

## **Forecasting with dynamic spatial panel data: practical implementation methods**

**Bernard Fingleton\***

*University of Cambridge, UK*

*Received: 8 October 2014  
Revised: 5 December 2014  
Accepted: 6 December 2014*

---

### **Abstract**

This paper addresses the issue of how in practice one might predict given that the data generating process is a dynamic spatial panel model. One period ahead predictions are calculated using various alternative predictors, some of which are misspecified, and their relative performance evaluated.

*Keywords:* dynamic spatial panel, prediction, random effects, Monte Carlo simulation

*JEL Classification Codes:* C33, C53

---

### **1. Introduction**

Prediction based on spatial panel models has recently come to the forefront as a technically feasible and useful contribution to spatial economics and regional science. Baltagi, Bresson and Pirotte (2012) write about forecasting with spatial panel data with spatial error correlation where the DGP is a simple error components regression model with an autoregressive or moving average spatial dependence process. Monte Carlo analysis shows that the “dynamic” predictor performs well in comparison to predictions which ignore the endogeneity, the spatial correlation in the disturbances and/or ignore individual heterogeneity. Baltagi, Fingleton and Pirotte (2013) consider a dynamic specification with spatial dependence coming from two sources, an endogenous spatial lag and an autoregressive error process, and give the appropriate linear predictor based on Chamberlain (1984) and Sevestre and Trognon (1996). In this paper various prediction equations are compared given dynamic and static spatial panel data generating processes.

---

\* E-mail: bf100@cam.ac.uk.

Citation: Fingleton, B. (2014) Forecasting with dynamic spatial panel data: practical implementation methods, *Economics and Business Letters*, 3(4), 194-207.

## 2. The spatial panel data generating process

The (dynamic) data generating process (DGP) is

$$y_{it} = \gamma y_{it-1} + \rho_1 \sum_{j=1}^N w_{ij} y_{jt} + x_{it} \beta + \varepsilon_{it} \quad i = 1, \dots, N; t = 1, \dots, T, \quad (1)$$

in which  $y_{it}$  denotes the value of the dependent variable for individual (location)  $i$  at time  $t$ , and  $x_{it}$  is the  $(1 \times K)$  vector of exogenous variables with  $(K \times 1)$  coefficient vector  $\beta$ . Also  $w_{ij}$  is the value in the  $i$ 'th row and  $j$ 'th column of predetermined matrix  $W_N$  of known spatial weights with dimension  $(N \times N)$ . Conventionally this matrix has 0s on the main diagonal and  $\gamma$  and  $\rho_1$  are scalar parameters to be estimated. Observe that there are some conditions defining an appropriate parameter space. Thus  $(I_N - \rho_1 W_N)$  with  $I_N$  the  $(N \times N)$  identity matrix must be non-singular and therefore have a non-zero determinant, which is the case when  $\rho_1$  is within the interval  $\left] \frac{1}{r_{min}}, \frac{1}{r_{max}} \right[$  where  $r_{min}$  is the most negative real characteristic root of  $W_N$  and  $r_{max}$  is the maximum and with  $W_N$  typically row normalised so that each row sums to 1,  $r_{max} = 1$ . Similarly stationarity requires that  $|\gamma| < 1$  and dynamic stability is achieved if the largest absolute eigenvalue of  $(I_N - \rho_1 W_N)^{-1} \gamma$  is smaller than one.

The random error process also involves a contemporaneous spatial autoregressive mechanism such that

$$\varepsilon_{it} = \rho_2 \sum_{j=1}^N m_{ij} \varepsilon_{jt} + u_{it} \quad (2)$$

in which  $m_{ij}$  is equivalent to, but not necessarily the same as,  $w_{ij}$  and accordingly matrix  $M_N$  has the same properties as matrix  $W_N$ . The conditions for  $\rho_2$  are equivalent to those for  $\rho_1$  but with respect to the real eigenvalues of  $M_N$  not  $W_N$  (which could be identical). The remainder  $u_{it}$  has two components, one ( $\mu_i$ ) is time invariant and individual-specific and the other transient component is denoted by  $v_{it}$ . Thus

$$u_{it} = \mu_i + v_{it}, \quad (3)$$

and it is assumed that each component is independently identically distributed so that  $\mu \sim iid(0, \sigma_\mu^2)$  and the remainder  $v$  is distributed as  $(0, \sigma_v^2)$ . Also both components are internally independent and independent of each other.

This equation is equivalent to a recurrent equation in matrix form, thus

$$y_t = \gamma y_{t-1} + \rho_1 W_N y_t + x_t \beta + \varepsilon_t \quad (4)$$

with

$$\varepsilon_t = \rho_2 M_N \varepsilon_t + u_t, \quad (5)$$

$$u_t = \mu + v_t. \quad (6)$$

So that

$$y_t = G_N^{-1} [\gamma y_{t-1} + x_t \beta + B_N^{-1} u_t] \quad (7)$$

in which  $y_t$  is an  $(N \times 1)$  vector,  $x_t$  is an  $(N \times K)$  matrix of exogenous explanatory variables,  $u_t$  is the  $(N \times 1)$  remainder term,  $\varepsilon_t$  is an  $(N \times 1)$  vector of errors,  $G_N = (I_N - \rho_1 W_N)$  and  $B_N = (I_N - \rho_2 M_N)$ . In order to obtain a simulated series consistent with the assumed DGP this can be solved recursively for  $t = 1, \dots, T + \tau$  starting with  $y_0$  and given  $\mu$  and  $v_t$ . To enable realisations of this process, in this instance the spatial matrix "j ahead j behind" given by Kelejian-Prucha (1999) is adopted, where  $j$  is a maximum of 5. The result is an  $(N \times N)$  matrix with non-zero weights equal to 0.1 summing to 1 both across rows and down columns. For simplicity, it is assumed throughout that  $M_N = W_N$ .

