Optimal Forecasting with Heterogeneous Panels: A Monte Carlo Study.

Lorenzo Trapani[*]
Cass Business School, London (U.K.)
Universita' di Bergamo (Italy)

Giovanni Urga
Cass Business School, London (U.K.)
Universita' di Bergamo (Italy)

First version: 20 February 2007. Second version: 15 March 2008. This version: 11 december 2008.

Abstract

We contrast the forecasting performance of alternative panel estimators, divided into three main groups: homogeneous, heterogeneous and shrinkage/Bayesian. Via a series of Monte Carlo simulations, the comparison is done using different levels of heterogeneity and cross sectional dependence, alternative panel structures in terms of T and N and the specification of the dynamics of the error term. To assess the predictive performance, we use traditional measures of forecast accuracy (Theil's U statistics, RMSE and MAE), the Diebold and Mariano's (1995) test, and the Pesaran and Timmerman's (1992) statistics on the capability of forecasting turning points. The main finding of our analysis is that when the level of heterogeneity is high, shrinkage/Bayesian estimators are preferred, whilst when there is low or mild heterogeneity homogeneous estimators are the ones with the best forecast accuracy.

[*]Corresponding author: Cass Business School, Faculty of Finance, Centre for Econometric Analysis, 106 Bunhill Row, London EC1Y 8TZ (U.K.). Tel. +/44/20/70405260; Fax.: +/44/20/70408881. e-mail: L.Trapani@city.ac.uk; http://www.cass.city.ac.uk/faculty/l.trapani/.

Optimal Forecasting with Heterogeneous Panels: A Monte Carlo Study.

We contrast the forecasting performance of alternative panel estimators, divided into three main groups: homogeneous, heterogeneous and shrinkage/Bayesian. Via a series of Monte Carlo simulations, the comparison is done using different levels of heterogeneity and cross sectional dependence, alternative panel structures in terms of T and N and the specification of the dynamics of the error term. To assess the predictive performance, we use traditional measures of forecast accuracy (Theil's U statistics, RMSE and MAE), the Diebold and Mariano's (1995) test, and the Pesaran and Timmerman's (1992) statistics on the capability of forecasting turning points. The main finding of our analysis is that when the level of heterogeneity is high, shrinkage/Bayesian estimators are preferred, whilst when there is low or mild heterogeneity homogeneous estimators are the ones with the best forecast accuracy.

J.E.L. Classification Numbers: C12, C13, C23, C33.

Keywords: Heterogeneity; Cross dependence; Forecasting; Monte Carlo simulations.

1 Introduction

Over the last two decades a variety of estimation techniques have been proposed to estimate parameters of interest when panel data are available: Arellano and Honore' (2001), Wooldridge (2002), Hsiao (2003), Arellano (2003) and Baltagi (2005) provide comprehensive surveys on the topic. It has become customary to group these techniques into three main groups: homogeneous, heterogeneous and Bayesian (or shrinkage) estimators. While the first class assumes poolability of the data in the panel, and therefore parameters homogeneity across the panel units, the second one rejects this hypothesis taking into account explicitly the presence of heterogeneity among units. The class of Bayesian estimators is viewable as a hybrid solution between the two other classes (see Maddala, Li and Srivatsava, 1994, and Pesaran, Hsiao and Tahmiscioglu, 1999). It becomes then crucial to understand which estimation method is the "best", in statistical terms, for the specific research interest (e.g. bias reduction, efficiency, forecasting accuracy...).

Recently, in several seminal empirical papers Professor Badi Baltagi and associates have focused on investigating which estimator is the "best" when the specified model has to be used for forecast purposes. Baltagi and Griffin (1997), Baltagi, Griffin and Xiong (2000), Baltagi, Bresson, Griffin and Pirotte (2003) and Baltagi, Bresson and Pirotte (2002) apply dynamic panel specifications to industrial level data and find that the predictive ability of homogeneous estimators outperforms the predictive ability of heterogeneous and Bayesian estimators over any forecast horizon. Amongst the homogeneous estimators, GLS and within-2SLS emerge as the best estimators for forecasting purposes, especially when we forecast over a long time span. The superiority of the homogeneous estimators can sound quite reasonable when the panel is short, and when the degree of heterogeneity across units is limited, but it is rather puzzling when the time length T of the panel is large or when the degree of heterogeneity is high. This genuine empirical finding is particularly interesting because the model where we impose homogeneity is in general rejected by the data. A first interpretation of this apparent counter-intuitive empirical regularity is that a model that is "simple and parsimonious" offers a better forecasting performance. However, using a different dataset, Baltagi, Bresson and Pirotte (2004) find that Bayesian estimators provide the best forecasting performance.

It becomes therefore worth investigating whether these results hold generally speaking or if they are properties of the data considered in the works cited above, or, possibly, if the outcome of the comparison among the estimators forecasting performance depends on the number of units N and the time length of the panel T, and on the degree of the parameters heterogeneity across units. Our main objective in this work is to compare via a broad Monte Carlo simulation exercise the forecasting accuracy of several estimators belonging to each of the three classes (homogeneous, heterogeneous and shrinkage) for a routinely applied model (the dynamic specification with one or more exogenous covariates) under various circumstances. Such "circumstances" are the pair (N, T), the level of heterogeneity among units, the dynamic specification of the error term, and the existence and degree of cross sectional dependency across units. These issues are of paramount importance in determining the properties of estimators.

An important related question that arises in these circumstances is how to assess forecasting performance of a model. In their papers, Baltagi and associates use the standard Root Mean Square Error (RMSE) to measure forecasting accuracy. However, the literature on forecasting has developed a quite critical attitude towards this classical statistical measure. Thus in addition to the method based on RMSE, in our Monte Carlo experiments we use also the approach based on different specifications for the loss function (Diebold and Mariano, 1995), and the non parametric statistic that evaluates the ability to forecast change points due to Pesaran and Timmermann (1992). Our main findings show that the degree of heterogeneity plays a crucial role, whilst other features of the data have a very limited impact on the predictive ability of various panel estimators. When heterogeneity is low or mild, homogeneous estimators have the best predictive ability, whereas when heterogeneity is high, shrinkage/Bayesian procedures are preferable.

The remainder of this paper is as follows. We set out the model we will be considering for our exercise, and briefly describe the estimation techniques and the predictive performance tests that we employ in our experiments (Section 2). We describe the details of the Monte Carlo experiments in Section 3, and report and comment the main results from the simulations in Section 4. Section 5 concludes.

2 ESTIMATION AND FORECASTING

2.1 Model

The data generating process (DGP) we employed for simulation is based on a dynamic specification and one strictly exogenous/predetermined variable:

$$y_{it} = \alpha_i + \beta_i y_{it-1} + \gamma_i x_{it} + u_{it} \tag{1}$$

where $i = 1, .., N$ and $t = 1, .., T$. Without loss of generality, the error term u_{it} is assumed to have no time specific effects since we focus on the impact of grouping across units. The possibility of having cross sectional dependence - i.e. the case $E[u_{it}u_{js}] \neq 0$ for some pair (i, j) - is not ruled out. Model (1) is the classical dynamic panel data specification, as discussed extensively in Baltagi (2005). It is also worth emphasizing that what we consider in our exercise are ex post forecasts, i.e. forecasts where the exogenous variable in model (1) is assumed known.

As far as estimation is concerned, we employed both homogeneous and heterogeneous estimators, performing an exercise similar to that in Baltagi and Griffin (1997), Baltagi, Griffin and Xiong (2000), Baltagi, Bresson, Griffin and Pirotte (2003) and Baltagi, Bresson and Pirotte (2002, 2004). Notice that whilst heterogeneous estimators are based on model (1), homogeneous estimators, assuming poolability of the data, are based on the following restricted specification of the DGP:

$$y_{it} = \alpha + \beta y_{it-1} + \gamma x_{it} + \varepsilon_{it}. \tag{2}$$

The error term ε_{it} is assumed to follow the well known one way specification:

$$\varepsilon_{it} = \mu_i + u_{it}, \tag{3}$$

where μ_i is the unobservable individual specific effect defined as

$$\mu_i = (\alpha_i - \alpha) + (\beta_i - \beta)\, y_{it-1} + (\gamma_i - \gamma)\, x_{it}, \tag{4}$$

and u_{it} is the remainder of the disturbance - see Baltagi (2005) for a thorough discussion. The results of pooling using model (2) on estimators are discussed in Pesaran and Smith (1995) and Hsiao, Pesaran and Tahmiscioglu (1999).

2.2 Homogeneous, heterogeneous and shrinkage/Bayesian estimators

We turn our discussion to estimation, referring to Baltagi (2005) for the details of each estimator.

2.2.1 Homogeneous estimators

The homogeneous estimators we consider fall into two main groups: least squares and instrumental variables estimators.

