nu^2 (I_N \otimes G)$, where $\Delta\nu_i^\top = (\nu_{i,3} - \nu_{i,3}, \ldots, \nu_{i,T} - \nu_{i,T-1})$ and

$$G = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}$$

  is $(T-2) \times (T-2)$ matrix.
- $\Delta\nu_{i,}$ is MA(1) with unit root.

Preliminary estimator

We define

$$W_i = \begin{bmatrix} [y_{i,1}] & 0 & \cdots & 0 \\ 0 & [y_{i,1}, y_{i,2}] & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [y_{i,1}, \ldots, y_{i,T-2}] \end{bmatrix}$$

- The matrix of instruments is $W = [W_1^\top, \ldots, W_N^\top]^\top$.
- The moment equations are given by $\mathrm{E}(W_i^\top \Delta\nu_{i,}) = 0$.
- Let us premultiply our simpler differenced model in vector form by $W^\top$, we get

$$W^\top \Delta y = W^\top (\Delta y_{-1})\delta + W^\top \Delta\nu. \quad (4)$$

The vector $\Delta y_i = (y_{i,3} - y_{i,2}, \ldots, y_{i,T} - y_{i,T-1})^\top$ and $\Delta y = ((\Delta y_1)^\top, \ldots, (\Delta y_N)^\top)^\top$.

GLS - generalized least squares

This slides serves only as a reminder and the notation is not consistent with the rest of this presentation.

- Let us have a model

$$y = X\beta + \varepsilon,$$

  where $\mathrm{E}(\varepsilon|X) = 0$ and $\mathrm{Var}(\varepsilon|X) = \Omega$ is a known matrix.
- To find an estimator we minimize the square of Mahalanobis distance

$$\hat{\beta} = \underset{b}{\operatorname{argmin}}(y - Xb)^\top \Omega^{-1}(y - Xb).$$

- The estimator has an explicit form

$$\hat{\beta} = (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} y.$$

- The estimator is unbiased, consistent, efficient and asymptotically normal.

One step estimator

- Performing GLS on (4) one gets Arellano and Bond preliminary one-step consistant estimator

$$\hat{\delta}_1 = \left[(\Delta y_{-1})^\top W \left(W^\top (I_N \bigotimes G) W \right)^{-1} W^\top (\Delta y_{-1}) \right]^{-1} \times \left[(\Delta y_{-1})^\top W \left(W^\top (I_N \bigotimes G) W \right)^{-1} W^\top (\Delta y) \right]. \quad (5)$$

- Let us remind that this estimator account for the differenced error term.

GMM - generalized method of moments

The two following slides serve only as a reminder and the notation is not consistent with the rest of this presentation.

- In order to apply GMM, we need to have "moment conditions", i.e., we need to know a vector-valued function $g(y, \theta)$ such that
$$\mathrm{E} g(y_i, \theta_0) = 0,$$
where y_i is a random variable. In our case $\mathrm{E}(W_i \Delta \nu_i) = 0$.
- The basic idea behind GMM is to replace the theoretical expected value $\mathrm{E}(\cdot)$ with its empirical analog — sample average
$$\frac{1}{n} \sum_{i=1}^{n} g(Y_i, \theta),$$
where Y_i are observed values.

GMM - generalized method of moments

- GMM estimator can be written as

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left(\frac{1}{n} \sum_{i=1}^{n} g(y_i, \theta) \right)^{\top} \Omega^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} g(y_i, \theta) \right),$$

  where Θ is some set of possible parameters and $\Omega = \mathrm{E}\left(g(y_i, \theta_0) g(y_i, \theta_0)^{\top} \right)$.
- In our case is $\Omega = \mathrm{E}\left((W\Delta\nu)^{\top} (W\Delta\nu) \right) = W^{\top} (I_N \otimes G) W$.
- Under suitable conditions this estimator is consistent, asymptotically normal, and asymptotically efficient in the class of all estimators that do not use any extra information aside from that contained in the moment conditions.

Second step estimator

- The optimal GMM estimator of δ for $N \to \infty$ and T fixed using only the moment restrictions $\mathrm{E}(W_i \Delta\nu_i) = 0$ yields the same expression as in (5) except that

$$W^\top (I_N \bigotimes G) W = \sum_{i=1}^{N} W_i^\top G W_i$$

is replaced by

$$V_n = \sum_{i=1}^{N} W_i^\top (\Delta\nu_i)(\Delta\nu_i)^\top W_i.$$

- The replacement is not surprising in the sense that V_N can serve as an estimator of $W^\top (I_N \otimes G) W$.