One approach to obtaining initial values  $y_0$ , following Baltagi, Fingleton and Pirotte (2013), is based on realised values of the unobservable individual effects  $\mu$  and  $\nu$ . In order to obtain such values, we draw at random from assumed normal distributions, where  $\mu \sim iid(0, \sigma_\mu^2)$  and  $\nu \sim iid(0, \sigma_\nu^2)$ , thus obtaining  $(N \times 1)$  vectors  $\mu$  and  $\nu_0$ . Then<sup>1</sup>

$$y_0 = \frac{1}{1-\gamma} P_N \mu + \frac{1}{\sqrt{1-\gamma^2}} P_N \nu_0. \quad (8)$$

Also with  $P_N = [B_N G_N]^{-1}$  the variance-covariance matrix is

$$V[y_0] = \left( \frac{\sigma_\mu^2}{(1-\gamma)^2} + \frac{\sigma_\nu^2}{1-\gamma^2} \right) P_N P_N'. \quad (9)$$

The lagged endogenous variable is correlated with the individual effects. So, in order to obtain individual effects  $\mu_j^* = E^*[\mu_j | y_0']$ ,  $j = 1, \dots, N$ , one calculates

$$E[y_0 \mu_j] = \frac{\sigma_\mu^2}{1-\gamma} P_N l_j. \quad (10)$$

in which  $l_j$  is the  $j$ 'th column of the  $(N \times N)$  identity matrix and

$$\lambda_j = V[y_0]^{-1} E[y_0 \mu_j]. \quad (11)$$

So that

$$\mu_j^* = \lambda_j' [y_0 - E[y_0]]. \quad (12)$$

The DGP can produce data consistent with either a dynamic process, in which  $\gamma \neq 0$ , or a static process where  $\gamma = 0$ . For the dynamic process, it is assumed that  $\gamma = 0.5$ ,  $\rho_1 = 0.333$ ,  $\rho_2 = 0.25$ ,  $\beta = 0.5$ ,  $\sigma_\mu^2 = 0.2$  and  $\sigma_\nu^2 = 0.04$ . The data series is generated by solving (7) recursively over  $t = 1, \dots, T + \tau$  with  $u_t = \mu^* + \nu_t$  and  $T = 11$ . In this it is assumed that  $x_{-10}$  is an  $(N \times 1)$  vector of 0s and discarded the first 10 observations, and  $x_t = 0.9x_{t-1} + e_t$ ; with  $e_t$  drawn at random from an  $N(0,1)$  distribution. For the static process, we maintain the same assumptions with the exception that  $\gamma = 0$ . These assumptions are retained throughout the simulations used in the paper, and so the results are necessarily conditional on these.

### 3. Dynamic panel estimation

Given data across  $N$  locations and  $T$  times, the estimator for this dynamic specification given by Baltagi, Fingleton and Pirotte (2013) has been shown to outperform rival estimators. This estimator follows the path of Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991) which by differencing eliminates the individual effects  $\mu_i$  and which adopts a GMM approach with, in this case, appropriate spatial and non-spatial instruments satisfying moments conditions. This provides consistent initial estimates of  $\gamma$ ,  $\beta$  and  $\rho_1$  leading to consistent residuals which are the basis of the estimated  $\rho_2$ ,  $\sigma_\mu^2$  and  $\sigma_\nu^2$ . To obtain the latter, a modified version of the Kapoor, Kelejian and Prucha (2007) is adopted, so that consistent estimates are obtained of the autoregressive parameter of the error process and error component variances. In the final stages account is taken of the error process dependence leading to a final two-step spatial GMM estimator of  $\gamma$ ,  $\beta$  and  $\rho_1$ .

<sup>1</sup> This specification implicitly assumes that the contribution of initial values of the explanatory variable is zero and that the DGP of  $y_0$  is the same as the one of  $y_t$ ,  $t > 0$ .

Given a single realisation of the dynamic DGP, using the data for  $t = 1, \dots, T$ , and hence not basing estimation on the final realisation  $T + \tau$ , a typical outcome using the Baltagi, Fingleton and Pirotte (2013) estimator is as in Table 1.

Table 1. Parameter estimates: dynamic DGP and dynamic estimator

Parameters	Estimates	Standard errors	<i>t</i> -ratios
$\gamma = 0.5$	0.4992	0.0069	72.3858
$\rho_1 = 0.333$	0.3591	0.0315	11.3867
$\beta = 0.5$	0.4988	0.0055	90.7613
$\rho_2 = 0.25$	0.2155		
$\sigma_\mu^2 = 0.2$	0.2250		
$\sigma_v^2 = 0.04$	0.0899		

With the static DGP, it is anticipated that  $\hat{\gamma} \approx 0$ . As an illustration, Table 2 gives the outcome for a single realisation of the (static) DGP.

