Journal of Forecasting
J. Forecast. **27**, 153–173 (2008)
Published online 21 January 2008 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/for.1047

Forecasting with Panel Data

BADI H. BALTAGI*
Department of Economics and Center for Policy Research,
Syracuse University, Syracuse, New York, USA
Department of Economics, University of Leicester, Leicester, UK

ABSTRACT
This paper gives a brief survey of forecasting with panel data. It begins with a simple error component regression model and surveys the best linear unbiased prediction under various assumptions of the disturbance term. This includes various ARMA models as well as spatial autoregressive models. The paper also surveys how these forecasts have been used in panel data applications, running horse races between heterogeneous and homogeneous panel data models using out-of-sample forecasts.

KEY WORDS forecasting; BLUP; panel data; spatial dependence; serial correlation; heterogeneous panels

INTRODUCTION

The literature on forecasting is rich with time series applications, but until recently this has not been the case for panel data applications. Introductory textbooks on forecasting, such as Diebold (2004), have nothing on forecasting with panel data, and there is no paper on this subject in the companion to forecasting edited by Clements and Hendry (2005). This survey is aimed at making a contribution to this literature. It is a humble contribution focusing on what we know about the best linear unbiased predictor (BLUP) in the error component model, one of the most used econometric specifications in applied panel data. Although this has been widely studied in the statistics and biometrics literature, little discussion on this subject appears in the econometrics literature and what is there is scattered in journal articles and book chapters (see Baltagi, 2005). We then survey forecasting applications using panels, and most notably those applications that used forecasting to compare the performance of heterogeneous and homogeneous estimators using post-sample data. This survey has its limitations. It does not get into the large literature on 'forecast combination methods', which can serve as a good springboard to launch research in improving forecasting methods using panels (see Diebold and Lopez, 1996; Newbold and Harvey, 2002; Stock and Watson, 2004; among others). It also does not get into the related literature on 'forecasting economic aggregates from disaggregates' (see Hendry and Hubrich, 2006). This survey also does not do justice to the Bayesian literature on forecasting

* Correspondence to: Badi H. Baltagi, Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020, USA. E-mail: bbaltagi@maxwell.syr.edu
This paper was presented at the 8th Bundesbank Spring Conference on 'New Developments in Economic Forecasting' in Eltville, Germany, 5–6 May, 2006. I would like to thank the editor Massimiliano Marcellino and three anonymous referees for their helpful comments and suggestions.

and how it can improve forecasts using panels (see Zellner and Hong, 1989; Zellner *et al.*, 1991; Nandram and Petrucelli, 1997; Koop and Potter, 2003; Canova and Ciccarelli, 2004; among others). The next section surveys the BLUP in the error component model, while the third section focuses on out-of-sample forecasts comparing the performance of homogeneous and heterogeneous estimators using panel data. The last section recaps the limitations of this survey and suggests future work.

THE BEST LINEAR UNBIASED PREDICTOR

Consider a panel data regression model

$$y_{it} = \alpha + X'_{it}\beta + u_{it} \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{1}$$

with i denoting households, individuals, firms, countries, etc., and t denoting time. The i subscript therefore denotes the cross-section dimension, whereas t denotes the time series dimension; α is a scalar, β is $K \times 1$ and X_{it} is the itth observation on K explanatory variables. Most of the panel data applications utilize a one-way error component model for the disturbances (see Baltagi, 2005) with

$$u_{it} = u_i + \upsilon_{it} \tag{2}$$

where μ_i denotes the *unobservabl*e individual specific effect and v_{it} denotes the remainder disturbance. For example, in an earnings equation in labor economics, y_{it} will measure earnings of the head of the household, whereas X_{it} may contain a set of variables such as experience, education, union membership, sex, and race. Note that μ_i is time-invariant and it accounts for any individual specific effect that is not included in the regression. In this case we could think of it as the individual's unobserved ability. The remainder disturbance υ_{it} varies with individuals and time and can be thought of as the usual disturbance in the regression. This can be written as

$$y = \alpha\iota_{NT} + X\beta + u = Z\delta + u \tag{3}$$

where y is $NT \times 1$, X is $NT \times K$, $Z = [\iota_{NT}, X]$, $\delta' = (\alpha', \beta')$ and ι_{NT} is a vector of ones of dimension NT. Also

$$u = Z_\mu \mu + \upsilon \tag{4}$$

where $u' = (u_{11}, \ldots, u_{1T}, u_{21}, \ldots, u_{2T}, \ldots, u_{N1}, \ldots, u_{NT})$ with the observations stacked such that the slower index is over individuals and the faster index is over time. $Z_\mu = I_N \otimes \iota_T$, where I_N is an identity matrix of dimension N, ι_T is a vector of ones of dimension T and $\otimes$ denotes the Kronecker product. Z_μ is a selector matrix of ones and zeros, or simply the matrix of individual dummies that one may include in the regression to estimate the μ_i if they are assumed to be fixed parameters. $\mu' = (\mu_1, \ldots, \mu_N)$ and $\upsilon' = (\upsilon_{11}, \ldots, \upsilon_{1T}, \ldots, \upsilon_{N1}, \ldots, \upsilon_{NT})$. Note that $Z_\mu Z'_\mu = I_N \otimes J_T$, where J_T is a matrix of ones of dimension T, and $P = Z_\mu(Z'_\mu Z_\mu)^{-1}Z'_\mu$; the projection matrix on Z_μ reduces to $I_N \otimes \bar{J}_T$, where $\bar{J}_T = J_T/T$. P is a matrix which averages the observation across time for each individual, and $Q = I_{NT} - P$ is a matrix which obtains the deviations from individual means. For example, regressing y on the matrix of dummy variables Z_μ gets the predicted values Py, which have a typical element $\bar{y}_i = \Sigma_{t=1}^T y_{it}/T$

DOI: 10.1002/for

repeated T times for each individual. The residuals of this regression are given by Qy, which have a typical element $(y_{it} - \bar{y}_{i\cdot})$.

For the *fixed-effects* case, the μ_i are assumed to be fixed parameters to be estimated and the remainder disturbances stochastic with υ_{it} independent and identically distributed i.i.d. $(0,\sigma_\upsilon^2)$. The X_{it} are assumed independent of the υ_{it} for all i and t. The LSDV (least squares dummy variables) estimator performs ordinary least squares (OLS) on

$$y = \alpha \iota_{NT} + X\beta + Z_\mu \mu + \upsilon = Z\delta + Z_\mu \mu + \upsilon \tag{5}$$

Strictly speaking, the constant ι_{NT} has to be removed from the regression to evade the dummy variable trap, since ι_{NT} is spanned by Z_μ. Note that Z is $NT \times (K + 1)$ and Z_μ, the matrix of individual dummies, is $NT \times N$. If N is large, this will include too many individual dummies, and the matrix to be inverted by OLS is large and of dimension $(N + K)$. Alternatively, one can pre-multiply the model by Q and perform OLS on the resulting transformed model:

$$Qy = QX\beta + Q\upsilon \tag{6}$$

This uses the fact that $QZ_\mu = Q\iota_{NT} = 0$, since $PZ_\mu = Z_\mu$. In other words, the Q matrix wipes out the individual effects. This is a regression of $\tilde{y} = Qy$ with typical element $(y_{it} - \bar{y}_{i\cdot})$ on $\tilde{X} = QX$ with typical element $(X_{it,k} - \bar{X}_{i.,k})$ for the kth regressor, $k = 1,2, \ldots , K$. This involves the inversion of a $(K \times K)$ matrix rather than $(N + K) \times (N + K)$. The resulting OLS estimator is

$$\tilde{\beta}_{FE} = (X'QX)^{-1} X'Qy \tag{7}$$

with $\text{var}(\tilde{\beta}) = \sigma_\upsilon^2 (X'QX)^{-1} = \sigma_\upsilon^2 (\tilde{X}'\tilde{X})^{-1}$.

For the *random-effects* case $\mu_i \sim$ i.i.d. $(0, \sigma_\mu^2)$, $\upsilon_{it} \sim$ i.i.d. $(0, \sigma_\upsilon^2)$ and the μ_i are independent of the υ_{it}. In addition, the X_{it} are independent of the μ_i and v_{it}, for all i and t. The variance–covariance matrix is given by

$$\Omega = E(uu') = \sigma_\mu^2 (I_N \otimes J_T) + \sigma_\upsilon^2 (I_N \otimes I_T) = \sigma_1^2 P + \sigma_v^2 Q \tag{8}$$

where $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\upsilon^2$. This is the spectral decomposition representation of Ω, with σ_1^2 being the first unique characteristic root of Ω of multiplicity N and σ_υ^2 is the second unique characteristic root of Ω of multiplicity $N(T - 1)$. It is easy to verify, using the properties of P and Q, that

$$\Omega^{-1} = \frac{1}{\sigma_1^2} P + \frac{1}{\sigma_v^2} Q \tag{9}$$

and

$$\Omega^{-1/2} = \frac{1}{\sigma_1} P + \frac{1}{\sigma_v} Q \tag{10}$$

In fact, $\Omega^r = (\sigma_1^2)^r P + (\sigma_\upsilon^2)^r Q$, where r is an arbitrary scalar. Now we can obtain GLS as a weighted least squares. Fuller and Battese (1974) suggested pre-multiplying the regression equation by $\sigma_\upsilon \Omega^{-1/2} = Q + (\sigma_\upsilon/\sigma_1)P$ and performing OLS on the resulting transformed regression. In this case, $y^* = \sigma_\upsilon \Omega^{-1/2} y$ has a typical element $y_{it}^* = y_{it} - \theta \bar{y}_{i\cdot}$, where $\theta = 1 - (\sigma_\upsilon/\sigma_1)$. This transformed regression inverts a matrix of dimension $(K + 1)$ and can be easily implemented using any regression package.