Within the class of least squares estimators, we first consider six standard pooled estimators applied to model (2): OLS, which ignores unit specific effects; first difference OLS to wipe out the effect of (possible) serial correlation in the error term; Within(-groups) estimator, which allows for unit specific effects; Between(-groups) estimator; and WLS and WLS-AR(1), where unit specific effects are assumed to be random. It is known that none of these estimates is either unbiased or consistent (see Pesaran and Smith, 1995, and the review in Baltagi, 2005). This is due to the assumption, common to all these estimators, that regressors are exogenous. However, the model we consider is dynamic and thus though all the explanatory variables are uncorrelated with the error components, the presence of either serial correlation in the remainder error term ν_{it} or of a random unit effect such as μ_i makes the lagged dependent variable correlated with the error term and therefore leads to potentially inconsistent estimates. The asymptotic bias of OLS has been assessed by Sevestre and Trognon (1985); it is also well known (see Nickell, 1981), that Within estimator is consistent only when $T \to \infty$, being biased of order $O(1/T)$ for finite T. The random effect WLS estimator is also biased and inconsistent, as pointed out in Baltagi (2005).

To achieve consistency, we may focus on pooled estimators based on instrumental variables. Thus, we first employ a standard 2SLS, which is consistent but not efficient; no attempt was made to improve efficiency by taking into account the unit specific effects. We also consider Within 2SLS, which, like its least squares counterpart, wipes out unit specific effects by transforming the data in deviations across their mean, and the Between 2SLS. Thirdly, we apply 2SLS to the first differenced version of model (2); this estimator (that henceforth will be referred to as FD-2SLS) is due to Anderson and Hsiao (1982) and is meant to eliminate fixed and random effects. However, given that this estimation procedure may induce autocorrelation in the remainder error term $\nu_{it} - \nu_{it-1}$, we also employ the correction proposed by Keane and Runkle (1992) that allows for arbitrary types of serial correlation[1]. This is applied to both the specification in levels (leading to an estimator denoted as 2SLS-KR) and the first differenced model (obtaining another estimate referred to as FD-2SLS-KR). Also, we employ EC2SLS estimator - see Baltagi (2005) - and EC2SLS-AR(1) - see Baltagi, Griffin and Xiong (2000) - to potentially achieve more efficiency by taking account of possible serial correlation in the error term[2]. As a variant of EC2SLS, we also compute the G2SLS estimator due to Balestra and Varadharajan-Krishnakumar (1987). It is worth noticing that such estimator has the same asymptotic covariance matrix as EC2SLS - see Baltagi and Li (1992) - but its performance is different in finite samples[3].

[1] Such estimation technique can be applied only if $N > T$ - see Baltagi (2005).

[2] Note that these estimators, unlike standard 2SLS, also require an estimate of the variance components in order to be feasible.

[3] We also considered employing the Arellano and Bond (1991) estimation procedure, using a GMM estimation method on the specification in differences (whose outcome will

Finally, we considered the MLE (see Baltagi, 2005) using the iterative procedure suggested by Breusch (1987).

In total, we compare 18 homogeneous estimators.

2.2.2 Heterogeneous estimators

The estimators considered so far are all characterized by the assumption of poolability of the data. This is a valid assumption only if the parameters in model (1) are homogeneous across units. As pointed out by Pesaran and Smith (1995) with respect to the dynamic pooled model, when parameters are heterogeneous, pooling leads to biased estimates. Therefore, we turned our attention also onto heterogeneous estimators.

In our Monte Carlo experiments we considered OLS and 2SLS applied to each unit i, obtaining Individual OLS and 2SLS. Given the presence of a lagged dependent variable, both estimates are biased. We then consider an average of both estimates (obtaining Average OLS and 2SLS), as suggested by Pesaran and Smith (1995). Averaging individual estimates leads to a consistent estimator as long as both N and T tend to infinity. We also compute the Swamy (1970) estimator, which belongs to the class of GLS and is a weighted average of the least squares estimates, using as weights the estimated covariance matrix.

In total we compare 5 alternative heterogeneous estimators.

2.2.3 Shrinkage/Bayesian estimators

We employed a class of shrinkage/Bayesian estimators - see Maddala, Li and Srivastava (1994) - where each individual estimate is shrunk towards the pooled estimates by weighing it with weight depending on the corresponding covariance matrix. The authors claim that shrinkage type estimator are superior to the homogeneous and to the other heterogeneous estimators as far as predictive ability is concerned. The estimators we consider are the Empirical Bayes based on OLS initialization, the Empirical Bayes based on 2SLS estimation and their iterative counterparts. Finally, we implement the Hierarchical Bayes estimator using the same prior structure as in Hsiao, Pesaran and Tahmiscioglu (1999), which is found to have the best performance among heterogeneous estimators in terms of bias reduction, especially when T is small.

be labelled as FDGMM) and also the same set of instruments in first difference on a specification in levels (GMM). However GMM estimation procedures are only feasible when $N > T(k-2) + (T+3)/2$, where k is the number of parameters. These estimators would not have been feasible for all the cases we consider in our experiment, and we did not perform them. GAUSS code was anyway written and is available upon request.

In total, we compare 5 alternative Bayesian estimators.

2.3 Comparing forecasting performance

In this section we introduce the measures of forecasting performance we employ in our simulation exercise.

We employ three (classes of) measures of forecasting performance to assess the out-of-sample predicting ability of each estimator:

1. statistical measures of accuracy;
2. measure of statistical assessment of performance.
3. measures of the capability of predicting turning points.

The indicators we chose are, for each class:

1. mean and median absolute error (referred to as MAE and MdAE respectively), RMSE and Theil's U statistics, whose expressions are respectively

$$MAE_j \equiv \frac{1}{h}\sum_{i=1}^{h} |\hat{y}_{ji} - y_{ji}|$$

$$MdAE_j = median_{1\leq j\leq h} |\hat{y}_{ji} - median_{1\leq j\leq h}(y_{ji})|$$

$$RMSE_j \equiv \sqrt{\frac{1}{h}\sum_{i=1}^{h} (\hat{y}_{ji} - y_{ji})^2}$$

$$U_j \equiv \sqrt{\frac{\sum_{i=1}^{h} (\hat{y}_{ji} - y_{ji})^2}{\sum_{i=1}^{h} y_{ji}^2}}$$

where the index j refers to the j-th unit in the panel, h is the forecast horizon, $\hat{y}_{ji}$ is the forecast i steps ahead of y_{ji} and $median_{1\leq j\leq h}(y_{ji})$ is the median of the sample $\{y_{ji}\}_{j=1}^{h}$. To obtain a single overall measure of performance, we considered the average of each indicator across units, similarly to Baltagi and associates papers. These indicators are all based on the residuals from forecast, and widely employed in the realm of forecasting. We calculate these "classical" measures but we report and comment on the Theil's U statistics only, given its nature of relative measure which doesn't have the scaling problem of both RMSE and MAE. It is necessary to point out that using these indicators to assess forecasting accuracy has been widely criticized on the basis of statistical and economic considerations - for a general overview, see the review in Mariano (2002). From a statistical point of view, Clements and Hendry (1993) noted that the RMSE is not invariant to isomorphic transformations of models, and can therefore lead to contradictory results when applied to different (but isomorphic) representations of the same model. Moreover, Diebold and Lopez (1996) show that since RMSE depends only on the first two moments of the forecast distribution, it will suffer from serious shortcomings when such distribution is not adequately summarized by only two moments. The literature has criticized RMSE also on the grounds of economic considerations, arguing that predictive performance should be evaluated via the losses that arise from forecasting errors when certain decisions are made - see Leitch and Tanner (1991), Granger and Pesaran (2000a, 2000b), and the review by Pesaran and Skouras (2002). It has been shown that the RMSE is compatible with a quadratic loss function - see Pesaran and Skouras (2002) - but other specifications could be considered - see the discussions in Christoffersen and Diebold (1996) and Mariano (2002).

2. Diebold and Mariano's (1995) test is a widely used alternative to overcome the inadequacies of RMSE since it is based on a loss function approach without needing specify the functional form. This statistics - with the adjustment for small sample bias proposed by Harvey, Leybourne and Newbold (1997) - can be used for any forecasting horizon h and doesn't require gaussianity, zero-mean, serial or contemporaneous uncorrelation of the forecast errors, and under the null hypothesis of no difference between forecasting performances it is distributed as a standard normal. Formally, this statistic can be obtained as follows. Let $d_{ji}^k = \hat{y}_{ji} - y_{ji}$ be the forecast error at period i for series j when estimating parameters with an estimator k; assuming covariance stationarity and other regularity conditions, it is straightforward to show that
$$T^{-1/2}\left(\bar{d}_j - \mu_d\right) \Rightarrow N\left[0, 2\pi f\left(0\right)\right],$$
where $f\left(0\right)$ is the spectral density at frequency zero, $\mu_d = E\left(d_{ji}^k\right)$ and
$$\bar{d}_j = \sum_{i=1}^{h}\left[g\left(d_{ji}^1\right) - g\left(d_{ji}^2\right)\right]$$
with $g\left(\cdot\right)$ a generic loss function. Hence, the DM test is designed to compare the performance of two predictors; computationally, the sta-

tistic is set equal to

$$DM_j = \frac{\bar{d}_j}{\left[2\pi \hat{f}\,(0)\,/T\right]^{1/2}}.$$

The loss function we consider in order to compute the statistics is a quadratic one, which allows us to compare pairwise RMSEs.[4] This enables us to detect whether one estimator has a superior predictive ability compared to another one by a proper testing rather than by the pure comparison of RMSE values. Even in this case, we compute the test statistics for every unit of the panel and then take the average across units.