Table 2. Parameter estimates: static DGP and dynamic estimator

Parameters	Estimates	Standard errors	<i>t</i> -ratios
$\gamma = 0$	0.0020	0.0134	0.1482
$\rho_1 = 0.333$	0.3444	0.0436	7.8913
$\beta = 0.5$	0.5059	0.0073	68.8296
$\rho_2 = 0.25$	0.5935		
$\sigma_\mu^2 = 0.2$	0.1567		
$\sigma_v^2 = 0.04$	0.0376		

#### 4. Static panel estimation

We are generating both static and dynamic spatial panel data, and therefore a brief outline is given of the static panel estimator ( $\gamma = 0$ ) with a spatial lag ( $\rho_1 \neq 0$ ) and spatial error process  $\rho_2 \neq 0$ . As with the dynamic estimator the core of the method is Kapoor, Kelejian and Prucha et al. (2007), although their approach does not include the endogenous spatial lag, focusing on the spatial error process. It is a simple step to introduce the spatial lag into the estimation procedure, as outlined in various papers (Fingleton, 2008; Baltagi, Fingleton and Pirotte, 2011). There are three stages. First, given the presence of the endogenous spatial lag, instruments are used to give consistent initial estimates of  $\beta$  and  $\rho_1$  leading to consistent residuals. Second, these are the basis of the estimated  $\rho_2$ ,  $\sigma_\mu^2$  and  $\sigma_v^2$  via GMM. Third, these parameter estimates then allow elimination of error dependence via a Cochrane-Orcutt transformation, leading to final estimates of  $\beta$  and  $\rho_1$  and appropriate standard errors.

Given data for  $t = 1, \dots, T$  produced by a dynamic DGP, one expects the parameter estimates to be biased. In contrast with data produced via a static DGP process, the static panel estimator will typically lead to unbiased estimation. Typical outcomes for single realisations of the dynamic and static DGPs are presented in Tables 3 and 4.

Table 3. Parameter estimates: dynamic DGP and static estimator

Parameters	Estimates	Standard errors	t-ratios
$\gamma = 0.5$	-	-	-
$\rho_1 = 0.333$	0.5065	0.0671	7.5445
$\beta = 0.5$	0.8810	0.0356	24.7775
$\rho_2 = 0.25$	0.2770	0.0834	3.2602 *
$\sigma_\mu^2 = 0.2$	0.8969		
$\sigma_\nu^2 = 0.04$	0.2680		

\* Based on 100 Bootstrap replications, equal to estimate minus Bootstrap mean in units of standard deviation.

Table 4. Parameter estimates: static DGP and static estimator

Parameters	Estimates	Standard errors	t-ratios
$\gamma = 0$	-	-	-
$\rho_1 = 0.333$	0.3715	0.0699	5.3180
$\beta = 0.5$	0.5090	0.0153	33.2112
$\rho_2 = 0.25$	0.2997	0.0776	3.9111 *
$\sigma_\mu^2 = 0.2$	0.1834		
$\sigma_\nu^2 = 0.04$	0.0394		

\* Based on 100 Bootstrap replications, equal to estimate minus Bootstrap mean in units of standard deviation.

## 5. Prediction for dynamic and static panel data

Prediction is carried out using 5 methods, labelled A to E. Method A comes from the published literature, but relies on knowledge of normally unknown initial values. Method B approximates the initial values by the values observed at time 1. Method C also uses the observed data for initial values, but in addition attempts to estimate individual effects  $\mu$  from the residuals. Method D is a prediction method appropriate to static panel data, as is method E, but E is minus the contribution due to interdependence of disturbances.

### Prediction Method A

The source of this method is Baltagi, Fingleton and Pirotte (2013), which is derived from Chamberlain (1984) and Sevestre and Trognon (1996). The approach uses the same equations as used in the dynamic DGP and relies on the same initial predetermined values across individuals  $y_0 = (y_{10}, \dots, y_{N0})$  leading to an estimate of  $\mu_j^* = \lambda_j' [y_0 - E[y_0]]$ . The linear predictor  $y_{it}^*$  of  $y_{it}$  is conditional on  $y_0$  and on the whole sequence of an exogenous variable<sup>2</sup>  $(x_{11}, \dots, x_{N1}, \dots, x_{1T}, \dots, x_{NT})$ , and is given by

$$y_{it}^* = \gamma^t \sum_{j=1}^N h_{ij}^{(t)} y_{j0} + \sum_{l=1}^t \gamma^{l-1} \sum_{j=1}^N h_{ij}^{(l)} x_{jt-l+1} \beta + \sum_{l=1}^t \gamma^{l-1} \sum_{j=1}^N p_{ij}^{(l)} \mu_j^* \quad (13)$$

in which  $h_{ij}^{(l)}$  denotes cell  $(i,j)$  of matrix  $(G_N^{-1})^l$  and  $p_{ij}^{(l)}$  is cell  $(i,j)$  of  $((G_N^{-1})^l B_N^{-1})$ . Given estimates of  $\gamma$ ,  $\rho_1$ ,  $\rho_2$  and  $\beta$ , we can proceed to calculate  $y_{it}^*$ , given  $y_{j0}$  and  $\mu_j^*$ . The assumption is that initial values  $y_{j0}$  and individual effect  $\mu_j$  is uncorrelated with the explanatory variable  $x_{it}$ ,  $\forall i, t$ .

<sup>2</sup> For ease of exposition, we consider just a single exogenous variable.