Second step estimator

- This GMM estimator requires no knowledge concerning the initial conditions (matrix G) or the distributions of ν_i and μ_i.
- To operationalize this estimator, $\Delta\nu_i$ is replaced by differenced residuals obtained from the preliminary consistent estimator $\hat{\delta}_1$.
- The resulting estimator is the two-step Arellano and Bond GMM estimator

$$\hat{\delta}_2 = \left[(\Delta y_{-1})^\top W \hat{V}_N^{-1} W^\top (\Delta y_{-1})\right]^{-1} \left[(\Delta y_{-1})^\top W \hat{V}_N^{-1} W^\top (\Delta y)\right]. \quad (6)$$

- Note that $\hat{\delta}_2$ and $\hat{\delta}_1$ are asymptotically equivalent if the $\nu_{i,t}$ are i.i.d. with moments $(0, \sigma_\nu^2)$.

Testing for individual effects

- We will employ the simple autoregressive model (3)

$$y_{i,t} = \delta y_{i,t-1} + u_{i,t},\ i = 1, \dots, N;\ t = 1, \dots, T.$$

- Assume there are only three periods $T = 3$.
- Under the null hypothesis of no individual effects, the following orthogonality conditions hold

$$\mathrm{E}(y_{i,2}u_{i,3}) = 0 \quad \mathrm{E}(y_{i,1}u_{i,3}) = 0 \quad \mathrm{E}(y_{i,1}u_{i,2}) = 0.$$

Testing for individual effects

- Three conditions to identify one parameter, the remaining two over-identifying restrictions can be used to test for individual effects.
- We can reformulate these orthogonality restrictions as follows

$$E[y_{i,1}(u_{i,3} - u_{i,2})] = 0, \quad (7a)$$

$$E(y_{i,1}u_{i,2}) = 0, \quad (7b)$$

$$E(y_{i,2}u_{i,3}) = 0. \quad (7c)$$

- The first restriction can be used to identify δ even if there are individual effects in (3).
- The null hypothesis of no individual effects imposes only two additional restrictions (7b) and (7c).

Testing for individual effects

- Let us write
$$y^* = Y^*\delta + u^*, \tag{8}$$
where $y^{*\top} = (y_3^{*\top} - y_2^{*\top}, y_3^{*\top}, y_2^{*\top})$,
$Y^{*\top} = (y_2^{*\top} - y_1^{*\top}, y_2^{*\top}, y_1^{*\top})$ and $u^{*\top} = (u_3^{*\top} - u_2^{*\top}, u_3^{*\top}, u_2^{*\top})$.
- Because of dynamic structure in our model we will once more employ instrumental variables.
- Instrumental variables in one period may not qualify in earlier periods.
- Let $W = \text{diag}[W_i]$ for $i = 1, 2, 3$ be the matrix of instruments such that $\text{Plim}_{N\to\infty}(\frac{W^\top u^*}{N}) = 0$.

Testing for individual effects

- Perform GLS on (8) after premultiplying by $W^\top$.
- $\hat{u}_r^*$ denotes residuals from estimation of each equation separately.
- In this case, $\Omega = W^\top \mathrm{E}(u^* u^{*\top}) W$ is estimated by $\hat{\Omega} = \left(\sum_{i=1}^N \hat{u}_{i,r}^* \hat{u}_{i,s}^* W_{i,r}^\top W_{i,s}\right)$.
- The estimator of δ is

$$\hat{\delta} = \left[Y^{*\top} W \hat{\Omega}^{-1} W^\top Y^* \right]^{-1} \left[Y^{*\top} W \hat{\Omega}^{-1} W^\top y^* \right].$$

Testing for individual effects

- Let SSQ be the weighted sum of the squared transformed residuals

$$SSQ = \frac{(y^* - Y^*\hat{\delta})^\top W \hat{\Omega}^{-1} W^\top (y^* - Y^*\hat{\delta})}{N}.$$

- This has χ^2 distribution with degrees of freedom equal to the number of over-identifying restrictions as N grows.
- The test statistic

$$L = SSQ_R - SSW,$$

where SSQ_R is the sum of squared residuals when imposing the full set of orthogonality conditions implied by the null hypothesis, SSW is the sum of squared residuals that impose only those restrictions needed for the first differenced version (7a).

Testing for individual effects

- L has a χ^2 distribution with degrees of freedom equal to degrees of freedom of SSQ_R minus degrees of freedom of SSW.
- In our case we have three equations and one if we do not employ null hypothesis. I.e., 2 degrees of freedom for L.
- The same estimate of Ω should be used in both computations, and Ω should be estimated under the null hypothesis.
- When calculating parameters and test statistics under the alternative hypothesis the inverse of the submatrix corresponding to the differenced equations only is employed.