3. Forecasting performance could refer to something different from minimizing a loss function, such as the capability to capture the sign of changes in the series rather than its values - see Granger and Pesaran (2000b). We employ Pesaran and Timmerman's (1992) statistics, defined as

$$PT_j = \frac{\hat{P}_j - \hat{P}_j^*}{\sqrt{\hat{V}\left(\hat{P}_j\right) - \hat{V}\left(\hat{P}_j^*\right)}} \sim N(0,1)$$

where

$$\hat{P}_j = h^{-1}\sum_{i=1}^{h} sign\left(\hat{y}_{ji}y_{ji}\right), \quad \hat{P}_j^* = \hat{P}_{yj}\hat{P}_{xj} + \left(1-\hat{P}_{yj}\right)\left(1-\hat{P}_{xj}\right),$$

$$\hat{V}\left(\hat{P}_j\right) = h^{-1}\hat{P}_j^*\left(1-\hat{P}_j^*\right),$$

$$\begin{aligned}\hat{V}\left(\hat{P}_j^*\right) &= h^{-1}\left(2\hat{P}_{yj}-1\right)^2\hat{P}_{xj}\left(1-\hat{P}_{xj}\right) + h^{-1}\left(2\hat{P}_{xj}-1\right)^2\hat{P}_{yj}\left(1-\hat{P}_{yj}\right) + \\ &\quad +4h^{-2}\hat{P}_{yj}\hat{P}_{xj}\left(1-\hat{P}_{yj}\right)\left(1-\hat{P}_{xj}\right)\end{aligned}$$

$$\hat{P}_{xj} = h^{-1}\sum_{i=1}^{h} sign\left(\hat{y}_{ji}\right), \quad \hat{P}_{yj} = h^{-1}\sum_{i=1}^{h} sign\left(y_{ji}\right),$$

[4] The Diebold and Mariano testing procedure also requires a non parametric estimate of the spectral density of the difference of the loss associated with each predictor. The kernel we employ is the truncated rectangular one employed by Diebold and Mariano (1995), and the bandwidth we choose is specified as $m\,(h) = 1 + \lfloor \log\,(h) \rfloor$, where the operator $\lfloor \cdot \rfloor$ denotes the rounding to the closest integer.

where the function $sign(\cdot)$ takes the value of unity if its argument is positive and is equal to zero otherwise. Pesaran and Timmerman (1992) prove that this non parametric statistics is distributed as a standard normal under the null hypothesis that $\hat{y}_{ji}$ and y_{ji} are independent - and therefore that the predictor $\hat{y}_{ji}$ has no capability to forecast y_{ji}. Like in the previous point, here we compute the Pesaran and Timmerman statistics for each unit of the panel and then report its average value across units. Notice that this measure could be also employed as a descriptive measure to rank forecasting techniques (see *inter alia* Driver and Urga, 2004).

Having described the estimators considered and the methods of evaluating forecasting accuracy, in the next section we illustrate the design of the Monte Carlo experiment.

3 THE DESIGN OF THE MONTE CARLO EXPERIMENTS

We generate a sample of N units with length $T+T_0$, where T_0 is the number of initial values to be discarded to avoid dependence on the initial conditions (set equal to 0). We let the number of units N and the time dimension T assume various values.

The DGP we generate at each replication is the one given in model (1):

$$y_{it} = \alpha_i + \beta_i y_{it-1} + \gamma_i x_{it} + u_{it},$$

where:

- the parameters α_i, β_i and γ_i are generated as, respectively:

$$\alpha_i = \bar{\alpha} + HN_i^{\alpha},$$
$$\beta_i = \bar{\beta} + HU_i^{\beta},$$
$$\gamma_i = \bar{\gamma} + HN_i^{\gamma},$$

  where $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ are the mean values of the parameters, $N_i^{\cdot}$ denotes an independent (across i) extraction from a normal random variable and H is a parameter that controls heterogeneity across units, which will be useful throughout the set of simulations to assess the predictive performance of the estimators. Notice that whilst we employed standard normals for α_i and γ_i, β_i was simulated via a uniform distribution (U_i^{β}) with bounded support so as to rule out the possibility of having a value larger than (or equal to) unity;

- the disturbance u_{it} is, in a first set of experiments, assumed to follow a stationary, invertible Gaussian ARMA(1,1) specification defined by

$$u_{it} = \rho u_{it-1} + \zeta_{it} + \vartheta \zeta_{it-1},$$

and the parameters (ρ, ϑ) control the degree of autocorrelation of the error term in model (1). The error term is then rescaled by the factor $\lambda = \sqrt{(1+\vartheta)/(1-\rho)}$ to give it unit variance. Here there is no cross sectional dependence across units, since u_{it} is generated independently of u_{jt} for any pair (i, j). In a second set of experiments, we take into account the presence of cross sectional dependence by modelling the error (now denoted as $u_{it}^{'}$) as

$$u_{it}^{'} = u_{it} + \zeta_i f_t,$$

where f_t is a standard normal independent over t and ζ_i is a uniformly distributed random variable whose support is chosen as $[0, 0.2]$ to model small cross section dependence and $[-1, 3]$ to represent a large amount of cross section dependence. This part of the experiments to modelling cross sectional dependence follows the same lines as Pesaran (2007);

- the explanatory variable x_{it} is generated with the following DGP:

$$x_{it} = \alpha_i + \beta_i + \delta x_{it-1} + \eta_{it}, \tag{5}$$

where the error term η_{it} is a Gaussian white noise generated independently of u_{it}. Thus, the presence of the term $\alpha_i + \beta_i$ introduces a correlation between x_{it} and the error term (3) in the random effect specification (2)

$$\varepsilon_{it} = \mu_i + u_{it}.$$

This correlation is such that $E(x_{it}u_{it}) = 0$ for any i since all the quantities on the right hand side of (5) are generated independently of u_{it} - and hence x_{it} endogeneity is ruled out - and $E(x_{it}\mu_i) \neq 0$.[5] This two results make x_{it} a strictly exogenous variable and a valid instrument for GMM estimation *a la* Arellano and Bond (1991) thanks to its correlation with the unit specific effect - see Baltagi (2005) for discussion.

[5] This can be seen combining (4) and (5), which (after some algebra) leads to

$$E(x_{it}\mu_i) = E[\alpha_i(\alpha - \alpha_i)] + E[\beta_i(\beta - \beta_i) y_{it-1}] \neq 0.$$

As our results reported in Section 4 show, the degree of heterogeneity H plays a pivotal role in determining the rank of the predictive abilities of the various estimation techniques. We considered two separate cases, namely $H = 0.1$ and $H = 0.9$ to represent the cases of "mild" heterogeneity and "high" heterogeneity respectively. We base our choice on the grounds of the empirical results in Baltagi and Griffin (1995), Baltagi, Griffin and Xiong (2000), Baltagi, Bresson, Griffin and Pirotte (2003) and Brucker and Silivertovs (2006), where some order statistics (mainly the maximum, the minimum, and the median) are reported for the individual estimated coefficients in the panel regression. The discrepancy between maximum and mininum estimated values differ (sometimes substantially) depending on the type of heterogeneous estimator considered, and the models employed in the papers cited above have different specifications and therefore a different number of parameters to our exercise. However, it is possible to derive some guidelines as to the degree of heterogeneity in the data considered in the empirical exercises in Baltagi and Griffin (1997), Baltagi, Griffin and Xiong (2000), Baltagi, Bresson, Griffin and Pirotte (2003) and Brucker and Silivertovs (2006). Assuming that unit specific parameters have a normal distribution around the estimated medians and standard deviation H, and letting n be the number of cross sectional units, the (expected) maximum is given by $H\left(a_n\gamma + b_n\right)$, where $\gamma = 0.5772$ is Euler's constant and

$$\begin{aligned} a_n &= 1/\sqrt{2\log n}, \\ b_n &= \frac{1}{\sqrt{2\log n}} - \frac{1}{2}\frac{\log\log n + \log\left(4\pi\right)}{\sqrt{2\log n}}. \end{aligned}$$

We compute the level of heterogeneity H for each estimator as

$$H = \frac{d_{Mm}}{2\left(a_n\gamma + b_n\right)},$$

where d_{Mm} is the difference between the maximum and the minimum estimate. We subsequently calculated an average level of heterogeneity based on averaging (1) the levels of heterogeneity found using different estimators and (2) the levels of heterogeneity found estimating different parameters when models have more than one exogenous variable. Based on this approach, we find that the levels of heterogeneity found in the datasets used by Baltagi, Griffin and Xiong (2000) and Baltagi, Bresson, Griffin and Pirotte (2003) are equal to 0.176 and 0.183 respectively. Baltagi and Griffin (1995) and Brucker and Silivertovs (2006) find higher levels of H, equal to 0.323 and 0.428 respectively. Thus, in our experiments we set $H = 0.1$ to represent low levels of heterogeneity, $H = 0.5$ to represent a mild level of heterogeneity, and we also choose $H = 0.9$ to explore the robustness and sensitivity of the rankings of the forecasting performance relative to various estimators.