Therefore to proceed, the estimates (similar to Table 1 but re-estimated for each Monte Carlo simulation) are substituted into the linear predictor

$$\hat{y}_t^* = \hat{G}_N^{-1} [\hat{\gamma} \hat{y}_{t-1}^* + x_t \hat{\beta} + \hat{B}_N^{-1} \hat{\mu}^*] \tag{14}$$

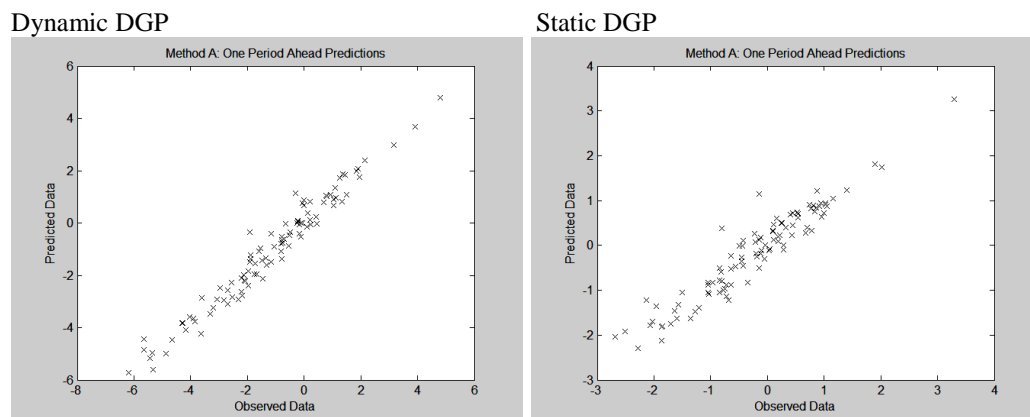
which is solved for  $t = 1, \dots, T + \tau; \tau = 1$ , where  $\hat{y}_0^* = y_0, \hat{G}_N = (I_N - \hat{\rho}_1 W_N),$

$\hat{B}_N = (I_N - \hat{\rho}_2 M_N)$  and with  $\hat{P}_N = [\hat{B}_N \hat{G}_N]^{-1}$ , this gives

$$\hat{\lambda}_j = \left[ \left( \frac{\hat{\sigma}_\mu^2}{(1-\hat{\gamma})^2} + \frac{\hat{\sigma}_v^2}{(1-\hat{\gamma}^2)} \right) \hat{P}_N \hat{P}_N' \right]^{-1} \frac{\hat{\sigma}_\mu^2}{1-\hat{\gamma}} \hat{P}_N l_j \text{ leading to } \hat{\mu}_j^* = \hat{\lambda}_j' [y_0 - E[y_0]].$$

For a single realization, and for  $\tau = 1$ , in Figure 1 a comparison is made of the resulting  $(N \times 1)$  vectors with the vectors created by the dynamic and static DGPs. Table 5 summarises Method A prediction errors at  $T + \tau; \tau = 1$  from 500 replications.

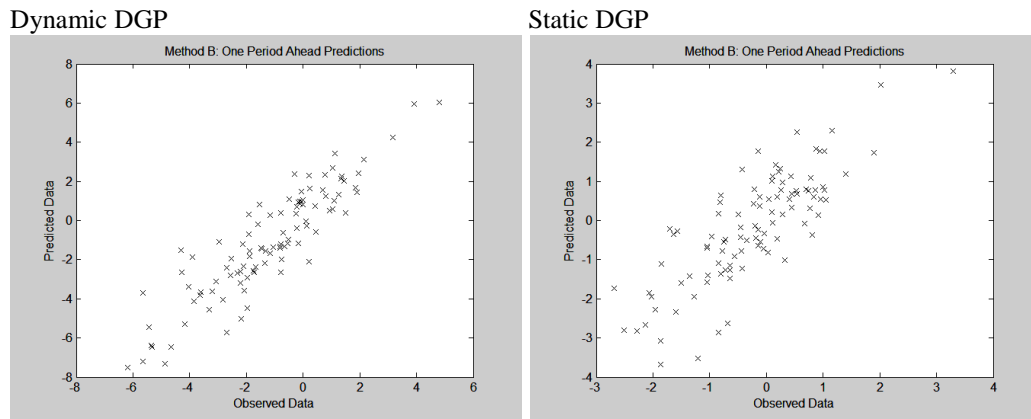
Figure 1. Prediction method A: one period ahead predictions



*Prediction Method B*

This is identical to prediction method A except that instead of using  $y_0$ , which in practice will be unknown, one can use  $y_1$ , the observed data at time 1, assuming it will be similar to  $y_0$ . Thus one can solve equation (14) with  $\hat{y}_0^* = y_1$  as the initial  $y$  observation. Debarsy, Ertur and LeSage (2012: pp. 161) argue that it is reasonable to condition on the initial period, since their focus is on “interpretation not estimation of these models”, and likewise here one is not concerned with parameter estimation, but with interpretation based on extant parameter estimates. For a single replication, and for time  $T + \tau; \tau = 1$ , Figure 2 plots the resulting  $(N \times 1)$  vectors against the DGP vectors, showing a relatively poor performance for these single realizations. Table 5 summarises prediction errors for method B at  $T + \tau; \tau = 1$  from 500 replications.

Figure 2. Prediction method B: one period ahead predictions



Prediction method C

Method B uses proxy initial values  $y_1$  so that  $\hat{\mu}_j^* = \hat{\lambda}'_j [y_1 - E[y_1]]$ ,  $\forall j$ . In method C an attempt is made to estimate individual effects  $\mu$  from the residuals observed over time. To show this we commence with the single cross-section

$$y_t = \gamma y_{t-1} + \rho_1 W_N y_t + x_t \beta + \varepsilon_t. \tag{16}$$

So that

$$B_N^{-1} u_t = y_t - \gamma y_{t-1} - \rho_1 W_N y_t - x_t \beta. \tag{17}$$

and

$$u_t = \mu + v_t = B_N [y_t - \gamma y_{t-1} - \rho_1 W_N y_t - x_t \beta], \tag{18}$$