DOI: 10.1002/for

The best quadratic unbiased (BQU) estimators of the variance components arise naturally from the spectral decomposition of Ω. In fact, $Pu \sim (0, \sigma_1^2 P)$ and $Qu \sim (0, \sigma_\nu^2 Q)$ and

$$\hat{\sigma}_1^2 = \frac{u'Pu}{tr(P)} = T\sum_{i=1}^{N} \bar{u}_{i.}^2 / N \tag{11}$$

and

$$\hat{\sigma}_\nu^2 = \frac{u'Qu}{tr(Q)} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(u_{it} - \bar{u}_{i.})^2}{N(T-1)} \tag{12}$$

provide the BQU estimators of σ_1^2 and σ_ν^2, respectively.

Suppose we want to predict S periods ahead for the ith individual. For the GLS model, knowing the variance–covariance structure of the disturbances, Goldberger (1962) showed that the best linear unbiased predictor (BLUP) of $y_{i,T+S}$ is

$$\hat{y}_{i,T+S} = Z'_{i,T+S}\hat{\tilde{\delta}}_{GLS} + w'\Omega^{-1}\hat{u}_{GLS} \quad \text{for } s \geqslant 1 \tag{13}$$

where $\tilde{u}_{GLS} = y - Z\tilde{\tilde{\delta}}_{GLS}$ and $w = E(u_{i,T+S}u)$. Note that for period $T + S$

$$u_{i,T+S} = \mu_i + \nu_{i,T+S} \tag{14}$$

and $w = \sigma_\mu^2(l_i \otimes \iota_T)$, where l_i is the ith column of I_N; i.e. l_i is a vector that has 1 in the ith position and zero elsewhere. In this case

$$w'\Omega^{-1} = \sigma_\mu^2(l'_i \otimes l'_T)\left[\frac{1}{\sigma_1^2}P + \frac{1}{\sigma_\nu^2}Q\right] = \frac{\sigma_\mu^2}{\sigma_1^2}(l'_i \otimes l'_T) \tag{15}$$

since $(l'_i \otimes l'_T)P = (l'_i \otimes l'_T)$ and $(l'_i \otimes l'_T)Q = 0$. The typical element of $w'\Omega^{-1}\tilde{u}_{\mathrm{GLS}}$ becomes $((T\sigma_\mu^2/\sigma_1^2)\bar{\tilde{u}}_{i.,\mathrm{GLS}})$, where $\bar{\tilde{u}}_{i.,GLS} = \Sigma_{t=1}^{T}\hat{u}_{it,GLS}/T$. Therefore, the BLUP for $y_{i,T+S}$ corrects the GLS prediction by a fraction of the mean of the GLS residuals corresponding to that ith individual. This predictor was considered by Wansbeek and Kapteyn (1978), Lee and Griffiths (1979) and Taub (1979). The BLUP are optimal assuming true values of the variance components. In practice, these are replaced with estimated values that yield empirical BLUP. Kackar and Harville (1984) propose inflation factors that account for the additional uncertainty introduced by estimating these variance components.

Baillie and Baltagi (1999) consider the practical situation of prediction from the error component regression model when the variance components are not known. They derive both theoretical and simulation evidence as to the relative efficiency of four alternative predictors: (i) an ordinary predictor, based on the optimal predictor but with Maximum Likelihood Estimates (MLEs) replacing population parameters; (ii) a truncated predictor that ignores the error component correction, but uses MLEs for its regression parameters; (iii) a misspecified predictor which uses OLS estimates of the regression parameters; and (iv) a fixed-effects predictor which assumes that the individual effects are fixed parameters that can be estimated. The asymptotic formulae for mean square error (MSE) prediction are derived for all four predictors. Using numerical and simulation results, these are shown to perform adequately in realistic sample sizes ($N = 50$ and 500 and $T = 10$ and 20). Both the analytical and sampling results show that there are substantial gains in MSE prediction by using the ordinary predictor instead of the misspecified or the truncated predictors, especially with increasing $\rho = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ values. The reduction in MSE is about tenfold for $\rho = 0.9$ and a little more than twofold for $\rho = 0.6$ for various values of N and T. The fixed-effects predictor performs remarkably well, being a close second to the ordinary predictor for all experiments. Simulation evidence confirms

DOI: 10.1002/for

the importance of taking into account the individual effects when making predictions. The ordinary predictor and the fixed effects predictor outperform the truncated and misspecified predictors and are recommended in practice.

It is important to note that BLUP is a statistical methodology that has been used extensively in animal breeding (see Henderson, 1975; Harville, 1976, 1985). It is used to estimate genetic merits. For example, in animal breeding, one predicts the production of milk by daughter cows based on their lineage. Robinson (1991) is a good review of BLUP and how it can be used to derive the Kalman filter, the method of Kriging used for ore reserve estimation, credibility theory used to work out insurance premiums, removing noise from images and for small-area estimation. Robinson argues that BLUP is a method of estimating random effects. While BLUP was developed via a frequentist approach to statistics, it has a Bayesian interpretation; see Harville (1976), who showed that Bayesian posterior mean predictors with a diffuse prior are equivalent to BLUP. Robinson (1991, p. 30) adds that one of the reasons why the estimation of random effects has been neglected by the classical school of thought is that: 'The idea of estimating random effects seems suspiciously Bayesian to some Classical statisticians ... adding that ... the adherents of each school emphasize the differences rather than the similarities.' One of the commentators of the paper paraphrases I. J. Good's memorable aphorism: 'To a Bayesian, all things are Bayesian.' He argues that a summary of Robinson's paper could be 'To a non-Bayesian, all things are BLUPs.' For an application in actuarial science to the problem of predicting future claims of a risk class, given past claims of that and related risk classes, see Frees *et al.* (1999, 2001); also, Battese *et al.* (1988) for predicting county crop areas with survey and satellite data using an error component model.

What does the BLUP look like for the ith individual, S periods ahead for the two-way model? For the two-way error components disturbances:

$$u_{it} = u_i + \lambda_t + \nu_{it} \quad i = 1, \dots, N \quad t = 1, \dots, T \tag{16}$$

with $\mu_i \sim$ i.i.d.$(0, \sigma_\mu^2)$, $\lambda_t \sim$ i.i.d.$(0, \sigma_\lambda^2)$ and $\upsilon_{it} \sim$ i.i.d.$(0, \sigma_\upsilon^2)$ independent of each other. In addition, X_{it} is independent of μ_i, λ_t and υ_{it} for all i and t. The variance–covariance matrix is given by

$$\Omega = E(uu') = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\lambda^2(J_N \otimes I_T) + \sigma_\nu^2(I_N \otimes I_T) \tag{17}$$

The disturbances are homoskedastic with $\text{var}(u_{it}) = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\upsilon^2$ for all i and t:

$$\begin{aligned} \text{cov}(u_{it}, u_{js}) &= \sigma_\mu^2 \quad i = j, t \neq s \\ &= \sigma_\lambda^2 \quad i \neq j, t = s \end{aligned}$$

and zero otherwise. For period $T + S$:

$$u_{i,T+S} = \mu_i + \lambda_{T+S} + \nu_{i,T+S} \tag{18}$$

and

$$\begin{aligned} E(u_{i,T+S}u_{jt}) &= \sigma_\mu^2 \quad \text{for } i = j \\ &= 0 \quad\ \ \text{for } i \neq j \end{aligned} \tag{19}$$

DOI: 10.1002/for

and $t = 1, 2, \ldots, T$. Hence, $w = E(u_{i,T+S}u) = \sigma_\mu^2(l_i \otimes \iota_T)$ remains the same for the two-way model as in the one-way model, where l_i is the ith column of I_N. However, Ω^{-1} is different, and the typical element of $w'\Omega^{-1}\hat{u}_{GLS}$ where $\hat{u}_{GLS} = y - Z_{\delta GLS}$ is

$$\frac{T\sigma_\mu^2}{(T\sigma_\mu^2+\sigma_\nu^2)}\left(\bar{\hat{u}}_{i,\text{GLS}}-\bar{\hat{u}}_{..,\text{GLS}}\right)+\frac{T\sigma_\mu^2}{(T\sigma_\mu^2+N\sigma_\lambda^2+\sigma_\nu^2)}\bar{\hat{u}}_{..,\text{GLS}} \tag{20}$$

where $\bar{\hat{u}}_{i.,\text{GLS}}=\Sigma_{t=1}^{T}\hat{u}_{it,\text{GLS}}/T$ and $\bar{\hat{u}}_{..,\text{GLS}}=\Sigma_i\Sigma_t\hat{u}_{it,\text{GLS}}/NT$. In general, $\bar{\hat{u}}_{..,\text{GLS}}$ is not necessarily zero. The GLS normal equations are $Z'\Omega^{-1}\hat{u}_{\text{GLS}} = 0$. However, if Z contains a constant, then $\iota'_{NT}\Omega^{-1}\hat{u}_{\text{GLS}} = 0$, and using the fact that $\iota'_{NT}\Omega^{-1} = \iota'_{NT}/(T\sigma_\mu^2 + N\sigma_\lambda^2 + \sigma_\upsilon^2)$, one gets $\bar{\hat{u}}_{..,\text{GLS}} = 0$. Hence, for the two-way model, if there is a constant in the model, the BLUP for $y_{i,T+S}$ corrects the GLS prediction by a fraction of the mean of the GLS residuals corresponding to that ith individual:

$$\hat{y}_{i,T+S} = Z'_{i,T+S}\hat{\delta}_{\text{GLS}}+\left(\frac{T\sigma_\mu^2}{T\sigma_\mu^2+\sigma_\nu^2}\right)\bar{\hat{u}}_{i..,\text{GLS}} \tag{21}$$

This looks exactly like the BLUP for the one-way model but with a different Ω.