As far as the other parameters in our simulation exercise, we considered the following values:

- we ran 5000 iterations for each simulation, and 2500 iterations (500 of which in the burn-in period) for every Gibbs sampling algorithm - on the ground of the results in Hsiao, Pesaran and Tahmiscioglu (1999);
- as far as the autocorrelation structure is concerned, we considered (ρ, ϑ) to be equal either to $(0, 0)$ or to $(0.9, 0.9)$. These two pairs represent the cases of non autocorrelation and of near integration, respectively;
- the number of initial observations to be discarded was set equal to $T_0 = 100$;
- the forecasting horizon is set equal to $h = 10$, though our results can be extended to the cases $h = 1$ and $h = 5$, as in various papers by Baltagi and associates.

4 SIMULATION RESULTS

In this section we report and comment the full set of results from the various Monte Carlo experiments using the three forecasting accuracy tests. We consider three different degrees of heterogeneity given by $H = 0.1$, 0.5 and 0.9 respectively; two different specifications for the error dynamics, namely (ρ, ϑ) were set equal to $(0, 0)$ and $(0.9, 0.9)$; in addition to the case of no cross dependence, two alternative degree of cross sectional dependence are considered, namely the case of "mild" cross dependence (ζ_i $[0, 0.2]$) and one with "large" cross sectional dependence (ζ_i is now $[-1, 3]$). Finally, the pairs of (T, N) we consider are $(5, 10)$, $(5, 20)$, $(10, 20)$, $(10, 50)$, $(20, 50)$ and $(50, 50)$.

The presentation of the full set of experiments are reported in details in a companion paper (Trapani and Urga, 2005).

4.1 Theil's U statistic

In this section we report the rankings of the various estimation techniques based on Theil's U statistic[6]. Each table is divided in three panels. We report

[6] The full set of results are reported in Trapani and Urga (2005), Tables A1-A12. We also computed RMSEs and the MAEs for each simulation. The findings remain unchaged. The results are available upon request.

the statistics for the homogeneous, heterogeneous and shrinkage/Bayesian estimators respectively.

The main findings can be summarized as follows.

Our results show that heterogeneity plays a very important role and has a strong impact on the outcomes of the simulation exercises. The results we obtained for the case of mild heterogeneity, $H = 0.5$, were essentially the same as those with low levels of heterogeneity, $H = 0.1$, and therefore we report and comment only the latter case. When the degree of heterogeneity is low (columns with $H = 0.1$ in the Tables) and the amount of dependence among units is mild, homogeneous estimators prevail; this is a common feature found in the case of mild cross dependence (Tables 1, 3, 5, 7, 9, 11). Such findings are in the line with what reported in Baltagi and associates. Note that the results from homogeneous estimators are very closed to those obtained from the class of shrinkage/Bayesian estimators. However, by increasing the level of heterogeneity ($H = 0.9$) homogeneous estimators are outstaged by the shrinkage/Bayesian estimators, as columns 2 and 4 in Tables 1, 3, 5, 7, 9, 11 show. While the statistics from shrinkage/Bayesian estimators do not change very much with respect to the case of low heterogeneity, we note massive changes affecting homogeneous estimators, which is particularly evident in Tables 1 and 3 where Theil's U becomes on average three times bigger than in the case where $H = 0.1$. It is important to emphasize that this result reinforces and complements the empirical findings reported in the papers by Baltagi and Griffin (1997), Baltagi, Griffin and Xiong (2000), Baltagi, Bresson, Pirotte (2002), Baltagi, Bresson, Griffin and Pirotte (2003), and Brucker and Silivertovs (2006), where homogeneous estimators seem to dominate over heterogeneous and shrinkage. Low/mild levels of heterogeneity, as found in the papers cited above, entail the superiority of homogeneous estimators over heterogeneous and shrinkage. This occurs especially for small T, whereby the performance of individual estimators is very poor, as one could expect and as found in the empirical papers cited above. On the contrary, increasing the degree of heterogeneity improves the performance of shrinkage estimators over homogeneous estimators. This occurs for all values of T, even when heterogeneous estimators do not perform well; see also columns 2 and 4 Tables 1 and 3 with $T = 5$, where, in spite of high heterogeneity, homogeneous estimators have a better performance than their homogeneous counterparts. This is consistent with the analysis by Maddala, Li and Srivastava (1994), who advocate the use of shrinkage estimators to both take heterogeneity into account and smooth away the bias of heterogeneous estimators. It is worth noting that even under high levels of heterogeneity, heterogeneous estimators never outperform the shrinkage ones when limited cross dependence is present, whose predictive ability actually improves ceteris paribus as heterogeneity increases. The results change when cross dependence increases to high levels.

Other findings that emerge from our simulations are reported below.

The impact of cross sectional dependence is also quite substantial. In the case of mild cross dependence our findings are very much in line with what reported in the existing applied literature for the case of mild heterogeneity - see Baltagi and Griffin (1997), Baltagi, Griffin and Xiong (2000), Baltagi, Bresson, Pirotte (2002), Baltagi, Bresson, Griffin and Pirotte (2003), and Brucker and Silivertovs (2006). When instead we consider the case of large contemporaneous correlation, the statistics change dramatically, as one may see from Tables 2, 4, 6, 8, 10 and 12. Irrespective of the level of heterogeneity considered and other characteristics of the panel (namely the combination of size of T and N and the dynamics of the error terms), the estimators that show the best forecasting accuracy are always the shrinkage/Bayesian ones, with the only exception of very short panels with $T = 5$ where individual estimates are unreliable (this clearly emerges from Tables 2 and 4). It is worth noticing that the presence of cross sectional dependence has an impact on the shrinkage/Bayesian estimators in the sense that the statistics get worse as cross dependence gets larger, and this seems to suggest that an increasing presence of cross dependence makes forecasting in general more difficult in this case.

The time and cross sectional dimensions of panels do not play a substantial role. Few estimators are sensitive to T and N, especially when in the small sample case. In general most estimators are not sensitive to the values of the pair (T, N), and this is particularly evident in the case of the Hierarchical Bayes estimator, whose prediction outcome is almost invariant with T; this emerges from the comparison of the results in Tables 1 and 3 with those in tables 5, 7, 9 and 11, where the value of Theil's U is essentially the same for shrinkage estimators. This result confirms previous findings reported in Hsiao, Pesaran and Tahmiscioglu (1999) about the very small bias of this estimator even with small samples.

The error term dynamics does have an impact on the choice of the best estimator when cross dependence is mild. However, first difference homogeneous estimators outperform all other estimators in presence of low heterogeneity and (ρ, ϑ) equal to 0.9, as it is evident from the third columns of Tables 1, 3, 5, 7, 9 and 11. Both the presence of high heterogeneity or high cross sectional dependence make the dynamics of the error term irrelevant and as in most other cases seen so far the shrinkage/Bayesian estimators dominate. We refer the reader to column 4 in Tables 1, 3, 5, 7, 9 and 11 to observe the effect of large heterogeneity in presence of highly persistent dynamics; columns 3 and 4 in Tables 2, 4, 6, 8, 10 and 12 show that when cross-sectional dependence is high, first difference estimators no longer deliver the best performance among homogeneous estimators.

A final note about a case where there is evidence of small sample problem, related to the time series dimension T and arising when we implement Individual OLS and 2SLS. For $T = 5$ (see Tables 1-4), the Theil's statistics is never lower than 10^4, and therefore their forecasting capability is totally implausible. This also affects the performance of shrinkage estimators, whose magnitude of the Theil's statistics is much larger (at least of a factor 10^2) than that of the best estimators. Thus, for the case of a short panel ($T = 5$ in our case), our results contradict the findings reported in Maddala, Li and Srivastava (1994).

4.2 Diebold and Mariano's (1995) test

The outcome of Diebold and Mariano test is represented by a lower triangular matrix of dimensions (28×28) for each experiments. Since the amount of output generated by this part of the exercise exceeds a reasonable number of pages, we decide not to report it. They are available upon request from authors.

The main results reinforce the conclusions reported in the previous section for the Theil's U statistic.