$$\mu = B_N [G_N y_t - \gamma y_{t-1} - x_t \beta] - v_t, \tag{19}$$

$$v_t \sim N(0, \sigma_v^2). \tag{20}$$

In order to calculate  $\hat{\mu}$ , one uses observed data (from the DGP) for the sequence of  $y$ s in equation (19) together with the parameter estimates ( $\hat{\gamma}, \hat{\beta}$ , etc.), using the data for the period  $t = 2, \dots, T$ , on each occasion drawing an  $(N \times 1)$  vector  $v_t$  at random from the  $N(0, \hat{\sigma}_v^2)$  distribution. This gives  $T - 1$  (different) estimates of  $\mu$ , so we take the across time mean as an estimate of the time-invariant  $(N \times 1)$  vector  $\mu$ , given  $E[v_t] = 0$ . Also the estimate is scaled so that its variance is equal to  $\hat{\sigma}_\mu^2$ .

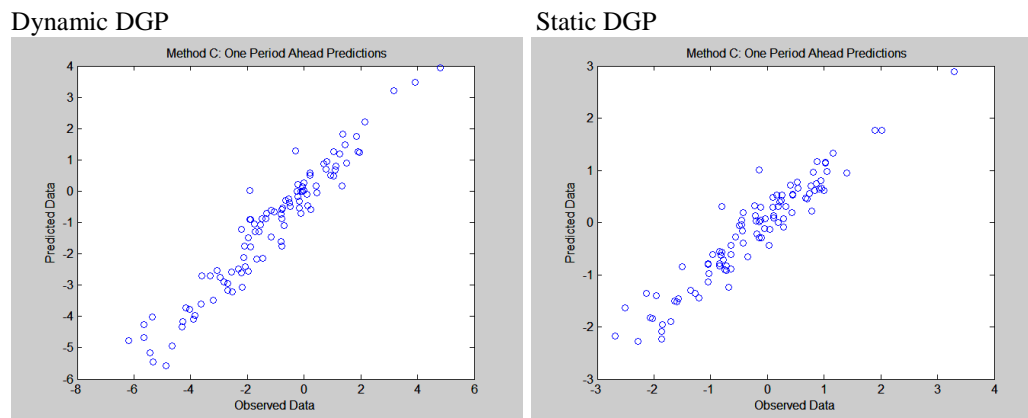
This estimated  $\mu$ , denoted by  $\bar{\mu}$ , is then used in the recurrent equation with initial observed values  $y_1$  thus

$$\hat{y}_t = \hat{G}_N^{-1} [\hat{\gamma} \hat{y}_{t-1} + x_t \hat{\beta} + \hat{B}_N^{-1} \bar{\mu}] \tag{21}$$

which solves recursively over  $t = 2, \dots, T + \tau$ .

Figure 3 shows that the approach works well for the dynamic DGP, and also for the static DGP. Of course these are single realizations, so one should look at 500 replications (Tables 5 and 6) to confirm these initial observations.

Figure 3. Prediction method C: one period ahead predictions



### Prediction method D

The static model prediction equation is the BLUP predictor of Goldberger (1962), which for  $\gamma = \rho_1 = 0$  is given by

$$\hat{y}_{i,T+\tau} = x_{i,T+\tau} \hat{\beta} + \frac{\hat{\sigma}_\mu^2}{\hat{\sigma}_1^2} (\iota'_T \otimes l'_i) \hat{\varepsilon} \quad (22)$$

where  $x_{i,T+\tau}$  is the  $(1 \times K)$  vector of explanatory variable values for individual  $i$  at time  $T + \tau$ ,  $\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2$  and  $l_i$  is the  $i$ 'th column of  $I_N$ , where  $\sigma_\mu^2$  is the variance of the time-constant random process for individuals and  $\sigma_v^2$  is the variance of the idiosyncratic errors varying across individuals and across time. This means that for individual  $i$ , one adds  $(T\hat{\sigma}_\mu^2/\hat{\sigma}_1^2)\bar{\varepsilon}_i$  where  $\bar{\varepsilon}_i = \sum_{t=1}^T \varepsilon_{it}/T$ . Thus the usual GLS forecast is modified by adding a fraction of mean of the GLS residuals corresponding to the  $i$ 'th individual. Interestingly, Baltagi, Bresson and Pirotte (2012) show that this also applies even if  $\rho_2 \neq 0$  but where  $\gamma = \rho_1 = 0$ .

In the case where  $\rho_1 \neq 0$ , following Fingleton (2009) and Baltagi, Fingleton and Pirotte (2011) the linear predictor becomes

$$\hat{y}_{T+\tau} = \hat{G}_N^{-1} \left( x_{T+\tau} \hat{\beta} + \frac{\hat{\sigma}_\mu^2}{\hat{\sigma}_1^2} (\iota'_T \otimes I_N) \hat{\varepsilon} \right) \quad (23)$$

in which  $\iota_T$  is a  $(T \times 1)$  vector of 1s. This<sup>3</sup> is equal to