How would one forecast with a two-way fixed-effects model with both country and time effects? After all, future coefficients of time dummies cannot be estimated unless more structure can be placed on the model. One example is the study by Schmalensee *et al.* (1998), which forecast world carbon dioxide emissions through 2050 using national-level panel data over the period 1950–1990. This consisted of 4018 observations. In 1990, this data covered 141 countries, which accounted for 98.6% of the world's population. This paper estimated a reduced form model relating per capita CO_2 emissions from energy consumption to a flexible functional form of real GDP per capita using time and period fixed effects. Schmalensee *et al.* (1998) forecast the time effects using a linear spline model with different growth rates prior to 1970 and after 1970, i.e., $\lambda_t = \gamma_1 + \gamma_2 t + \gamma_3(t - 1970).1[t \geqq 1970]$, with the last term being an indicator function which is 1 when $t \geqq 1970$. Also, using a nonlinear trend model including a logarithmic term, i.e., $\lambda_t = \delta_1 + \delta_2 t + \delta_3 \ln(t - 1940)$. Although these two time effects specifications had essentially the same goodness-of-fit performance, they resulted in different out-of-sample projections. The linear spline projected the time effects by continuing the estimated 1970–1990 trend to 2050, while the nonlinear trend projected a flattening trend consistent with the trend deceleration from 1950 to 1990. An earlier study by Holtz-Eakin and Selden (1995) employed 3754 observations over the period 1951–1986. For their main case, they simply set the time effect at its value in the last year in their sample.

Serial correlation

So far, we have derived Goldberger's (1962) BLUP of $y_{i,T+S}$ for the one-way error component model without serial correlation. For ease of reference, we reproduce the one-period-ahead forecast for the ith individual:

$$\hat{y}_{i,T+1} = Z'_{i,T+1}\hat{\hat{\delta}}_{\text{GLS}} + w'\Omega^{-1}\hat{u}_{\text{GLS}} \tag{22}$$

where $\hat{u}_{\text{GLS}} = y - Z\hat{\hat{\delta}}_{\text{GLS}}$ and $w = E(u_{i,T+1}u)$. For the AR(1) model with no error components, a standard result is that the last term reduces to $\rho\hat{u}_{i,T}$, where $\hat{u}_{i,T}$ is the Tth GLS residual for the ith individual. For the one-way error component model without serial correlation (see Taub, 1979), the last term reduces to $[T\sigma_\mu^2/(T\sigma_\mu^2 + \sigma_\upsilon^2)]\bar{\hat{u}}_{i.}$, where $\bar{\hat{u}}_{i.} = \Sigma_{t=1}^{T}\hat{u}_{it}/T$ is the average of the ith individual's GLS residuals. Baltagi and Li (1992) showed that when *both* error components and serial correlation are present, i.e.

$$\upsilon_{it} = \rho \upsilon_{i,t-1} + \varepsilon_{it} \tag{23}$$

$|\rho| < 1$ and $\varepsilon_{it} \sim$ i.i.d.$(0, \sigma_\varepsilon^2)$. The μ_i are independent of the υ_{it} and $\upsilon_{i0} \sim (0, \sigma_\varepsilon^2/(1-\rho^2))$. The last term reduces to

$$w'\Omega^{-1}\hat{u}_{\text{GLS}} = \rho \hat{u}_{i,T} + \left(\frac{(1-\rho)^2\sigma_\mu^2}{\sigma_\omega^2}\right)\left[\omega \hat{u}_{i1}^* + \sum_{t=2}^{T} \hat{u}_{it}^*\right] \tag{24}$$

where u_{it}^* denotes the Prais–Winsten-transformed residuals

$$\begin{aligned} u_{it}^* &= \sqrt{1-\rho^2}\, u_{i1} \quad \text{for } t = 1 \\ &= u_{it} - \rho u_{i,t-1} \quad \text{for } t = 2, \ldots, T \end{aligned}$$

with $\omega = \sqrt{(1+\rho)/(1-\rho)}$, $\sigma_\omega^2 = d^2\sigma_\mu^2(1-\rho)^2 + \sigma_\varepsilon^2$, and $d^2 = w^2 + (T-1)$. Note that $\hat{u}_{i1}^*$ receives an w weight in averaging across the ith individual's residuals. (i) If $\sigma_\mu^2 = 0$, so that only serial correlation is present, the prediction correction term reduces to $\rho\tilde{u}_{i,T}$. Similarly, (ii) if $\rho = 0$, so that only error components are present, this reduces to $[T\sigma_\mu^2/(T\sigma_\mu^2 + \sigma_\upsilon^2)]\bar{\tilde{u}}_{i.}$.

For the one-way error component model with remainder disturbances following an AR(2) process, i.e.

$$u_{it} = \rho_1 v_{i,t-1} + \rho_2 v_{i,t-2} + \varepsilon_{it} \tag{25}$$

where $\varepsilon_{it} \sim \text{IIN}(0, \sigma_\varepsilon^2)$, $|\rho_2| < 1$ and $|\rho_1| < (1-\rho_2)$. Baltagi and Li (1992) find that the last term reduces to

$$w'\Omega^{-1}\hat{u}_{\text{GLS}} = \rho_1 \hat{u}_{i,T-1} + \rho_2 \hat{u}_{i,T-2} + \left[\frac{(1-\rho_1-\rho_2)^2\sigma_\mu^2}{\sigma_\omega^2}\right]\left[\omega_1 \hat{u}_{i1}^* + \omega_2 \hat{u}_{i2}^* + \sum_{t=3}^{T} \hat{u}_{it}^*\right] \tag{26}$$

where

$$\omega_1 = \sigma_\varepsilon/\sigma_v(1-\rho_1-\rho_2) \qquad \omega_2 = \sqrt{(1+\rho_2)/(1-\rho_2)}$$

$$\sigma_\omega^2 = d^2\sigma_\mu^2(1-\rho_1-\rho_2)^2 + \sigma_\varepsilon^2$$

$$d^2 = \omega_1^2 + \omega_2^2 + (T-2)$$

and

$$\hat{u}_{i1}^* = (\sigma_\varepsilon/\sigma_v)\,\hat{u}_{i1}$$

$$\hat{u}_{i2}^* = \sqrt{1-\rho_2^2}\,[\hat{u}_{i2} - (\rho_1/(1-\rho_2))\,\hat{u}_{i1}]$$

$$\hat{u}_{it}^* = \hat{u}_{it} - \rho_1 \hat{u}_{i,t-1} - \rho_2 \hat{u}_{i,t-2} \quad \text{for } t = 3, \ldots, T$$

Note that if $\rho_2 = 0$, this predictor reduces to that of the AR(1) model with RE. Also, note that for this predictor the first two residuals are weighted differently when averaging across the ith individual's residuals.

For the one-way error component model with remainder disturbances following the specialized AR(4) process for quarterly data, i.e., $\upsilon_{it} = \rho\upsilon_{i,t-4} + \varepsilon_{it}$, where $|\rho| < 1$ and $\varepsilon_{it} \sim \text{IIN}(0, \sigma_\varepsilon^2)$, Baltagi and Li (1992) find that the last term reduces to

$$w'\Omega^{-1}\hat{u}_{\text{GLS}} = \rho\hat{u}_{i,T-3} + \left[\frac{(1-\rho)^2\sigma_\mu^2}{\sigma_\alpha^2}\right]\left[\omega\sum_{t=1}^{4}\hat{u}_{it}^{*} + \sum_{t=5}^{T}\hat{u}_{it}^{*}\right] \tag{27}$$

where $\omega = \sqrt{(1+\rho)/(1-\rho)}$, $\sigma_\omega^2 = d^2(1-\rho)^2\sigma_\mu^2 + \sigma_\varepsilon^2$, $d^2 = 4\omega^2 + (T-4)$ and

$$\begin{aligned} u_{it}^{*} &= \sqrt{1-\rho^2}\,u_{it} \quad \text{for } t = 1, 2, 3, 4 \\ &= u_{it} - \rho u_{i,t-4} \quad \text{for } t = 5, 6, \ldots, T \end{aligned}$$

Note, for this predictor, that the first four quarterly residuals weighted by w when averaging across the ith individual's residuals.