Models with Exogenous Variables

- Let us return to the model

$$y_{i,t} = \delta y_{i,t-1} + x_{i,t}^{\top}\beta + u_{i,t},\ i = 1, \dots, N; t = 1, \dots, T,$$

  $u_{i,t} = \mu_i + \nu_{i,t}$.
- If $\mathrm{E}(x_{i,t}\nu_{i,s}) = 0$ for all $t, s = 1, 2, \dots, T$, but where all the $x_{i,t}$ are correlated with μ_i, then all the $x_{i,t}$ are valid instruments for the first-differenced equation of our model.
- Therefore, $[x_{i,1}^{\top}, x_{i,2}^{\top}, \dots, x_{i,T}^{\top}]$ should be added to each diagonal element of W_i.

Models with Exogenous Variables

- If $\mathrm{E}(x_{i,t}\nu_{i,s}) \neq 0$ for $s < t$ and zero otherwise, then only $[x_{i,1}^{\top}, x_{i,2}^{\top}, \ldots, x_{i,s-1}^{\top}]$ should be added to each diagonal element of W_i.
- In both cases we estimate the model as in (4)

$$W^{\top}\Delta y = W^{\top}(\Delta y_{-1})\delta + W^{\top}(\Delta X)\beta + W^{\top}\delta\nu,$$

  where ΔX is the stacked $N(T-2) \times K$ matrix of observations on $\Delta x_{i,t}$.
- Further we continue in a same fashion as for (5) and (6).

Outline

Data and model

- We will estimate a dynamic demand model for cigarettes based on panel data from 46 American states.
- The data are updated from 1963–92.
- The estimated equation is

$$\ln C_{i,t} = \alpha + \beta_1 \ln C_{i,t-1} + \beta_2 \ln P_{i,t} + \beta_3 \ln Y_{i,t} + \beta_4 \ln Pn_{i,t} + u_{i,t},$$

where the subscript i denotes the ith state ($i = 1, \ldots, 46$) and the subscript t denotes the tth year ($t = 1, \ldots, 30$).

Variables

- $C_{i,t}$ is real per capita sales of cigarettes by persons of smoking age (14 years and older). This is measured in packs of cigarettes per head.
- $P_{i,t}$ is the average retail price of a pack of cigarettes measured in real terms.
- $Y_{i,t}$ is the average retail price of a pack of cigarettes measured in real terms.
- $Pn_{i,t}$ denotes the minimum real price of cigarettes in any neighboring state (smuggling effect).

Error term

- The disturbance term is specified as a two-way error component model

$$u_{i,t} = \mu_i + \lambda_t + \nu_{i,t} \quad i = 1, \ldots, 46; \quad t = 1, \ldots, 30.$$

- μ_i denotes a state-specific effect.
- λ_t denotes a year-specific effect.
- The time-period effects are assumed fixed parameters to be estimated as coefficients of time dummies for each year in the sample.

Policy interventions and health warnings

1. The imposition of warning labels by the Federal Trade Commission effective January 1965.
2. The application of the Fairness Doctrine Act to cigarette advertising in June 1967, which subsidized antismoking messages from 1968 to 1970.
3. The Congressional ban on broadcast advertising of cigarettes effective January 1971.

State specific effect

1. States with Indian reservations like Montana, New Mexico and Arizona are among the biggest losers in tax revenues from non-Indians purchasing tax-exempt cigarettes from the reservations.
2. Florida, Texas, Washington and Georgia are among the biggest losers of revenues due to the purchasing of cigarettes from tax-exempt military bases in these states.
3. Utah, which has a high percentage of Mormon population (a religion which forbids smoking), has a per capita sales of cigarettes in 1988 of 55 packs, a little less than half the national average of 113 packs.
4. Nevada, which is a highly touristic state, has a per capita sales of cigarettes of 142 packs in 1988, 29 more packs than the national average.

Table

Variable	$\ln C_{i,t-1}$	$\ln P_{i,t}$	$\ln Y_{i,t}$	$\ln Pn_{i,t}$
OLS	0.97	-0.090	0.024	-0.03
Within	0.83	-0.299	0.034	0.10
2SLS	0.85	-0.205	0.052	-0.02
GMM-one-step	0.84	-0.377	-0.016	0.14
GMM-two-step	0.80	-0.379	-0.020	0.24

Comments

- OLS, which ignores the state and time effects, yields a low short-run price elasticity of -0.09. However, the coefficient of lagged consumption is 0.97.
- Tests emphasize the importance of including state and time effects in the cigarette demand equation.
- GMM-one-step and GMM-two-step (Arellano and Bond) give very similar results.
- GMM-one-step and GMM-two-step have the highest values of $\ln P_{i,t}$.

Bibliography

D. Holz-Eakin
Testing for individual effects in dynamic models using panel data.
Journal of Econometrics, 1988.

B. H. Baltagi
Econometric analysis of panel data.
John Wiley & Sons, 2005.

End

Time for your questions

End

Thank you for your attention