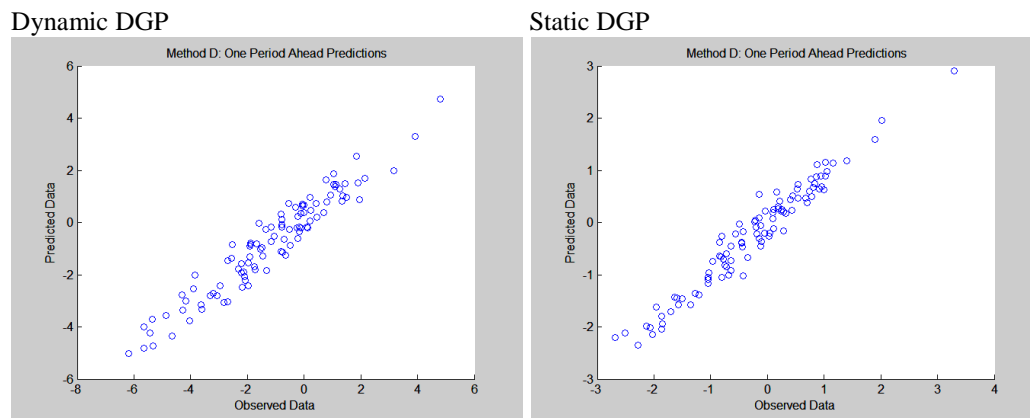
$$\hat{y}_{i,T+\tau} = \sum_{k=1}^K \hat{\beta}_k \sum_{j=1}^N \hat{h}_{ij} x_{kj,T+\tau} + \frac{T\hat{\sigma}_\mu^2}{\hat{\sigma}_1^2} \sum_{j=1}^N \hat{h}_{ij} \hat{\varepsilon}_j \quad (24)$$

in which  $h_{ij}$  is the  $i,j$ 'th element of  $G_N^{-1}$  (Baltagi, Fingleton and Pirotte, 2011, 2013). Given either a dynamic DGP or a static DGP, the static panel estimator for each Monte-Carlo simulation of a dynamic or static process (giving estimates similar to Tables 3 and 4) is applied. The relative performance of (24) compared with other prediction equations is given in Fingleton (2009). Figure 4 suggests (in a single realization) that Method D is (slightly) better at predicting outcomes from a static rather than a dynamic DGP.

<sup>3</sup> A proof that this is a Best Linear Unbiased Predictor (BLUP) is available at [www.spatialeconomics.ac.uk/textonly/SERC/publications/download/sercdp0095.pdf](http://www spatialeconomics.ac.uk/textonly/SERC/publications/download/sercdp0095.pdf).



Figure 4. Prediction method D: one period ahead predictions



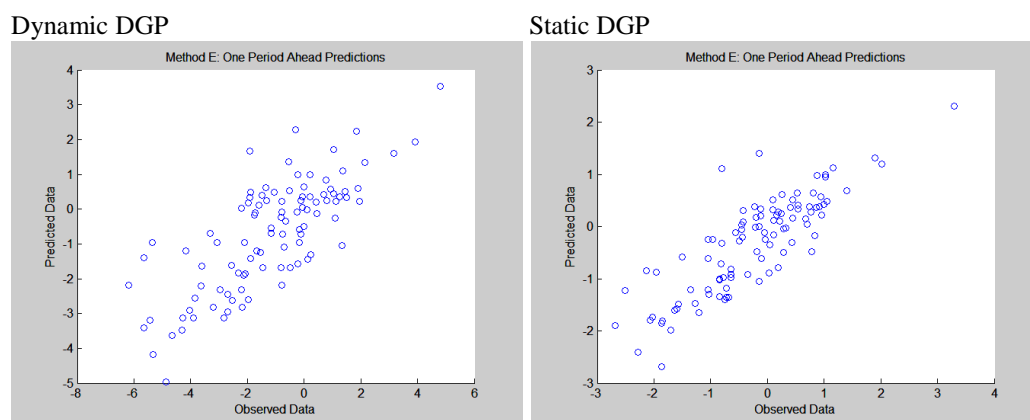
*Prediction method E*

This is simply the static panel estimator minus the Goldberger (1962) correction, hence

$$\hat{y}_{i,T+\tau} = \hat{G}_N^{-1}(x_{i,T+\tau}\hat{\beta}) \tag{25}$$

Comparing Figures 4 and 5 one can see an evident loss of precision due to the absence of the Goldberger correction.

Figure 5: Prediction method E: one period ahead predictions



*Summary of prediction outcomes*

The foregoing comparisons were simply indicative. To provide more valid evidence, Tables 5 and 6 summarise outcomes of 500 Monte-Carlo simulations for each prediction method with both dynamic and static DGP. In the Appendix the complete distribution of the prediction error from the Monte-Carlo simulations is shown, where

$$prediction\ error = \sum_{i=1}^N (y_{i,T+\tau} - y_{i,T+\tau}^*)^2 \tag{26}$$

Table 5 shows that method C provided the most accurate predictions for the data obtained via the dynamic DGP. Method A is not a practical option in that the true initial values are typically unobserved, but it does provide a yardstick for comparison. The contrast between methods A and B shows that proxying the true, but unknown, initial values by the observed

data in this case by itself leads to a severe reduction in accuracy. However this is not strictly due to the initial values used per se, but is attributable to the effect it has on the individual effects. This is shown by the optimal method C, which like B also uses the observed data for the initial values, but which uses residuals to obtain estimates of the individual effects. Methods D and E are relatively poor predictors, since they are misspecified for the dynamic DGP. Despite this, method D is preferable to method B.

Table 5. Summary statistics for out-of-sample prediction errors with Dynamic DGP, 500 replications

Method	Mean	Median	RMSE
A	38.9423	28.9348	0.5925
B	161.128	147.473	1.2556
C	23.8635	22.5812	0.4800
D	55.3232	52.9774	0.7391
E	198.214	191.547	1.3985

Table 6 shows outcomes for the static DGP. Not unexpectedly, method D is clearly the best approach. Interestingly, method C (overparametrized for the static DGP) proved superior to method E, which omits the Goldberger correction.