Finally, for the one-way error component model with remainder disturbances following an MA(1) process, i.e.

$$\upsilon_{it} = \varepsilon_{it} + \lambda\varepsilon_{i,t-1}$$

where $\varepsilon_{it} \sim \text{IIN}(0, \sigma_\varepsilon^2)$ and $|\lambda| < 1$, Baltagi and Li (1992) find that

$$w'\Omega^{-1}\hat{u}_{\text{GLS}} = -\lambda\left(\frac{a_{T-1}}{a_T}\right)^{1/2}\hat{u}_{iT}^{*} + \left[1 + \lambda\left(\frac{a_{T-1}}{a_T}\right)^{1/2}\alpha_T\right]\left(\frac{\sigma_\mu^2}{\sigma_\omega^2}\right)\left[\sum_{t=1}^{T}\alpha_t\hat{u}_{it}^{*}\right] \tag{28}$$

where $a_t = 1 + \lambda^2 + \ldots + \lambda^{2t}$ with $a_0 = 1$, $\sigma_\omega^2 = d^2\sigma_\mu^2 + \sigma_\varepsilon^2$ and $d^2 = \Sigma_{t=1}^{T}\alpha_t^2$, and the $\tilde{u}_{it}^{*}$, can be solved for recursively as follows:

$$\tilde{u}_{i1}^{*} = (a_0/a_1)^{1/2}\,\tilde{u}_{i1}$$

$$\tilde{u}_{it}^{*} = \lambda(a_{t-2}/a_{t-1})^{1/2}\hat{u}_{i,t-1}^{*} + (a_{t-1}/a_t)^{1/2}\hat{u}_{i,t}\ t = 2, \ldots, T$$

If $\lambda = 0$, then $a_t = \alpha_t = 1$ for all t, the prediction correction term reduces to the predictor for the error component model with no serial correlation. If $\sigma_\mu^2 = 0$, the predictor reduces to that of the MA(1) process.

These results can be extended to the MA(q) case (see Baltagi and Li, 1994) and the autoregressive moving average ARMA(p, q) case on the υ_{it} (see MaCurdy, 1982, and more recently Galbraith and Zinde-Walsh, 1995). For an extension to the two-way model with serially correlated disturbances, see Revankar (1979), who considers the case where the λ_t also follow an AR(1) process; also, Karlsson and Skoglund (2004), who consider the two-way error component model with an ARMA process on the time-specific effects. For an extension to the unequally spaced panel data regression model with AR(1) remainder disturbances, see Baltagi and Wu (1999).

Frees and Miller (2004) forecast the sale of state lottery tickets using panel data from 50 postal (ZIP) codes in Wisconsin observed over 40 weeks. The first 35 weeks of data are used to estimate the model and the remaining 5 weeks are used to assess the validity of model forecasts. Using the mean absolute error criteria and the mean absolute percentage error criteria, the best forecasts were given by the error component model with AR(1) disturbances followed by the fixed-effects model with AR(1) disturbances.

Spatial correlation

Consider the spatial panel data model

$$y_{it} = x_{it}'\beta + \varepsilon_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T \tag{29}$$

DOI: 10.1002/for

(see Anselin, 1988, p. 152), where the disturbance vector for time t is given by

$$\varepsilon_t = \mu + \phi_t \tag{30}$$

with $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$, $\mu = (\mu_1, \ldots, \mu_N)'$ denotes the vector of individual effects and $\phi_t = (\phi_{1t}, \ldots, \phi_{Nt})'$ are the remainder disturbances which are independent of μ. The ϕ_ts follow the spatial error dependence model

$$\phi_t = \lambda W \phi_t + \nu_t \tag{31}$$

where W is the matrix of known spatial weights of dimension $N \times N$ with zero diagonal elements and row normalized elements that sum to 1. λ is the spatial autoregressive coefficient, $\upsilon_t = (\upsilon_{1t}, \ldots, \upsilon_{Nt})'$ is i.i.d.$(0, \sigma_\upsilon^2)$ and is independent of ϕ_t and μ.

For the random effects model, the μ_is are i.i.d.$(0, \sigma_\mu^2)$ and are independent of the ϕ_{it}s (see Anselin, 1988). For this model, we need to derive the variance–covariance matrix. Let $B = I_N - \lambda_W$, then the disturbances in equation (31) can be written as follows: $\phi_t = (I_N - \lambda_W)^{-1}\upsilon_t = B^{-1}\upsilon_t$. Substituting for ϕ_t, we get

$$\varepsilon = (\iota_T \otimes I_N)\mu + \left(I_T \otimes B^{-1}\right)\nu \tag{32}$$

where ι_T is a vector of ones of dimension T and I_N is an identity matrix of dimension N. The variance–covariance matrix is

$$\Omega = E(\varepsilon\varepsilon') = \sigma_\mu^2(\iota_T\iota_T' \otimes I_N) + \sigma_\nu^2\left(I_T \otimes (B'B)^{-1}\right) \tag{33}$$

Let $\Psi = \dfrac{1}{\sigma_\nu^2}\Omega = \dfrac{\sigma_\mu^2}{\sigma_\nu^2}(\iota_T\iota_T' \otimes I_N) + \left(I_T \otimes (B'B)^{-1}\right)$ and $\theta = \dfrac{\sigma_\mu^2}{\sigma_\nu^2}$, then

$$\Psi = \bar{J}_T \otimes (T\theta I_N) + I_T \otimes (B'B)^{-1} = \bar{J}_T \otimes V + E_T \otimes (B'B)^{-1} \tag{34}$$

where $V = T\theta I_N + (B'B)^{-1}$ and $E_T = I_T - \bar{J}_T$. It is easy to verify that

$$\Psi^{-1} = \bar{J}_T \otimes V^{-1} + E_T \otimes (B'B) \tag{35}$$

(see Anselin, 1988, p. 154). In this case, GLS using Ω^{-1} yields $\hat{\beta}_{\mathrm{GLS}}$. Note that the computation is simplified, since the $NT \times NT$ matrix Ψ^{-1} is based on inverting two lower-order matrices, V and B, both of dimensions $N \times N$.

If $\lambda = 0$, so that there is no spatial autocorrelation, then $B = I_N$ and Ω becomes the usual error component variance–covariance matrix:

$$\Omega_{\mathrm{RE}} = E(\varepsilon\varepsilon') = \sigma_\mu^2(\iota_T\iota_T' \otimes I_N) + \sigma_\nu^2(I_T \otimes I_N) \tag{36}$$

Applying GLS using this Ω_{RE} yields the random effects (RE) estimator, which we will denote by $\hat{\beta}_{\mathrm{RE}}$.

Baltagi and Li (2004) derived the BLUP correction term when both error components and spatial autocorrelation are present. In this case $w = E(\varepsilon_{i,T} + S^{\varepsilon}) = E[(\mu_i + \phi_{i,T+S})\varepsilon] = \sigma_\mu^2(\iota_T \otimes l_i)$ since the ϕs are not correlated over time. Using $\Omega^{-1} = \dfrac{1}{\sigma_\nu^2}\Psi^{-1}$, we get

DOI: 10.1002/for

$$w'\Omega^{-1} = \frac{\sigma_\mu^2}{\sigma_v^2}(\iota_T' \otimes l_i')\left[\left(\bar{J}_T \otimes V^{-1}\right) + \left(E_T \otimes (B'B)\right)\right] = \theta\left(\iota_T' \otimes l_i'V^{-1}\right) \tag{37}$$

since $\iota_T' E_T = 0$. Therefore

$$w'\Omega^{-1}\hat{\varepsilon}_{\mathrm{GLS}} = \theta\left(\iota_T' \otimes l_i'V^{-1}\right)\hat{\varepsilon}_{\mathrm{GLS}} = \theta l_i'V^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{t,\mathrm{GLS}} = T\theta\sum_{j=1}^{N}\delta_j\bar{\varepsilon}_{j.,\mathrm{GLS}} \tag{38}$$

where δ_j is the jth element of the ith row of V^{-1} and $\bar{\varepsilon}_{j.,\mathrm{GLS}} = \Sigma_{t=1}^{T}\bar{\varepsilon}_{tj,\mathrm{GLS}}/T$. In other words, the BLUP adds to $x_{i,T+S}'\hat{\beta}_{\mathrm{GLS}}$ a weighted average of the GLS residuals for the N regions averaged over time. The weights depend upon the spatial matrix W and the spatial autocorrelation coefficient λ. To make this predictor operational, we replace $\hat{\beta}_{\mathrm{GLS}}$, θ and λ by their estimates from the RE-spatial MLE.

When there is no spatial autocorrelation, i.e., $\lambda = 0$, the BLUP correction term reduces to the Taub (1979) predictor term of the RE model. Also, when there are no random effects, so that $\sigma_\mu^2 = 0$, then $\theta = 0$ and the BLUP prediction term drops out completely. In this case, Ω reduces to $\sigma_\upsilon^2(I_T \otimes (B'B)^{-1})$ and GLS on this model, based on the MLE of λ, yields the pooled spatial estimator. The corresponding predictor is labelled the pooled spatial predictor.

If the fixed-effects model with spatial autocorrelation is the true model, then the problem is to predict

$$y_{i,T+S} = x_{i,T+S}'\beta + \mu_i + \phi_{i,T+S} \tag{39}$$

with $\phi_{T+S} = \lambda W\phi_{T+S} + \upsilon_{T+s}$. Unlike the usual FE case, $\lambda \neq 0$ and the μ_is and β have to be estimated from MLE, i.e., using the FE-spatial estimates. The disturbance vector can be written as $\phi = (I_T \otimes B^{-1})\upsilon$, so that $w = E(\phi_{i,\mathrm{T}+S}\phi) = 0$ since the υs are not serially correlated over time. So the BLUP for this model looks like that for the FE model without spatial correlation except that the μ_i's and β are estimated assuming $\lambda \neq 0$. The corresponding predictor is labelled the FE-spatial predictor.