When the degree of heterogeneity is small and there is mild cross sectional dependence, there is no evidence of statistically significant difference between shrinkage and homogeneous estimators, and therefore either class of estimators can be used irrespective of any feature of the data; this is evident from column 1 in Tables 1, 3, 5, 7, 9, and 11. On the other hand, when H is large, shrinkage/Bayesian estimators have a significantly better performance, especially for the small T case, and therefore the conclusion that they should be preferred follows from columns 2 and 4 in the Tables referred above. Only when the error component dynamics is characterized by a nearly integrated behavior the performance of homogeneous first difference estimators is significantly better than that of shrinkage estimators based on the model in levels (column 3 in the Tables).

When cross sectional dependence is large, for T larger than 5 there is no significant difference in the performance of homogeneous versus shrinkage/Bayesian estimators - see Tables 6, 8, 10 and 12. However, once again in the nearly integrated case the first difference homogeneous estimators dominate (see column 3 of Tables). When $T = 5$, Tables 2 and 4 show that even though none of the estimators has a significantly better performance than the others, however there is statistical evidence that Hierarchical Bayes is more powerful as heterogeneity increases. When T increases, the difference between homogeneous and heterogeneous estimators gets significant, and the latter group performs better, especially when heterogeneity gets bigger (this is particularly evident in column 2 in Tables 8,10 and 12). In presence of low heterogeneity, there is virtually no difference between estimators, Hierarchical Bayes included. Such finding illustrates that as long as heterogeneity is limited across units the choice of estimators is not crucial for forecasting. This is true especially when $T \geq 10$ (Tables 6, 8, 10 and 12). It is worth noticing that the presence of serial correlation in the error term doesn't affect these findings. The main findings so far are reinforced when the number of units is large (i.e. $N = 50$, see Tables 8, 10 and 12). Here too the presence of heterogeneity is crucial in marking the difference between pooled and heterogeneous estimators, in favor of the latter.

The impact of the error dynamics here follows a similar pattern as in the case of Theil's U statistics.

4.3 Pesaran and Timmermann's (1992) test

In this section, we describe the results of our Monte Carlo for the Pesaran and Timmermann's (1992) statistic which measures the capability to forecast turning points, reported at the bottom of Tables 1-12. The full set of results are reported in Trapani and Urga (2005), Tables B1-B12.

Since Pesaran and Timmermann's test is asymptotically distributed as a standard normal under the null hypothesis of no capability to detect turning points, the values in Tables can be interpreted either as raw numbers to rank estimators (the larger the value of the statistics, the higher the turning points detection capability), or we may compare them with quantiles of the normal distribution to test whether each estimator predicting capability is significant or not.

The main findings can be summarized as follows.

The impact of heterogeneity on the capability of forecasting turning points produces similar results as in Theil's U statistic case under mild cross sectional dependence; moreover, as shown in Tables 1, 3, 5, 7, 9 and 11, there exists always an estimator which has a statistically significant capability of detecting turning points. Low heterogeneity leads to the choice of homogeneous estimators, whilst with high levels of heterogeneity (like it is the case when considering Theil's U statistics) the capability of homogeneous estimators to detect turning points worsens. Conversely, shrinkage/Bayesian estimators become the best ones as heterogeneity increases, always outperforming heterogeneous estimators. It is worth noticing the following interesting regularity: in presence of large heterogeneity there is an improvement in predicting turning point, as can be seen from the higher values attained by Pesaran and Timmermann statistic.. The presence of heterogeneity always improves the predictive ability of heterogeneous and shrinkage/Bayesian estimators, with the latter being always the best when heterogeneity is high (the statistic is always different from zero, the only exception being the case of large cross sectional dependence). This pattern changes when the amount of contemporaneous dependence across units increases, and it makes homogeneous estimators less capable to forecast turning points even in the presence of near homogeneity ($H = 0.1$); Tables 4, 6, 8, 10 and 12 do not contain any case of estimators capable to predict turning points.

Other results as far as Pesaran and Timmermann's statistics is concerned are reported below.

As far as the impact of *cross sectional dependence* is concerned, as already pointed out in the previous point, in presence of mild levels of cross dependence it is always possible to find an estimator whose turning point prediction ability is statistically significant, but when we have large cross sectional dependence it is virtually impossible to find an estimator capable of predicting turning points, with a few exceptions in the class of Bayesian estimators.

The time series size T has an impact on the Pesaran and Timmermann's statistic, which has greater predictive performance when T increases. This does not apply when we evaluate to the cross sectional dimension N of the panel. For instance, when $T = 5$ and cross dependence is small, it is still possible to find estimators that are significantly capable of identifying turning points. In this case, an increase in N has the effect of improving the forecasting performance. Note that for $T = 5$ and $N = 20$, the predictive ability of Individual estimators is significant and very close to be the best among all estimators, albeit these estimates are computed for each unit with a degree of freedom equal to 2. This outcome is completely different with respect to the previous case, and it should lead to the conclusion that predictive performance measured with Theil's U statistics is different and unrelated with this aspect of forecasting performance.

The impact of the *error dynamics* has some commonalities with the Theil's U statistic case. Here too a nearly integrated error term results in having a better predictive performance on the side of first difference homogeneous estimators when heterogeneity is limited; in this case as well the presence of either heterogeneity or cross sectional correlation makes predictive performance worse. The presence of a nearly integrated dynamics makes homogeneous estimators based on the first differenced model the best, as shown by the third column in all Tables. However, their significance is heavily affected by the presence of cross dependence: the tests are significant when there is mild or no cross dependence but insignificant when the system exhibits a large degree of covariance among units.

4.4 The main features of our findings

In this final section, we summarize the main features of the various experiments commented above. Tables 1-12 report a summary of the three sets of statistics. Each of the tables is divided in three panels. The first one reports the best estimators according to Theil's U statistics. In the second panel, using the Diebold and Mariano (1995) test (DM), we report the comparison between the second best and the best estimator between the above estimators (DM1) and between the best estimator and the best Bayesian estimator (DM2), Finally, the last panel (PT) reports the best estimator according to the Pesaran and Timmerman (1992) statistic.

[Insert somewhere here Table 1-12]

The most important finding in this paper is the impact of heterogeneity on the predictive ability of alternative panel estimators, which reinforces and extends previous findings in the empirical literature - see e.g. Baltagi and Griffin (1997), Baltagi, Griffin and Xiong (2000), Baltagi, Bresson, Pirotte (2002, 2004), Baltagi, Bresson, Griffin and Pirotte (2003) and Brucker and Silivertovs (2006). The main features of the impact of heterogeneity are:

1. mild levels of heterogeneity lead to similar results as in Baltagi and Griffin (1997), Baltagi, Griffin and Xiong (2000), Baltagi, Bresson, Pirotte (2002), Baltagi, Bresson, Griffin and Pirotte (2003), and Brucker and Silivertovs (2006), whilst as heterogeneity increases shrinkage estimators have a better performance (and the forecasting capability of homogeneous estimators worsens), leading to similar results as in Baltagi, Bresson and Pirotte (2004);

2. particularly, increases in heterogeneity lead to both a worsening in the perfomance of homogeneous estimators and also an improvement in the accuracy of predictions based on shrinkage estimators, which reinforces the previous conclusion;

3. the same pattern occurs for other measures of forecasting accuracy as Pesaran and Timmermann's (1991) statistic to detect the capability of predicting turning points.

Other important findings are the following ones.

When cross sectional dependence is mild, the best class of estimators is the homogeneous one when heterogeneity H is limited or the shrinkage/Bayesian when heterogeneity is set to a large value. This regularity always takes place, irrespective of any other feature of the data.

When cross sectional dependence is large, the best estimators are almost always the shrinkage/Bayesian ones for T larger than 5. This does not hold when the error term exhibits a nearly integrated dynamics, as in such case estimates based on the first differenced data achieve the best performance. When $T = 5$ it should be pointed out the poor performance of both heterogeneous (which is likely to be due to the limited degree of freedom in each equation) and shrinkage/Bayesian estimators, mainly due to that in this case their prior is not designed to take account of the presence of contemporaneous correlation.

5 CONCLUDING REMARKS

In this paper, we compare the predictive performance of several homogeneous, heterogeneous and shrinkage/Bayesian estimators. We analyze the forecasting performance of 28 alternative estimators by varying the degree of heterogeneity and cross dependence in the panel, and by considering various combination of T and N, and using alternative specifications for error dynamics.

Our simulations show that the relative performance of the alternative estimators is affected by heterogeneity, being independent of the error term dynamics and of the time and cross sectional dimensions (N, T). The main conclusion is that heterogeneity greatly affects the performance of the various estimators. Whilst mild levels of heterogeneity entail the superiority of homogeneous estimators, as heterogeneity increases the shrinkage/Bayesian estimators and the Hierarchical Bayes estimator in general show a better forecasting performance across all experiments, regardless of sample size (T, N) and error dynamics. Another important finding is that the degree of cross dependence leaves the ranking of estimators virtually unchanged, even though the predictive performances of different estimators become more similar as cross dependence increases.