Table 6. Summary statistics for out-of-sample prediction errors with Static DGP, 500 replications

Method	Mean	Median	RMSE
A	10.1216	9.8257	0.3164
B	60.1664	60.1281	0.7741
C	7.6827	7.4247	0.2756
D	5.3960	5.3775	0.2314
E	30.5929	29.9836	0.5506

## 6. Conclusions

The conclusions are prefixed with the warning that these results are merely an initial, provisional foray into the problem, in that they are conditional on assumed parameter values and interaction matrices used in the DGP and outcomes across a range of alternatives have not yet been explored. For example, it would be interesting to explore the effect of assuming a higher level of temporal persistence, such as  $\gamma = 0.9$ . Thus the contribution is essentially methodological, pointing to the relevant literature and bringing into focus issues of relevance and how one might wish to proceed in studying the problem, rather than being definitive or final interpretations. With this important caveat in mind, it is suggested that given data of unknown provenance, one should estimate the dynamic model in order to test the null hypothesis  $\gamma = 0$  and thus see whether the dynamic specification is appropriate. If it is, then it seems that the best approach appears to be Method C. However if we fail to reject  $\gamma = 0$  then Method D is the best.

Note however that for a dynamic DGP, Method C cannot always be guaranteed to perform better than Method A, it depends on how closely the assumed initial values  $y_1$  correlate to the unknown initial values  $y_0$ . In the case of the simulations reported here, the correlations are quite high, with a mean equal to 0.67, so that in this instance Method C is more accurate than Method A. However if one maximises the correlation by replacing  $y_1$  by  $y_0$  in Method B, then Methods A and B are effectively identical. Also using  $y_0$  in place of  $y_1$  in Method C produces even superior predictions. In practice, without knowledge of  $y_0$ , one does not know how good the correlation is between  $y_1$  and  $y_0$ . However one way to improve predictions via Methods B and C would be to remain open to optional starting values rather than remain rigidly with  $y_1$ . This could be achieved by holding back data and predicting the data ex post, choosing starting values which optimize the ex post prediction. With those optimal starting values, then one might proceed with greater confidence to make ex ante predictions. However this is beyond the scope of this present paper and we leave this to further research. To summarize, because it uses additional information, Method C appears to be less vulnerable to the influence of poorly correlated starting values than is Method B, so on the current evidence is the preferred choice, although it will not always be as accurate as Method A, which cannot however be used in practice.

*Acknowledgements.* Professor Alain Pirotte provided some suggestions in writing this paper, but I am solely responsible for any errors it may contain.

## References

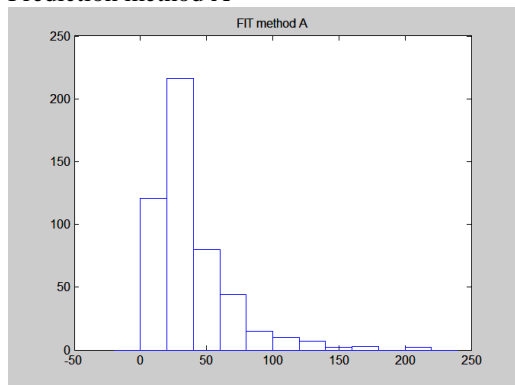
- Anderson, T.W. and Hsiao, C. (1981) Estimation of dynamic models with error components, *Journal of the American Statistical Association*, 76, 598–606.
- Anderson, T.W. and Hsiao, C. (1982) Formulation and estimation of dynamic models using panel data, *Journal of Econometrics*, 18, 47–82.
- Arellano, M. and Bond, S. (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment, *Review of Economic Studies*, 58, 277–297.
- Fingleton, B. (2009) Prediction using panel data regression with spatial random effects, *International Regional Science Review*, 32, 195–220.
- Baltagi, B.H., Fingleton, B. and Pirotte, A. (2013) Estimating and forecasting with a dynamic spatial panel model, *Oxford Bulletin of Economics and Statistics*, 76, 112–138.
- Baltagi, B.H., Fingleton, B. and Pirotte, A. (2011) *Estimating and forecasting with a dynamic spatial panel model*, Spatial Economics Research Centre (LSE), Discussion Paper 95.
- Baltagi, B.H., Bresson, G. and Pirotte, A. (2012) Forecasting with spatial panel data, *Computational Statistics and Data Analysis*, 56, 3381–3397.
- Chamberlain, G. (1984) *Panel data*, in Griliches Z. and Intriligator, M. (eds) *The Handbook of Econometrics*, ch. 22, North-Holland, Amsterdam, 1247–1318.
- Debarsy, N., Ertur, C. and LeSage, J.P. (2012) Interpreting dynamic space–time panel data models, *Statistical Methodology*, 9, 158–171.
- Goldberger, A.S. (1962) Best linear unbiased prediction in the generalized linear regression model, *Journal of the American Statistical Association*, 57, 369–375.
- Kapoor, M., Kelejian, H.H. and Prucha, I.R. (2007) Panel data models with spatially correlated error components, *Journal of Econometrics*, 140, 97–130.

- Kelejian, H.H., Prucha, I.R. (1999) A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review*, 40, 509–533.
- Sevestre, P. and Trognon, A. (1996) *Dynamic linear models*, in Mâtyàs L. and Sevestre P. (eds.) *The Econometrics of Panel Data: A Handbook of Theory with Applications*, ch. 7, Kluwer Academic Publishers, Dordrecht, 121–144.