Baltagi and Li (2004) consider the problem of prediction in a panel data regression model with spatial autocorrelation in the context of a simple demand equation for cigarettes. This is based on a panel of 46 states over the period 1963–1992. The spatial autocorrelation due to neighboring states and the individual heterogeneity across states is taken explicitly into account. They compare the performance of several predictors of the states' demand for cigarettes for 1 year and 5 years ahead. The estimators whose predictions are compared include OLS, fixed effects ignoring spatial correlation, fixed effects with spatial correlation, random-effects GLS estimator ignoring spatial correlation and random-effects estimator accounting for the spatial correlation. Based on RMSE forecast performance, estimators that take into account spatial correlation and heterogeneity across the states perform the best. The FE-spatial estimator gives the lowest RMSE for the first 4 years and is only surpassed by the RE-spatial in the fifth year. Overall, both the RE-spatial and FE-spatial estimators perform well in predicting cigarette demand.

For examples of prediction of random effects in a spatial generalized linear mixed model, see Zhang (2002), who applied this technique to disease mapping of plant roots on a 90-acre farm in Washington state. In many applications in epidemiology, ecology and agriculture, predicting the random effects of disease at unsampled sites requires modeling the spatial dependence continuously. This is especially important for data observed at point locations, where interpolation is needed to predict values at unsampled sites. Zhang implements this minimum mean squared error prediction through the Metropolis–Hastings algorithm.

HETEROGENEOUS PANELS

The underlying assumption behind pooling the observations across individuals and time is the homogeneity of the regression coefficients. The latter is a testable assumption using a Chow F-test, which can allow for an error component variance–covariance matrix in a random-effects model or varying intercepts in a fixed-effects model (see Baltagi, 2005). The pooled model represents a behavioral equation with the same parameters across individuals and over time, given by equation (3). The unrestricted model, however, is a heterogeneous model with different parameters across individuals or time. In particular, for macro panel data with large T, one can allow for a different set of regression coefficients for each country:

$$y_i = Z_i \delta_i + u_i, \quad i = 1, \ldots, N \tag{40}$$

where y_i is $(T \times 1)$, $Z_i = [\iota_T, X_i]$, X_i is $(T \times K)$, $\delta_i' = (\alpha_i, \beta_i')$, and u_i is $(T \times 1)$. The null hypothesis is

$$H_0 : \delta_i = \delta, \; \forall i = 1, \ldots, N$$

So, under H_0, we can write the restricted model as $y = Z\delta + u$. The unrestricted model can also be written as

$$y = Z^*\delta^* + u = \begin{pmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & Z_N \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_N \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix} \tag{41}$$

where $Z = Z^*I^*$ with $I^* = (\iota_N \otimes I_K)$. This is feasible when T is large and most likely may reject the poolability of the data. In fact, Robertson and Symons (1992) and Pesaran and Smith (1995) questioned the poolability of the data across heterogeneous units. Instead, they argue in favor of heterogeneous estimates that can be combined to obtain homogeneous estimates if the need arises. To make this point, Robertson and Symons (1992) studied the properties of some panel data estimators when the true model is static and *heterogeneou*s but the estimated model is taken to be dynamic and *homogeneous*. This is done for both stationary and nonstationary regressors. The basic conclusion is that severe biases can occur in dynamic estimation even for relatively small parameter variation.

Pesaran and Smith (1995) generalize this to the case of a heterogeneous dynamic panel data model given by

$$y_{it} = \lambda_i y_{i,t-1} + \beta_i x_{it} + u_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T \tag{42}$$

where λ_i is i.i.d.$(\lambda, \sigma_\lambda^2)$ and β_i is i.i.d.(β, σ_β^2). Further λ_i and β_i are independent of y_{is}, x_{is} and u_{is} for all s. The objective in this case is to obtain consistent estimates of the mean values of λ_i and β_i. Pesaran and Smith (1995) present four different estimation procedures:

1. aggregate time series regressions of group averages;
2. cross-section regressions of averages over time;

DOI: 10.1002/for

3. pooled regressions allowing for fixed or random intercepts; or
4. separate regressions for each group, where coefficient estimates are averaged over these groups.

They show that when T is small (even if N is large), all the procedures yield inconsistent estimators. When both N and T are large, Pesaran and Smith (1995) show that the cross-section regression procedure will yield consistent estimates of the mean values of λ and β.

Heterogeneous estimators

One heterogeneous estimator is the random coefficient model studied extensively by Swamy (1970). The model is given by

$$y_i = Z_i\delta_i + u_i, \quad i = 1, \dots, N \tag{43}$$

with

$$y_i \sim N\left(Z_i\delta_i, \sigma_i^2 I_T\right) \tag{44}$$

In this case

$$\delta_i = \bar{\delta} + \varepsilon_i, \quad \delta_i \sim N\left(\bar{\delta}, \Delta\right) \tag{45}$$

with $\text{cov}(\delta_i, \delta_j) = 0$, $i \neq j$). Substituting (45) into (43) yields

$$y_i = Z_i\bar{\delta} + \upsilon_i \tag{46}$$

where $\upsilon_i = Z_i\varepsilon_i + u_i$.

Stacking all NT observations, we have

$$y = Z\bar{\delta} + \upsilon \tag{47}$$

where $\upsilon = Z^*\varepsilon + u$. The covariance matrix for the composite disturbance term υ is bloc-diagonal, $\text{diag}(\Sigma_i)$ where

$$\Sigma_i = Z_i\Delta Z_i' + \sigma_i^2 I_T$$

The best linear unbiased estimator of $\bar{\delta}$ for (47) is the GLS estimator

$$\hat{\bar{\delta}}_{\text{GLS}} = \sum_{i=1}^{N}\Theta_i\hat{\delta}_{i,\text{OLS}} \tag{48}$$

where

$$\hat{\delta}_{i,\text{OLS}} = (Z_i'Z_i)^{-1}Z_i'y_i \tag{49}$$

and

$$\Theta_i = \left(\sum_{i=1}^{N}\left[\Delta + \sigma_i^2(Z_i'Z_i)^{-1}\right]^{-1}\right)^{-1}\left[\Delta + \sigma_i^2(Z_i'Z_i)^{-1}\right]^{-1} \tag{50}$$

DOI: 10.1002/for

The covariance matrix for the GLS estimator is

$$V\left[\hat{\bar{\delta}}_{\mathrm{GLS}}\right]=\left(\sum_{i=1}^{N}\left[\Delta+\sigma_i^2(Z_i'Z_i)^{-1}\right]^{-1}\right)^{-1} \tag{51}$$

Swamy suggested using $\hat{\delta}_{\delta i,\mathrm{OLS}}$ and their residuals $\hat{e}_i$ $(= y_i - Z_i\hat{\delta}_{i,\mathrm{OLS}})$ to obtain unbiased estimators of σ_i^2 and Δ. Pesaran and Smith (1995) and Pesaran *et al.* (1999) advocate alternative estimators which they call respectively the Mean Group estimator and the Pooled Mean Group estimator.

The Mean Group estimator is obtained by estimating the coefficients of each cross-section separately by OLS and then taking an arithmetic average:

$$\tilde{\bar{\delta}}=\frac{1}{N}\sum_{i=1}^{N}\hat{\delta}_{i,OLS}. \tag{52}$$

When $T \to \infty$, $\tilde{\delta}_{i,\mathrm{OLS}} \to \delta_i$ and (52) will be consistent when N also goes to infinity.

Pesaran *et al.*'s (1999) Pooled Mean Group estimator constrains the long-run coefficients to be the same among individuals. Details of this estimator are given in that paper. Since this is a forecasting survey, we only refer to the various heterogeneous estimators used and focus on their implementation in forecasting applications; for an extensive discussion of these estimators, see Baltagi *et al.* (2006), Pesaran *et al.* (1996).

Maddala *et al.* (1997) applied classical, empirical Bayes and Bayesian procedures to the problem of estimating short-run and long-run elasticities of residential demand for electricity and natural gas in the USA for 49 states over 21 years (1970–1990). Since the elasticity estimates for each state were the ultimate goal of their study they were faced with three alternatives. The first was to use individual time series regressions for each state. These gave bad results, were hard to interpret, and had several wrong signs. The second option was to pool the data and use panel data estimators. Although the pooled estimates gave the right signs and were more reasonable, Maddala *et al.* (1997) argued that these estimates were not valid because the hypothesis of homogeneity of the coefficients was rejected. The third option, which they recommended, was to allow for some (but not complete) heterogeneity or (homogeneity). This approach led them to their preferred shrinkage estimator, which gave them more reasonable parameter estimates.

Pre-testing and Stein rule methods

Choosing a pooled estimator if we do not reject H_0: $\delta_i = \delta$ for all i, and the heterogeneous estimator if we reject H_0, leads to a pre-test estimator. This brings into question the appropriate level of significance to use with this preliminary test. In fact, the practice is to use significance levels much higher than 5% (see Maddala and Hu, 1996).