Our findings provide a clear guideline to practitioners when panel data are available for forecasting purposes. As a preliminary stage, individual estimators could be used and an indication as to the degree of heterogeneity could be derived e.g. considering the standard deviation of individual coefficients around their mean. Homogeneous estimators should be employed in presence of low and mild levels of heterogeneity, whilst when the degree of heterogeneity is high, shrinkage/Bayesian procedures deliver a superior forecasting performance.

Acknowledgements

We wish to thank the Editor, Michael Clements, an Associate Editor and an anonymous referee for their helpful comments which helped to greatly improve the paper. We also wish to thank participants in the 2004 Royal Economic Society Conference (Swansea, 7-9 April 2004), the 11th International Panel Data Conference (Texas A&M University, June 4-6, 2004), the ESRC Econometric Study Group Annual Conference (Bristol, 15-17 July, 2004) and seminars and workshops at Cass Business School, Bergamo University, European Central Bank for discussions and comments. Roy Batchelor, Valentina Corradi and Ron Smith provided very useful comments and suggestions which helped to improve the paper. However, the usual disclaimer applies. L. Trapani acknowledges financial support from Cass Business School under the RAE Development Fund scheme and ESRC Postdoctoral Fellowship Scheme (PTA-026-27-1107).

References

[1] Anderson, T.W., Hsiao, C. (1982). "Formulation and Estimation of Dynamic Models with an Error Component". *Journal of Econometrics*, *18*, 47-82.

[2] Arellano, M. (2003). *Panel Data Econometrics.* Oxford University Press, Oxford.

[3] Arellano, M., Honore', B. (2001). "Panel Data Models: Some Recent Developments", in Heckman, J.J. and E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, chapter 53, North-Holland.

[4] Arellano, M., Bond, S. (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". *Review of Economic Studies*, *58*, 277-97.

[5] Balestra, P., Vadharajan-Krishnakumar, J. (1987). "Full Information Estimation of a System of Simultaneous Equations with Error Component Structure". *Econometric Theory*, *3*, 223-246.

[6] Baltagi, B.H. (2005). *Econometric Analysis of Panel Data.* Wiley, 3rd edition.

[7] Baltagi, B.H., Bresson, G., Griffin, J.M., Pirotte, A. (2003). "Homogeneous, Heterogeneous or Shrinkage Estimators? Some Empirical Evidence from French Regional Gasoline Consumption". *Empirical Economics*, *28*, 795-811.

[8] Baltagi, B.H., 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*, 375-82.

[9] Baltagi, B.H., Bresson, G., Pirotte, A. (2004). "Tobin q: Forecast Performance for Hierarchical Bayes, Shrinkage, Heterogeneous and Homogeneous Panel Data Estimators". *Empirical Economics*, *29*, 107-113.

[10] Baltagi, B.H., Griffin, J.M. (1997). "Pooled Estimators vs. their Heterogeneous Counterparts in the Context of Dynamic Demand for Gasoline". *Journal of Econometrics*, *77*, 303-27.

[11] Baltagi, B.H., Griffin, J.M., Xiong, W. (2000). "To Pool or not to Pool: Homogeneous versus Heterogeneous Estimators Applied to Cigarette Demand". *The Review of Economics and Statistics*, *82(1)*, 117-26.

[12] Baltagi, B.H., Li, Q. (1992). "A Note on the Estimation of Simultaneous Equations with Error Components". *Econometric Theory*, *8*, 113-119.

[13] Breusch, T.S. (1987). "Maximum Likelihood Estimation of Random Effects Models". *Journal of Econometrics*, 36, 383-389.

[14] Brucker, H, Siliverstovs, B. (2006). "On the Estimation and Forecasting of International Migration: How Relevant is Heterogeneity across Countries?". *Empirical Economics*, *31*, 735-754.

[15] Clements, M.P., Hendry, D. (1993). "On the Limitations of Comparing Mean Square Forecast Errors (with Discussion)". *Journal of Forecasting*, *12*, 617-76.

[16] Christoffersen, P.F., Diebold, F.X. (1996). "Further Results on Forecasting and Model Selection under Asymmetric Loss". *Journal of Applied Econometrics*, *11*, 561-71.

[17] Diebold, F.X., Mariano, R. (1995). "Comparing Predictive Accuracy". *Journal of Business Economics and Statistics*, *13*, 253-63.

[18] Diebold, F.X., Lopez, J.A. (1996). "Forecast Evaluation and Combination", in G.S. Maddala and C.R. Rao (eds.), *Handbook of Statistics*, Amsterdam, North-Holland, 241-268.

[19] Driver, C., Urga, G. (2004). "Transforming Qualitative Survey Data: Performance Comparison for the UK". *Oxford Bulletin of Economics and Statistics*, *66*, 71-89.

[20] Granger, C.W.J., Pesaran, M.H. (2000a). "A Decision-based Approach to Forecast Evaluation", in W.S. Chan, W.K. Li, H. Tong (eds.), *Statistics and Finance: an Interface*, Imperial College Press, London..

[21] Granger, C.W.J., Pesaran, M.H. (2000b). "Economic and Statistical Measures of Forecast Accuracy". *Journal of Forecasting*, *19*, 537-60.

[22] Harvey, D., Leybourne, S., Newbold, P. (1997). "Testing the Equality of Prediction Mean Squared Errors". *International Journal of Forecasting*, *13*, 281-291.

[23] Hsiao, C. (2003). *Analysis of Panel Data.* Cambridge University Press, Cambridge, 2nd Edition,.

[24] Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K. (1999). "Bayes Estimation of Short-Run Coefficient in Dynamic Panel Data", in Hsiao, C., Lahiri, K., Lee, L.-F. and Pesaran, M.H. (eds.), *Analysis of Panels and Limited Dependent Variable Models*, Cambridge University Press, Cambridge.

[25] Keane, M.P., Runkle, D.E. (1992). "On the Estimation of Panel Data Models with Serial Correlation when Instruments are not Strictly Exogenous". *Journal of Business and Economic Statistics*, *10*, 1-9.

[26] Leitch, G., Tanner, J.E. (1991). "Economic Forecast Evaluation: Profits versus the Conventional Error Measures". *American Economic Review*, *81(3)*, 580-90.

[27] Maddala, G.S., Li, H., Srivastava, V.K. (1994). "A Comparative Study of Different Shrinkage Estimators for Panel Data Models". *Annals of Economics and Finance*, *2*, 1-30.

[28] Mariano, R. (2002). "Testing Forecasting Accuracy", in M.P. Clements and D.F. Hendry (eds.), *A Companion to Economic Forecasting*, Blackwell, 284-98.

[29] Nickell, S. (1981). "Biases in Dynamic Models with Fixed Effects". *Econometrica*, *16*, 1-32.

[30] Pesaran, M.H. (2007). "A Simple Panel Unit Root Test in the Presence of Cross Section Dependence". *Journal of Applied Econometrics*, *22(2)*, 265-312.

[31] Pesaran, M.H., Hsiao, C., Tahmiscioglu, A.K. (1999). "Bayes Estimation of Short-Run Coefficient in Dynamic Panel Data", in Hsiao, C., Lahiri, K., Lee, L.-F. and Pesaran, M.H. (eds.), *Analysis of Panels and Limited Dependent Variable Models*, Cambridge University Press, Cambridge.

[32] Pesaran, M.H., Smith, R. (1995). "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels". *Journal of Econometrics*, *68*, 79-113.

[33] Pesaran, M.H., Skouras, S. (2002). "Decision-Based Methods for Forecast Evaluation", in M.P. Clements and D.F. Hendry (eds.), *A Companion to Economic Forecasting*, Blackwell, 241-67.

[34] Pesaran, M.H., Timmermann, A. (1992). "A Simple non Parametric Test of Predictive Performance". *Journal of Business and Economic Statistics* *10*, 461-65.

[35] Sevestre, P., Trognon, A. (1985). "A Note on Autoregressive Error Component Models". *Journal of Econometrics*, *28*, 231-245.

[36] Swamy, P.A.V.B. (1970). "Efficient Inference in a Random Coefficient Regression Model". *Econometrica*, *38*, 311-23.

[37] Trapani, L., Urga, G. (2005). "Assessing Predictive Performance of Homogenous, Heterogeneous and Shrinkage Estimators for Heterogeneous Panels: A Monte Carlo Study". CEA-WP-01-2005, Centre for Econometric Analysis, Cass Business School, London.

[38] Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, Massachusetts.