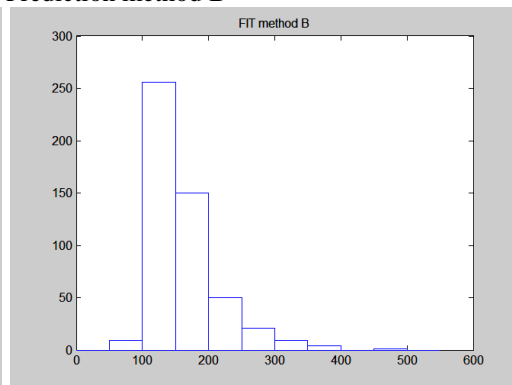
## Appendix

### A1. Distribution of outcomes from Monte-Carlo simulations with dynamic DGP

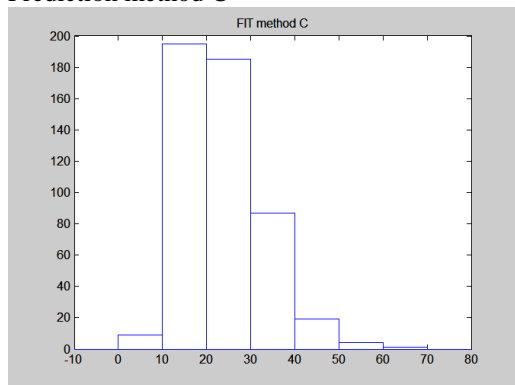
Prediction method A



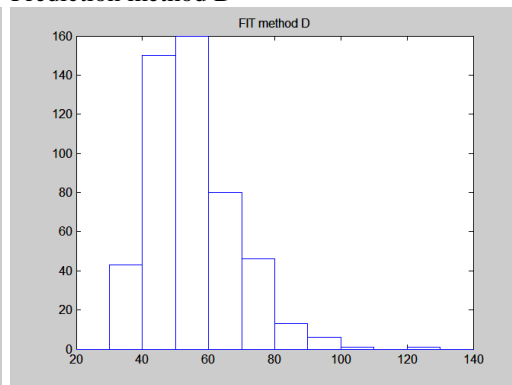
Prediction method B



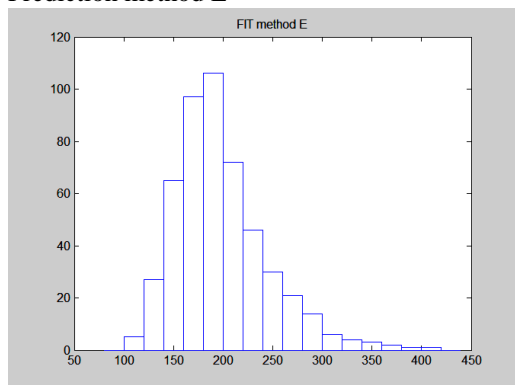
Prediction method C



Prediction method D

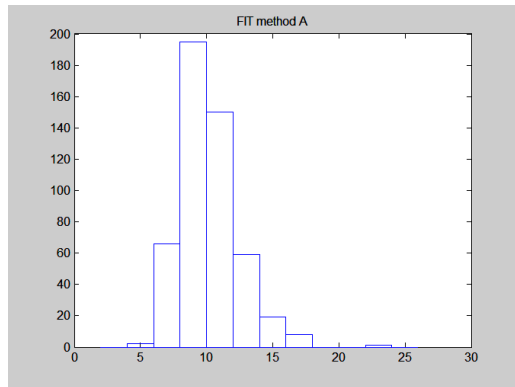


Prediction method E

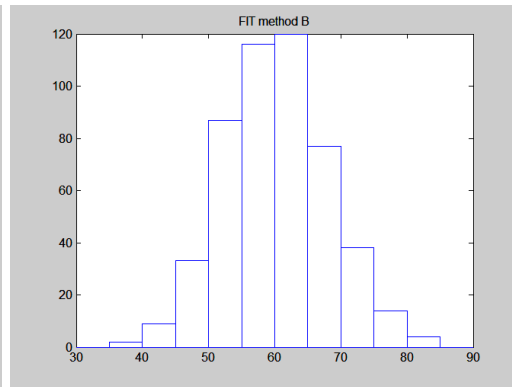


A.2 Distribution of outcomes from Monte-Carlo simulations with static DGP

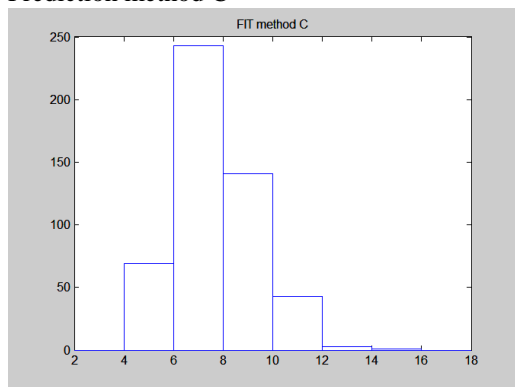
Prediction method A



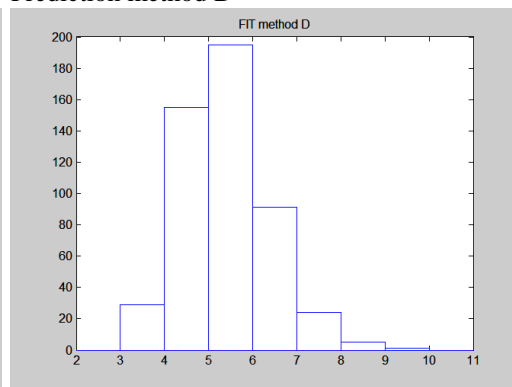
Prediction method B



Prediction method C



Prediction method D



Prediction method E