Another problem with the pre-testing procedure is that its sampling distribution is complicated (see Judge and Bock, 1978). Also, these pre-test estimators are dominated by Stein rule estimators under quadratic loss function. Using a wilderness recreation demand model, Ziemer and Wetzstein (1983) show that a Stein rule estimator gives better forecast risk performance than the pooled ($\hat{\delta}_{\mathrm{OLS}}$) or individual estimators ($\tilde{\delta}_{i,\mathrm{OLS}}$). The Stein rule estimator is given by

$$\hat{\delta}_i^S=\left(\frac{c}{F_{\mathrm{obs}}}\right)\hat{\delta}_{\mathrm{OLS}}+\left(1-\frac{c}{F_{\mathrm{obs}}}\right)\hat{\delta}_{i,\mathrm{OLS}} \tag{53}$$

The optimal value of the constant c suggested by Judge and Bock (1978) is

$$c=\frac{(N-1)K-2}{N(T-K)+2}$$

DOI: 10.1002/for

Note that $\hat{\delta}_i^S$ shrinks $\hat{\delta}_{i,OLS}$ towards the pooled estimator $\hat{\delta}_{OLS}$. When N is large, the factor c is roughly $K/(T - K)$. The Bayesian and empirical Bayesian methods imply shrinking towards a weighted mean of the $\hat{\delta}_i$ and not the pooled estimator $\hat{\delta}$.

Forecasting applications

In the context of dynamic demand for gasoline across 18 OECD countries over the period 1960–1990, Baltagi and Griffin (1997) argued for pooling the data as the best approach for obtaining reliable price and income elasticities. They also pointed out that pure cross-section studies cannot control for unobservable country effects, whereas pure time series studies cannot control for unobservable oil shocks or behavioral changes occurring over time. Baltagi and Griffin (1997) compared the homogeneous and heterogeneous estimates in the context of gasoline demand based on the plausibility of the price and income elasticities as well as the speed of adjustment path to the long-run equilibrium. They found considerable variability in the parameter estimates among the heterogeneous estimators some giving implausible estimates, while the homogeneous estimators gave similar plausible short-run estimates that differed only in estimating the long-run effects. Baltagi and Griffin (1997) also compared the forecast performance of these homogeneous and heterogeneous estimators over 1, 5 and 10 years horizon. Their findings show that the homogeneous estimators outperformed their heterogeneous counterparts based on mean squared forecast error. This result was replicated using a panel data set of 21 French regions over the period 1973–1998 by Baltagi *et al.* (2003). Unlike the international OECD gasoline data set, the focus on inter-regional differences in gasoline prices and income within France posed a different type of dataset for the heterogeneity versus homogeneity debate. The variation in these prices and income was much smaller than international price and income differentials. This in turn reduces the efficiency gains from pooling and favors the heterogeneous estimators, especially given the differences between the Paris region and the rural areas of France. Baltagi *et al.* (2003) showed that the time series estimates for each region are highly variable, unstable and offer the worst out-of-sample forecasts. Despite the fact that the shrinkage estimators proposed by Maddala *et al.* (1997) outperformed these individual heterogeneous estimates, they still had a wide range and were outperformed by the homogeneous estimators in out of sample forecasts. Baltagi *et al.* (2000) carried out this comparison for a dynamic demand for cigarettes across 46 US states over 30 years (1963–1992). Once again the results showed that the homogeneous panel data estimators beat the heterogeneous and shrinkage type estimators in RMSE performance for out-of-sample forecasts. In another application, Driver *et al.* (2004) utilize the Confederation of British Industry's (CBI) survey data to measure the impact of uncertainty on UK investment authorizations. The panel consists of 48 industries observed over 85 quarters 1978(Q1) to 1999(Q1). The uncertainty measure is based on the dispersion of beliefs across survey respondents about the general business situation in their industry. The heterogeneous estimators considered are OLS and 2SLS at the industry level, as well as the unrestricted SUR estimation method. Fixed effects, random effects, pooled 2SLS and restricted SUR are the homogeneous estimators considered. The panel estimates find that uncertainty has a negative, non-negligible effect on investment, while the heterogeneous estimates vary considerably across industries. Forecast performance for 12 out-of-sample quarters 1996(Q2) to 1999(Q1) are compared. The pooled homogeneous estimators outperform their heterogeneous counterparts in terms of RMSE.

Baltagi *et al.* (2002) reconsidered the two US panel datasets on residential electricity and natural-gas demand used by Maddala *et al.* (1997) and compared the out-of-sample forecast performance of the homogeneous, heterogeneous and shrinkage estimators. Once again the results show that when the data is used to estimate heterogeneous models across states, individual estimates offer the

DOI: 10.1002/for

worst out-of-sample forecasts. Despite the fact that shrinkage estimators outperform these individual estimates, they are outperformed by simple homogeneous panel data estimates in out-of-sample forecasts. Admittedly, these are additional case studies, but they do add to the evidence that simplicity and parsimony in model estimation offered by the homogeneous estimators yield better forecasts than the more parameter consuming heterogeneous estimators.

Hsiao and Tahmiscioglu (1997) use a panel of 561 US firms over the period 1971–1992 to study the influence of financial constraints on company investment. They find substantial differences across firms in terms of their investment behavior. When a homogeneous pooled model is assumed, the impact of liquidity on firm investment is seriously underestimated. The authors recommend a mixed fixed and random coefficients framework based on the recursive predictive density criteria.

Baltagi *et al.* (2004) reconsider the Tobin q investment model studied by Hsiao *et al.* (1999) using a slightly different panel of 337 US firms over the period 1982–1998. They contrast the out-of-sample forecast performance of nine homogeneous panel data estimators and 11 heterogeneous and shrinkage Bayes estimators over a 5-year horizon. Results show that the average heterogeneous estimators perform the worst in terms of MSE, while the hierarchical Bayes estimator suggested by Hsiao *et al.* (1999) performs the best. Homogeneous panel estimators and iterative Bayes estimators are a close second.

Using data on migration to Germany from 18 source countries over the period 1967–2001, Brucker and Siliverstovs (2006) compare the performance of homogeneous and heterogeneous estimators using out-of-sample forecasts. They find that the mean group estimator performs the worst, while a fixed-effects estimator performs the best in RMSE for 5 years and 10 years ahead forecasts. In general, the heterogeneous estimators performed poorly. They attribute this to the unstable regression parameters across the 18 countries, such that the gains from pooling more than offset the biases from the inter-country heterogeneity.

Rapach and Wohar (2004) show that the monetary model of exchange rate determination performs poorly on a country-by-country basis for US dollar exchange rates over the post-Bretton Woods period for 18 industrialized countries for quarterly data over the period 1973:1–1997:1. However, they find considerable support for the monetary model using panel procedures. They reject tests for the homogeneity assumptions inherent in panel procedures. Hence, they are torn between obtaining panel-cointegrating coefficient estimates that are much more plausible in economic terms than country-by-country estimates. Yet these estimates might be spurious since they are rejected by formal statistical test for pooling. Rapach and Wohar (2004) perform an out-of-sample forecasting exercise using the panel and country-by-country estimates employing the RMSE criteria for a 1-, 4-, 8-, 12- and 16-step-ahead quarters. For the 1-step- and 4-step-ahead, the RMSEs of the homogeneous and heterogeneous estimates are similar. At the 8-step-ahead horizon, homogeneous estimates generate better forecasts in comparison to five of the six heterogeneous estimates. At the 16-step horizon, the homogeneous estimates have RMSE that is smaller than each of the heterogeneous estimates. In most cases the RMSE is reduced by 20%. The authors concluded that while there are good reasons to favor the panel estimates over the country-by-country estimates of the monetary model, there are also good reasons to be suspicious of these panel estimates since the homogeneity assumption is rejected. Despite this fact, they argue that panel data estimates should not be dismissed based on tests for homogeneity alone, because they may eliminate certain biases that plague country-by-country estimates. In fact, panel estimates of the monetary model were more reliable and generated superior forecasts to those of country-by-country estimates. Rapach and Wohar (2004) suspicion of panel data estimates come from Monte Carlo evidence that show that ‘it is not improbable to find evidence in support of the monetary model by relying on panel estimates, even when the true

DOI: 10.1002/for

data generating process is characterized by a heterogeneous structure that is not consistent with the monetary model'. Other papers in this vein are Mark and Sul (2001) and Groen (2005). The latter paper utilizes a panel of vector error correction models based on a common long-run relationship to test whether the euro exchange rates of Canada, Japan and the USA have a long-run link with monetary fundamentals. Out-of-sample forecasts show that this common long-run exchange model is superior to both the naive random walk-based forecasts and the standard cointegrated VAR model-based forecasts, especially for horizons of 2–4 years.

Hoogstrate *et al.* (2000) investigate the improvement of forecasting performance using pooling techniques instead of single-country forecasts for N fixed and T large. They use a set of dynamic regression equations with contemporaneously correlated disturbances. When the parameters of the models are different but exhibit some similarity, pooling may lead to a reduction in the mean squared error of the estimates and the forecasts. They show that the superiority of the pooled forecasts in small samples can deteriorate as T grows. They apply these results to growth rates of 18 OECD countries over the period 1950–1991 using an AR(3) model and an AR(3) model with leading indicators put forth by Garcia-Ferrer *et al.* (1987) and Zellner and Hong (1989). They find that the median MSFE of OLS-based pooled forecasts is smaller than that of OLS-based individual forecasts and that a fairly large T is needed for the latter to outperform the former. They argue that this is due to the reduction in MSE due to imposing a false restriction (pooling). However, for a large enough T, the bias of the pooled estimates increases without bound and the resulting forecasts based on unrestricted estimates will outperform the forecasts based on the pooled restricted estimates.