Table 1: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(5,10)	(5,10)	(5,10)	(5,10)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	OLS **0.4276**	FDGMM 1.4166	FD-2SLS **0.3257**	GMM 1.4278
	Heterogeneous	Swamy 0.4319	Average OLS 1.7823	Average OLS 0.4520	Average OLS 1.5321
	Shrinkage	It. Bayes 0.4853	It. Bayes **0.5307**	It. Bayes 0.3895	It. Bayes **0.4832**
DM	DM1	OLS vs Swamy (0.3190)	It. Bayes vs FDGMM (-2.1966)(**)	FD-2SLS vs It. Bayes (2.0660)(**)	It. Bayes GMM (-2.3024)(**)
	DM2	It. Bayes vs. OLS (0.8880)			
PT	Homogeneous	FDGMM **1.6805(*)**	Within 1.7426(*)	FD-2SLS-KR **2.8158(**)**	Within 1.8785(*)
	Heterogeneous	Average OLS 1.3451	Ind. OLS 2.2208(**)	Ind. OLS 2.0630(**)	Ind. OLS 2.7630(**)
	Shrinkage	Bayes OLS 1.4275	It. Bayes **2.3312(**)**	I.B. OLS 2.1101(**)	I.B. OLS **2.7696(**)**

Note: This Table reports results for the case $(N, T) = (5, 10)$ under "mild" cross dependence (the support of ζ_i is [0,0.2]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0, 0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 2: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(5,10)	(5,10)	(5,10)	(5,10)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	MLE **<u>0.7618</u>**	FDGMM **<u>1.1896</u>**	Within-2SLS **<u>0.8000</u>**	FDGMM **<u>1.2822</u>**
	Heterogeneous	Swamy 0.7646	Average OLS 2.2297	Swamy 0.8303	Average OLS 1.6972
	Shrinkage	It. Bayes 134.84	It. Bayes 61.45	It. Bayes 53.46	It. Bayes 13.27
DM	DM1	Swamy vs. MLE (0.1387)	Average OLS vs. FDGMM (4.1188)(**)	Swamy vs. Within-2SLS (-0.0170)	Average OLS vs. FDGMM (4.1035)(**)
	DM2	It. Bayes vs. MLE (0.8286)	It. Bayes vs. FDGMM (-1.2647)	It. Bayes vs. Within-2SLS (0.0293)	It. Bayes vs. FDGMM (-1.0731)
PT	Homogeneous	FDGMM 0.9682	Within 1.4618	FDGMM 1.1872	Within 1.5026
	Heterogeneous	Ind. OLS 0.5177	Ind. OLS 1.5606	Ind. OLS 0.6647	Ind. OLS **<u>1.7540(*)</u>**
	Shrinkage	Bayes OLS 0.5809	It. Bayes **<u>1.6961(*)</u>**	I. B. OLS 0.7912	It. Bayes **<u>1.8367(*)</u>**

Note: This Table reports results for the case $(N, T) = (5, 10)$ under "large" cross dependence (the support of ζ_i is [-1,3]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0, 0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 3: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(5,20)	(5,20)	(5,20)	(5,20)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	OLS **0.4405**	Between 1.2063	FD-2SLS **0.2923**	Between 1.3196
	Heterogeneous	Swamy 0.4427	Swamy 1.4297	Average OLS 0.4628	Average OLS 1.2699
	Shrinkage	It. Bayes 0.4752	It. Bayes **0.4988**	It. Bayes 0.4037	It. Bayes **0.4899**
DM	DM1	Swamy vs. OLS (0.2034)	Between vs. It. Bayes (-2.2382)(**)	It. Bayes vs. FD-2SLS (2.4040)(*)	Average OLS vs. It. Bayes (1.1141)
	DM2	It. Bayes vs. OLS (0.7101)			
PT	Homogeneous	WLS **1.8422(*)**	Between-2SLS 1.1574	FD-2SLS-KR **2.8153(**)**	Between-2SLS 1.2569
	Heterogeneous	Swamy 1.8025(*)	Ind. OLS 1.7207(*)	Average OLS 2.2450(**)	Ind. OLS 2.5424(**)
	Shrinkage	It. Bayes 1.7747(*)	It. Bayes **1.8769(*)**	It. Bayes 2.2987(**)	I. B. OLS **2.5547(**)**

Note: This Table reports results for the case $(N, T) = (5, 20)$ under "mild" cross dependence (the support of ζ_i is [0,0.2]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0, 0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 4: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(5,20)	(5,20)	(5,20)	(5,20)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	MLE **0,7763**	FDGMM **1.1415**	OLS **0.8136**	FDGMM **1.2668**
	Heterogeneous	Swamy **0.7763**	Swamy 1.3863	Swamy 0.8369	Swamy 1.4654
	Shrinkage	It. Bayes 4.2762	It. Bayes 6.3261	It. Bayes 121.25	It. Bayes 10.4824
DM	DM1	Swamy vs. MLE (0.1870)	Swamy vs. FDGMM (1.1881)	Swamy vs. OLS (0.5656)	Swamy vs. FDGMM (1.3239)
	DM2	It. Bayes vs. MLE (0.7609)	It. Bayes vs. FDGMM (-0.9985)	It. Bayes vs. OLS (0.2129)	It. Bayes vs. FDGMM (-0.7957)
PT	Homogeneous	FD-2SLS 0.9331	Between -2SLS 0.8203	FD-2SLS-KR 1.2184	Between-2SLS 0.8228
	Heterogeneous	Swamy 0.7311	Average OLS 0.3977	Average OLS 0.7358	Ind. OLS 1.2785
	Shrinkage	I. B. OLS 0.7217	It. Bayes 1.2584	I. B. OLS 0.8544	I. B. OLS 1.3933

Note: This Table reports results for the case $(N,T) = (5,20)$ under "large" cross dependence (the support of ζ_i is [-1,3]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0,0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 5: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(10,20)	(10,20)	(10,20)	(10,20)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	WLS **0.4367**	GMM 0.8725	FD-2SLS **0.2911**	EC2SLS-AR(1) 0.8408
	Heterogeneous	Average 2SLS 0.4393	Ind. OLS 0.4755	Ind. 2SLS 0.4495	Ind. OLS 0.4462
	Shrinkage	I. B. OLS 0.4592	It. Bayes **0.4342**	It. Bayes 0.4243	It. Bayes **0.3874**
DM	DM1	Average 2SLS vs. WLS (0.1913)	Ind. OLS vs. It. Bayes (0.3638)	It. Bayes vs. FD-2SLS (2.5319)(**)	Ind. OLS vs. It. Bayes (0.5813)
	DM2	I. B. OLS vs. WLS (0.1424)			
PT	Homogeneous	Within **1.8706(*)**	Between-2SLS 1.1768	FD-2SLS-KR **2.8141(**)**	FD-2SLS 1.993(**)
	Heterogeneous	Average OLS 1.8543(*)	Ind. OLS 1.9827(**)	Average 2SLS 2.2883(**)	Ind. 2SLS 2.6172(**)
	Shrinkage	I. B. OLS 1.8304(*)	It. Bayes **2.0116(**)**	It. Bayes 2.2921(**)	Bayes 2SLS **2.6198(**)**

Note: This Table reports results for the case $(N, T) = (10, 20)$ under "mild" cross dependence (the support of ζ_i is [0,0.2]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0, 0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 6: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(10,20)	(10,20)	(10,20)	(10,20)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	2SLS-KR 0.7216	GMM 1.0270	WLS 0.7561	GMM 1.0339
	Heterogeneous	Average OLS 0.7319	Ind. OLS 0.7430	Average 2SLS 0.7597	Ind. OLS 0.7086
	Shrinkage	I. B. 2SLS **<u>0.7165</u>**	Bayes OLS **<u>0.6358</u>**	I. B. 2SLS **<u>0.7281</u>**	Bayes OLS **<u>0.6294</u>**
DM	DM1	2SLS-KR vs. I. B. 2SLS (-0.1842)	Ind. OLS vs. Bayes OLS (-0.3241)	WLS vs. I. B. 2SLS (-0.5509)	Ind. OLS vs. Bayes OLS (-0.5827)
	DM2				
PT	Homogeneous	FDGMM 0.9797	Between-2SLS 0.8761	FD-2SLS-KR 1.2275	Between-2SLS 0.8505
	Heterogeneous	Average 2SLS 0.7796	Ind. OLS 1.3693	Ind. OLS 0.8809	Ind. OLS 1.5320
	Shrinkage	I. B. OLS 0.8872	I. B. 2SLS 1.4408	I. B. OLS 1.0679	Bayes OLS 1.5962

Note: This Table reports results for the case $(N, T) = (10, 20)$ under "large" cross dependence (the support of ζ_i is [-1,3]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0, 0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 7: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(10,50)	(10,50)	(10,50)	(10,50)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	OLS **0.4244**	Between 0.9130	FD-2SLS **0.2743**	Between 0.9504
	Heterogeneous	Average OLS 0.4249	Ind. OLS 0.4250	Ind. OLS 0.4441	Ind. 2SLS 0.5004
	Shrinkage	I. B. 2SLS 0.4263	I. B. 2SLS **0.3566**	It. Bayes 0.4128	I. B. OLS **0.4191**
DM	DM1	Average OLS vs. OLS (0.1919)	Ind. OLS vs. I. B. 2SLS (-0.3208)	It.Bayes vs. FD-2SLS (2.5529)(**)	Ind. 2SLS vs. I. B. OLS (-0.4254)
	DM2	I. B. 2SLS vs. OLS (0.1294)			
PT	Homogeneous	Within **2.0593(*)**	FD-2SLS-KR 1.8359(*)	FD-2SLS **2.8517(**)**	Within 1.5140
	Heterogeneous	Average 2SLS 2.0511(**)	Ind. OLS 2.7020(**)	Average 2SLS 2.4851(**)	Ind. OLS 2.2132(**)
	Shrinkage	I. B. OLS 2.0343(**)	I. B. 2SLS **2.7046(**)**	It. Bayes 2.3819(**)	Bayes OLS **2.2358(**)**