Gavin and Theodorou (2005) use forecasting criteria to examine the macrodynamic behavior of 15 OECD countries observed quarterly over the period 1980–1996. They utilize a small set of familiar, widely used core economic variables (output, price level, interest rates and exchange rates), omitting country-specific shocks. They find that this small set of variables and a simple VAR common model strongly support the hypothesis that many industrialized nations have similar macroeconomic dynamics. In sample, they often reject the hypothesis that coefficient vectors estimated separately for each country are the same. They argue that these rejections may be of little importance if due to idiosyncratic events since macro time series are typically too short for standard methods to eliminate the effects of idiosyncratic factors. Panel data can be used to exploit the heterogeneous information in cross-country data, hence increasing the data and eliminating the idiosyncratic effects. They compare the forecast accuracy of the individual country models with the common models in a simulated out-of-sample experiment. They calculate four forecasts with increasing horizons at each point in time one quarter ahead and four quarters ahead. For the four equations, at every horizon, the panel forecasts are significantly more accurate more often than are the individual country model forecasts. The biggest differences are for the exchange rate and the interest rate. They conclude that the superior out-of-sample forecasting performance of the common model supports their hypothesis that market economies tend to have a common macrodynamic pattern related to a small number of variables.

Lahiri and Liu (2006) model inflation uncertainty using a dynamic heterogeneous panel data model. They examine the adequacy of EGARCH in explaining forecast uncertainty at the micro level and possible pitfalls from aggregate estimation. Using a panel of density forecasts from the survey of professional forecasters, they show that there is a strong relationship between forecast uncertainty and the level of inflation. They compare a hierarchical Bayes estimator, with empirical Bayes, pooled mean group, pooled OLS, fixed effects, conditional MLE, and an aggregate estimator. Their preferred estimator is the hierarchical Bayes estimator. The conventional time series estimator showed severe aggregation bias. They find that the persistence in forecast uncertainty is much less

DOI: 10.1002/for

than aggregate time series data would suggest. This study emphasizes the importance of individual heterogeneity when ARCH-type models are estimated using aggregate time series data.

For other uses of forecasting with panel data, see Fok *et al.* (2005), who show that forecasts of aggregates like total output or unemployment can be improved by considering panel models of disaggregated series covering 48 states. They use a panel version of a two-regime smooth transition autoregressive (STAR) type model to capture the nonlinear features that are often displayed by macroeconomic variables allowing the parameters that govern the regime switching to differ across states. Also, Mouchart and Rombouts (2005) use a clustering approach to the usual panel data model specification to nowcast from poor data, namely, very short time series and many missing values. Marcelino *et al.* (2003) consider a similar problem of forecasting from panel data with severe deficiencies. Using an array of forecasting models applied to 11 countries originally in the EMU, over the period 1982–1997, at both the monthly and quarterly level, they show that forecasts constructed by aggregating the country-specific models are more accurate than forecasts constructed using the aggregate data.

CAVEATS, RELATED STUDIES AND FUTURE WORK

This survey showed that although the performance of various panel data estimators and their corresponding forecasts may vary in ranking from one empirical example to another (see Baltagi and Griffin, 1997; Baltagi *et al.*, 2000, 2002, 2003, 2004; Driver *et al.*, 2004; Rapach and Wohar, 2004; Brucker and Siliverstovs, 2006), the consistent finding in all these studies is that homogeneous panel data estimators perform well in forecast performance mostly due to their simplicity, their parsimonious representation, and the stability of the parameter estimates. Average heterogeneous estimators perform badly due to parameter estimate instability caused by the estimation of several parameters with short time series. Shrinkage estimators did well for some applications, especially iterative Bayes and iterative empirical Bayes.

Much work remains to be done in forecasting with panels. This brief survey did not cover forecasting with panel VAR methods which are popular in macroeconomics (see Ballabriga *et al.*, 1998; Canova and Ciccarelli, 2004; Pesaran *et al.*, 2004; among others). Canova and Ciccarelli (2004) provide methods for forecasting variables and predicting turning points in panel Bayesian VARs. They allow for interdependencies in the cross-section as well as time variations in the parameters. Posterior distributions are obtained for hierarchical and for Minnesota-type priors and multi-step, multi-unit point and average forecasts for the growth rate of output in the G7 are provided. There is also the problem of forecasting with nonstationary panels; see Breitung and Pesaran (2008) for a recent survey of nonstationary panels; also Binder *et al.* (2005) for estimation and inference in short panel vector autoregressions with unit roots and cointegration and Hjalmarsson (2006) for predictive regressions with endogenous and nearly persistent regressors using panel data. For forecasting with micropanels, see Chamberlain and Hirano (1999), who suggested optimal ways of combining an individual's personal earnings history with panel data on the earnings trajectories of other individuals to provide a conditional distribution for this individual's earnings. Other applications to household survey data eliciting respondents' intentions or predictions for future outcomes, using panel data, include Keane and Runkle (1990) and Das *et al.* (1999), among others. This survey does not get into the large literature on 'forecast combination methods' (see Diebold and Lopez, 1996; Newbold and Harvey, 2002; Stock and Watson, 2004; among others). The latter study used forecast combination methods to forecast output growth in a seven-country quarterly economic data set covering 1959–

DOI: 10.1002/for

1999 using up to 73 predictors per country. This survey also does not get into the related literature on 'forecasting economic aggregates from disaggregates' (see Hendry and Hubrich, 2006). The latter study shows that including disaggregate variables in the aggregate model yields forecasts that outperform forecasting disaggregate variables and then aggregating those forecasts. Another related paper is Giacomini and Granger (2004), who compare the relative efficiency of different methods of forecasting the aggregate of spatially correlated variables. They show that ignoring spatial correlation even when it is weak leads to highly inaccurate forecasts. They also show that when a pooling condition is satisfied there is benefit in forecasting the aggregate directly.

Hopefully, this survey will encourage more work in this area and in particular on the evaluation of panel models using post-sample forecasting à la Diebold and Mariano (1995) and Granger and Huang (1997).

REFERENCES

Anselin L. 1988. *Spatial Econometrics: Methods and Models*. Kluwer: Dordrecht.

Baillie RT, Baltagi BH. 1999. Prediction from the regression model with one-way error components. In *Analysis of Panels and Limited Dependent Variable Models*, Hsiao C, Lahiri K, Lee LF, Pesaran H (eds). Cambridge University Press: Cambridge; 255–267.

Ballabriga FC, Sebastian M, Valles J. 1998. European asymmetries. *Journal of International Economics* **4**: 233–253.

Baltagi BH. 2005. *Econometric Analysis of Panel Data*. Wiley: Chichester.

Baltagi BH, Griffin JM. 1997. Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics* **77**: 303–327.

Baltagi BH, Li D. 2004. Prediction in the panel data model with spatial correlation. In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Anselin L, Florax RJGM, Rey SJ (eds). Springer: Berlin; 283–295.

Baltagi BH, Li Q. 1992. Prediction in the one-way error component model with serial correlation. *Journal of Forecasting* **11**: 561–567.

Baltagi BH, Li Q. 1994. Estimating error component models with general MA (q) disturbances. *Econometric Theory* **10**: 396–408.

Baltagi BH, Wu PX. 1999. Unequally spaced panel data regressions with AR (1) disturbances. *Econometric Theory* **15**: 814–823.

Baltagi BH, Griffin JM, Xiong W. 2000. To pool or not to pool: homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* **82**: 117–126.

Baltagi BH, Bresson G, Pirotte A. 2002. Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators: some empirical evidence from US electricity and natural-gas consumption. *Economics Letters* **76**: 375–382.

Baltagi BH, Bresson G, Griffin JM, Pirotte A. 2003. Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from French regional gasoline consumption. *Empirical Economics* **28**: 795–811.

Baltagi BH, Bresson G. Pirotte A. 2004. Tobin q: forecast performance for hierarchical Bayes, shrinkage, heterogeneous and homogeneous panel data estimators. *Empirical Economics* **29**: 107–113.

Baltagi BH, Bresson G, Pirotte A. 2005. To pool or not to pool? In *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Matyas L, Sevestre P (eds). Kluwer: Dordrecht (forthcoming).

Battese GE, Harter RM, Fuller WA. 1988. An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association* **83**: 28–36.

Binder M, Hsiao C, Pesaran MH. 2005. Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* **21**: 795–837.

Breitung J, Pesaran MH. 2008. Unit roots and cointegration in panels. In *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Matyas L, Sevestre P (eds). Kluwer: Dordrecht (forthcoming).

Brucker H, Siliverstovs B. 2006. On the estimation and forecasting of international migration: how relevant is heterogeneity across countries? *Empirical Economics* **31**: 735–754.

DOI: 10.1002/for

Canova F, Ciccarelli M. 2004. Forecasting and turning point predictions in a Bayesian panel VAR model. *Journal of Econometrics* **120**: 327–359.
Chamberlain G, Hirano K. 1999. Predictive distributions based on longitudinal earnings data. *Annales D'Économie et de Statistique* **55–56**: 211–242.
Clements M, Hendry D. 2005. *A Companion to Economic Forecasting*. Blackwell: Oxford.
Das M, Dominitz J, van Soest A. 1999. Comparing predictions and outcomes: theory and application to income changes. *Journal of the American Statistical Association* **94**: 75–85.
Diebold FX. 2004. *Elements of Forecasting*. South-Western: Cincinnati, OH.
Diebold FX, Lopez JA. 1996. Forecast evaluation and combination. In *Handbook of Statistics*, Maddala GS, Rao CR (eds). North-Holland: Amsterdam; 241–268.
Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–264.
Driver C, Imai K, Temple P, Urga A. 2004. The effect of uncertainty on UK investment authorisation: homogeneous vs. heterogeneous estimators. *Empirical Economics* **29**: 115–128.
Fok D, van Dijk D, Franses PH. 2005. Forecasting aggregates using panels of nonlinear time series. *International Journal of Forecasting* **21**: 785–794.
Frees EW, Miller TW. 2004. Sales forecasting using longitudinal data models. *International Journal of Forecasting* **20**: 99–114.
Frees EW, Young V, Luo Y. 1999. A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics* **24**: 229–247.
Frees EW, Young V, Luo Y. 2001. Credibility ratemaking using panel data models. *North American Actuarial Journal* **5**: 24–42.
Fuller WA, Battese GE. 1974. Estimation of linear models with cross-error structure. *Journal of Econometrics* **2**: 67–78.
Galbraith JW, Zinde-Walsh V. 1995. Transforming the error-component model for estimation with general ARMA disturbances. *Journal of Econometrics* **66**: 349–355.
Garcia-Ferrer A, Highfield RA, Palm F, Zellner A. 1987. Macroeconomic forecasting using pooled international data. *Journal of Business and Economic Statistics* **5**: 53–76.
Gavin WT, Theodorou AT. 2005. A common model approach to macroeconomics: using panel data to reduce sampling error. *Journal of Forecasting* **24**: 203–219.
Giacomini R, Granger CWJ. 2004. Aggregation of space-time processes. *Journal of Econometrics* **118**: 7–26.
Goldberger AS. 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* **57**: 369–375.
Granger CWJ, Huang L. 1997. Evaluations of panel data models: some suggestions from time series. *Discussion paper 97-10*, University of California, San Diego.
Groen JJJ. 2005. Exchange rate predictability and monetary fundamentals in a small multi-country panel. *Journal of Money, Credit, and Banking* **37**: 495–516.
Harville DA. 1976. Extension of the Gauss–Markov theorem to include the estimation of random effects. *Annals of Statistics* **4**: 384–395.
Harville DA. 1985. Decomposition of prediction error. *Journal of the American Statistical Association* **80**: 132–138.
Henderson CR. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**: 423–447.
Hendry DF, Hubrich K. 2006. Forecasting economic aggregates by disaggregates. *Working paper no. 589*, European Central Bank.
Hjalmarsson E. 2006. Predictive regressions with panel data. *International Finance Discussion Papers*, Board of Governors of the Federal Reserve System, Washington, DC.
Holtz-Eakin D, Selden TM. 1995. Stocking the fires? CO_2 emissions and economic growth. *Journal of Public Economics* **57**: 85–101.
Hoogstrate AJ, Palm FC, Pfann GA. 2000. Pooling in dynamic panel-data models: an application to forecasting GDP growth rates. *Journal of Business and Economic Statistics* **18**: 274–283.
Hsiao C, Tahmiscioglu AK. 1997. A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association* **92**: 455–465.

DOI: 10.1002/for

Hsiao C, Pesaran MH, Tahmiscioglu AK. 1999. Bayes estimation of short run coefficients in dynamic panel data models. In *Analysis of Panels and Limited Dependent Variable Models*, Hsiao C, Lahiri K, Lee LF, Pesaran MH (eds). Cambridge University Press: Cambridge, UK; 268–296.

Judge GG, Bock ME. 1978. *The Statistical Implications of Pre-test and Stein-Rule Estimators in Econometrics.* North-Holland: Amsterdam.

Kackar RN, Harville D. 1984. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* **79**: 853–862.

Karlsson S, Skoglund J. 2004. Maximum-likelihood based inference in the two-way random effects model with serially correlated time effects. *Empirical Economics* **29**: 79–88.

Keane MP, Runkle DE. 1990. Testing the rationality of price forecasts: new evidence from panel data. *American Economic Reviews* **80**: 714–735.

Koop G, Potter S. 2003. Forecasting in large macroeconomic panels using Bayesian model averaging. *Federal Reserve Bank of New York Staff Reports*, no. 163.

Lahiri K, Liu F. 2006. Modelling multi-period inflation uncertainty using a panel of density forecasts. *Journal of Applied Econometrics* **21**: 1111–1326.

Lee LF, Griffths WE. 1979. The prior likelihood and best linear unbiased prediction in stochastic coefficient linear models. Working paper, Department of Economics, University of Minnesota.

MaCurdy TA. 1982. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics* **18**: 83–114.

Maddala GS, Hu W. 1996. The pooling problem. In *The Econometrics of Panel Data: A Handbook of Theory with Applications*, Màtyàs L, Sevestre P. (eds). Kluwer: Dordrecht; 307–322.

Maddala GS, Trost RP, Li H, Joutz F. 1997. Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business and Economic Statistics* **15**: 90–100.

Marcelino M, Stock JH, Watson M. 2003. Macroeconomic forecasting in the EURO area: country specific versus area-wide information. *European Economic Review* **47**: 1–18.

Mark NC, Sul D. 2001. Nominal exchange rates and monetary fundamentals: evidence from a small post-Bretton Woods panel. *Journal of International Economics* **53**: 29–52.

Mouchart M, Rombouts JVK. 2005. Clustered panel data models: an efficient approach for nowcasting from poor data. *International Journal of Forecasting* **21**: 577–594.

Nandram B, Petruccelli JD. 1997. A Bayesian analysis of autoregressive time series panel data. *Journal of Business and Economic Statistics* **15**: 328–334.

Newbold P, Harvey DI. 2002. Forecast combination and encompassing. In *A Companion to Economic Forecasting*, Clements MP, Hendry DF (eds). Blackwell: Oxford; 268–283.

Pesaran MH, Smith R. 1995. Estimating long-run relationships from dynamic heterogenous panels. *Journal of Econometrics* **68**: 79–113.

Pesaran MH, Smith R, Im KS. 1996. Dynamic linear models for heterogenous panels. In *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Màtỳas L, Sevestre P (eds), Kluwer: Dordrecht; 145–195.

Pesaran MH, Shin Y, Smith R. 1999. Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association* **94**: 621–634.

Pesaran MH, Schuermann T, Weiner S. 2004. Modelling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business and Economics Statistics* **22**: 129–162.

Rapach DE, Wohar ME. 2004. Testing the monetary model of exchange rate determination: a closer look at panels. *Journal of International Money and Finance* **23**: 867–895.

Revankar NS. 1979. Error component models with serial correlated time effects. *Journal of the Indian Statistical Association* **17**: 137–160.

Robertson D, Symons J. 1992. Some strange properties of panel data estimators. *Journal of Applied Econometrics* **7**: 175–189.

Robinson GK. 1991. That BLUP is a good thing: the estimation of random effects. *Statistical Science* **6**: 15–32.

Schmalensee R, Stoker TM, Judson RA. 1998. World carbon dioxide emissions: 1950–2050. *Review of Economics and Statistics* **80**: 15–27.

Stock JH, Watson MW. 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* **23**: 405–430.

Swamy PAVB. 1970. Efficient inference in a random coefficient regression model. *Econometrica* **38**: 311–323.

DOI: 10.1002/for

Taub AJ. 1979. Prediction in the context of the variance-components model. *Journal of Econometrics* **10**: 103–108.

Wansbeek TJ, Kapteyn A. 1978. The separation of individual variation and systematic change in the analysis of panel data. *Annales de l'INSEE* **30–31**: 659–680.

Zellner A, Hong C. 1989. Forecasting international growth rates using Bayesian shrinkage and other procedures. *Journal of Econometrics* **40**: 183–202.

Zellner A, Hong C, Min C. 1991. Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameters, and pooling techniques. *Journal of Econometrics* **49**: 275–304.

Zhang H. 2002. On estimation and prediction for spatial generalized linear mixed models. *Biometrics* **58**: 129–136.

Ziemer RF, Wetzstein ME. 1983. A Stein-rule method for pooling data. *Economics Letters* **11**: 137–143.

Author's biography:
Badi H. Baltagi is distinguished Professor of Economics, and Senior Research Associate at the Center for Policy Research, Syracuse University. He received his Ph.D. in Economics at the University of Pennsylvania in 1979. He is also part-time Chair in Economics, University of Leicester. He is the author of *Econometric Analysis of Panel Data (Wiley)*; *Econometrics (Springer)*, and editor of *A Companion to Theoretical Econometrics (Blackwell)*; *Recent Developments in the Econometrics of Panel Data*, *Volumes I and II (Edward Elgar)*; *Nonstationary Panels*, *Panel Cointegration*, *and Dynamic Panels (Elsevier)*; *Panel Data Econometrics: Theoretical Contributions and Empirical Applications (Elsevier)* and author or co-author of over 100 publications, all in leading economics and statistics journals. Professor Baltagi is co-editor of *Empirical Economics*, and associate editor of *Journal of Econometrics* and *Econometric Reviews*. He is the replication editor of the *Journal of Applied Econometrics* and the series editor for *Contributions to Economic Analysis.* He is a fellow of the *Journal of Econometrics* and a recipient of the Multa and Plura Scripsit Awards from *Econometric Theory*.

Author's address:
Badi H. Baltagi, Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020, USA.

DOI: 10.1002/for