Note: This Table reports results for the case $(N,T) = (10, 50)$ under "mild" cross dependence (the support of ζ_i is [0,0.2]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0, 0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 8: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(10,50)	(10,50)	(10,50)	(10,50)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	2SLS-KR 0.7132	FD-2SLS-KR 1.0412	2SLS-KR 0.7443	FD-2SLS-KR 1.1162
	Heterogeneous	Average 2SLS 0.7260	Ind. OLS 0.7238	Average 2SLS 0.7509	Ind. OLS 0.6933
	Shrinkage	I. B. OLS **<u>0.7097</u>**	I. B. 2SLS **<u>0.7190</u>**	I. B. 2SLS **<u>0.7190</u>**	I. B. OLS **<u>0.6143</u>**
DM	DM1	I. B. OLS vs. 2SLS-KR (-0.1522)	Ind. OLS vs. I. B. 2SLS (-0.3148)	2SLS-KR vs. I. B. 2SLS (-0.4638)	Ind. OLS vs. I. B. OLS (-0.3885)
	DM2				
PT	Homogeneous	2SLS-KR 1.0630	Within 1.3108	FD-2SLS-KR 1.2919	Within 1.3212
	Heterogeneous	Average OLS 1.0218	Ind. OLS 1.4738	Average OLS 1.0929	Ind. OLS 1.6247
	Shrinkage	I. B. OLS 1.0756	Bayes OLS 1.5318	I. B. 2SLS 1.2143	Bayes OLS **<u>1.6810(*)</u>**

Note: This Table reports results for the case $(N, T) = (10, 50)$ under "large" cross dependence (the support of ζ_i is [-1,3]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0, 0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 9: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(20,50)	(20,50)	(20,50)	(20,50)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	WLS **0.4243**	EC2SLS-AR(1) 0.9244	FD-2SLS **0.2744**	EC2SLS-AR(1) 0.8767
	Heterogeneous	Average OLS 0.4256	Ind.OLS 0.3826	Average 2SLS 0.4458	Ind. 2SLS 0.3399
	Shrinkage	I.B. 2SLS 0.4254	I.B. OLS **0.3786**	I.B. 2SLS 0.4374	I.B. 2SLS **0.3323**
DM	DM1	I.B. 2SLS vs. WLS (0.1666)	I.B. OLS vs. Ind. OLS (-0.1734)	I.B. 2SLS vs. FD-2SLS (2.7041)(**)	Ind. 2SLS vs. I.B.2SLS (-0.3480)
	DM2				
PT	Homogeneous	Within **2.0342(**)**	Within 1.5275	FD-2SLS **2.8517(**)**	FD-2SLS-KR 1.8432(*)
	Heterogeneous	Average OLS 2.0290(**)	Ind. OLS 2.3158(**)	Average 2SLS 2.4560(**)	Ind. 2SLS 2.6993(**)
	Shrinkage	Bayes OLS 2.0191(**)	Bayes OLS **2.3199(**)**	I.B. 2SLS 2.3377(**)	I.B. 2SLS **2.7002(**)**

Note: This Table reports results for the case $(N,T) = (20,50)$ under "mild" cross dependence (the support of ζ_i is [0,0.2]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0,0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 10: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(20,50)	(20,50)	(20,50)	(20,50)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	2SLS 0.7055	FD-2SLS-KR 1.0049	WLS 0.7203	GMM 1.0140
	Heterogeneous	Average 2SLS 0.7050	Ind. OLS 0.5927	Average 2SLS 0.7282	Ind. OLS 0.5899
	Shrinkage	I.B. 2SLS **<u>0.6970</u>**	Bayes OLS **<u>0.5806</u>**	I.B. 2SLS **<u>0.7132</u>**	Bayes OLS **<u>0.5779</u>**
DM	DM1	Average 2SLS vs. I.B. 2SLS (-0.7593)	Ind. OLS vs. Bayes OLS (-0.4589)	WLS vs. I.B. 2SLS (-0.2570)	Ind. OLS vs. Bayes OLS (-0.3689)
	DM2				
PT	Homogeneous	2SLS-KR 1.0785	Within 1.3531	FD-2SLS-KR 1.2978	FD-2LSLS-KR 1.2116
	Heterogeneous	Average 2SLS 1.0502	Ind. 2SLS 1.6131	Average OLS 1.1505	Ind.OLS **<u>1.7262(**)</u>**
	Shrinkage	I.B. OLS 1.0797	Bayes OLS 1.6308	I.B. 2SLS 1.1829	Bayes OLS **<u>1.7403(**)</u>**

Note: This Table reports results for the case $(N, T) = (20, 50)$ under "mild" cross dependence (the support of ζ_i is [-1,3]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0, 0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 11: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(50,50)	(50,50)	(50,50)	(50,50)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	WLS **0.4217**	EC2SLS-AR(1) 0.8658	FD-2SLS **0.2954**	EC2SLS-AR(1) 0.7992
	Heterogeneous	Average 2SLS 0.4246	Ind. OLS 0.3982	Ind. 2SLS 0.4551	Ind. 2SLS 0.3848
	Shrinkage	Bayes OLS 0.4225	I.B. OLS **0.3952**	Bayes 2SLS 0.4495	I.B. 2SLS **0.3817**
DM	DM1	Bayes OLS vs. WLS (0.0333)	Ind. OLS vs. I.B. OLS (-0.1316)	Bayes 2SLS vs. FD-2SLS (2.7075)(**)	Ind. 2SLS vs. I.B. 2SLS (-0.1940)
	DM2				
PT	Homogeneous	Between-2SLS 2.0165(**)	Within 1.2593	FD-2SLS **2.8223(**)**	Within 1.3702
	Heterogeneous	Average OLS **2.0217(**)**	Ind. OLS 2.1208(**)	Average 2SLS 2.4043(**)	Average OLS **2.0217(**)**
	Shrinkage	I.B. OLS 1.9996(**)	I.B. OLS **2.1461(**)**	I.B. Bayes 2SLS 2.3046(**)	I.B. OLS 1.9996(**)

Note: This Table reports results for the case $(N, T) = (50, 50)$ under "mild" cross dependence (the support of ζ_i is [0,0.2]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0, 0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.

Table 12: Forecasting accuracy measures: Theil's U, Diebold and Mariano (1995) and Pesaran and Timmermann (1992) for homogeneous, heterogeneous and shrinkage/Bayesian estimators.

(T,N)		(50,50)	(50,50)	(50,50)	(50,50)
(ρ, ϑ, H)		(0,0,0.1)	(0,0,0.9)	(0.9,0.9,0.1)	(0.9,0.9,0.9)
Theil's U	Homogeneous	2SLS 0.6992	EC2SLS 1.0227	WLS 0.7172	EC2SLS 0.9964
	Heterogeneous	Average 2SLS 0.6989	Ind. OLS 0.5929	Average 2SLS 0.7251	Ind. OLS 0.5997
	Shrinkage	I.B. 2SLS **0.6931**	Bayes OLS **0.5870**	I.B. 2SLS **0.7153**	Bayes 2SLS **0.5937**
DM	DM1	Average 2SLS vs. I.B. 2SLS (-0.7603)	Ind. OLS vs. Bayes OLS (-0.3006)	WLS vs. I.B. 2SLS (-0.074)	Ind. OLS vs. Bayes 2SLS (-0.2384)
	DM2				
PT	Homogeneous	2SLS 1.0319	Within 1.0604	FD-2SLS-KR 1.2413	Within 1.0262
	Heterogeneous	Average 2SLS 1.0137	Ind. OLS 1.4924	Average 2SLS 1.1154	Ind. OLS 1.6103
	Shrinkage	I.B. 2SLS 1.0529	Bayes OLS 1.5384	I.B. 2SLS 1.1665	Bayes OLS **1.6500(*)**

Note: This Table reports results for the case $(N, T) = (50, 50)$ under "large" cross dependence (the support of ζ_i is [-1,3]), two different degrees of heterogeneity (low, with $H = 0.1$ and high with $H = 0.9$) and two different specifications for the error term dynamics (the white noise case with $(\rho, \vartheta) = (0, 0)$ and a nearly integrated one where $(\rho, \vartheta) = (0.9, 0.9)$). Forecasting horizon $h = 10$ periods ahead.