Assessing the Predictive Performance of Homogeneous, Heterogeneous and Shrinkage Estimators for Heterogeneous Panels: a Monte Carlo Study

Lorenzo Trapani*
Cass Business School and Bergamo University

Giovanni Urga
Cass Business School

22 January 2004
Very preliminary version - Please do not cite

* *Corresponding author*: Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ Tel.+/44/(0)20/70408089; Fax.+/44/(0)20/70408885; e-mail: L.Trapani@city.ac.uk. The usual disclaimer applies.

Abstract

This paper reports the results of a series of Monte Carlo exercises to contrast the forecasting performance of several panel data estimators. The comparison is done using different levels of heterogeneity, alternative panel structure in terms of T and N and using various error dynamics specification. We implement 28 estimators divided into three main groups: homogeneous, heterogeneous and shrinkage estimators. To assess the predictive performance, we use traditional measures of forecast accuracy (such as Theil's U statistics, RMSE and MAE) and we apply the Diebold and Mariano's (1995) test to check whether performances are significantly different. We also base our analysis on the capability of forecasting turning points, comparing the results from Pesaran and Timmerman's (1992) statistics. We find that homogeneous estimators perform well when heterogeneity is limited, while the shrinkage and Bayesian procedures have very good predictive power independently of the model's features.

J.E.L. Classification Numbers: C23.
Keywords: Panel data; Forecasting performance; Pooled estimators; Homogeneous, heterogeneous and shrinkage estimators; Monte Carlo.

1 INTRODUCTION

The issue of estimating the parameters of heterogeneous panels has been paid considerable attention by both the applied and the theoretical literature in the latest years, and a plethora of estimation techniques has been derived. Baltagi (2001), Arellano and Honore' (2001) and Arellano (2003) provide comprehensive surveys on the topic, and Baltagi and Griffin (1997) discuss the rationale behind the various estimation techniques. These estimation techniques differ very much between each other, not only as far as their computation is concerned, but also with respect to the assumptions they make on the model. It has then become customary to group them into three main classes: homogeneous, heterogeneous and shrinkage (or Bayesian) estimators. While the first group hypothesizes poolability of the data in the panel, and therefore parameters homogeneity across the panel units, the second one denies the validity of this assumption taking account explicitly of the presence of heterogeneity among units. The class of Bayesian estimators can be regarded as a hybrid solution between the two other classes (see Maddala, Li and Srivatsava, 1994, and Hsiao, Pesaran and Tahmiscioglu, 1999). It becomes then crucial to understand which estimation methodology is the "best", in statistical terms, for the specific research interest (e.g. bias reduction, efficiency, forecast accuracy...). Some previous literature elaborated on this point. Hsiao *et al.* (1999) analysed the bias of various estimators, obtainig the classical result that the Hyerarchical Bayes one achieves the best performance, and Cornwell and Rupert (1988) compared the efficiency of several IV estimators with an empirical exercise.

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 and Pirotte (2002) and Baltagi, Bresson, Griffin and Pirotte (2002) apply dynamic panel specifications to industrial level data and find that the predictive ability of homogeneous estimators outperforms that of heterogeneous and Bayesian estimators over any forecast horizon. GLS and within-2SLS, particularly, emerge as the estimators having the best out-of-sample performances, especially with respect to long forecasting horizons. Such superiority of homogeneous estimators may sound quite reasonable when the panel is short, and when the degree of heterogeneity across units is limited, but is rather puzzling when the time length of the panel T is large or when the degree of heterogeneity is high. This genuine empirical finding is particularly interesting especially in the light of the fact that the hypothesis of homogeneity and poolability is rejected by the data considered in the papers cited above. A first possible interpretation of this apparently contra-intuitive empirical regularity is that a model that is "simple and parsimonious" offers a better forecasting performance than one whose specification is richer but with a limited degree of freedom.

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 is sensitively dependent 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 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 couple (N, T), the degree of heterogeneity, 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. To do this, we develop a Monte Carlo exercise, running it for each possible scenario.

A further question that can arise in the light of some recent developments of the literature is how to assess predictive performance. In their contributions, Baltagi and associates use the RMSE as the measure of forecasting accuracy. However, the literature has developed a quite critical attitude towards classical statistical indicators, even though it allows for widely used results concerning testing - see the review in Mariano (2002). From the statistical point of view, it has been pointed out (Clements and Hendry, 1993) 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 summarised by only two moments. The literature has criticised RMSE also on the basis of economic considerations, arguing that predictive performance should be evaluated via the losses that arise from forecasts (and specifically their errors) when employed to make decisions - see Leitch and Tanner (1991). Hence, the economic and econometric literatures have approached the issue under a decision theoretic framework, where the best forecast is the one that minimizes losses - see Granger and Pesaran (2000a, 2000b), and the review by Pesaran and Skouras (2002). To briefly illustrate this theory, let e_{it} be the error made using forecast method i at time t, and $g(e_{it})$ the loss associated with this error. A forecasting method i will be superior to another one, say j, if

$$g(e_{it}) \leq g(e_{jt})$$

for all ts, the inequality holding for at least one t. The crucial point within this framework is choosing the loss function $g(\cdot)$ is the crucial point; this issue has no general answer since it depends on the decision problem one is dealing with. A well-known and widely used specification for $g(\cdot)$ is the quadratic one, i.e. $g(e_{it}) = e_{it}^2$, and Pesaran and Skouras (2002) show that this loss function is proportional to the RMSE. However, other specifications can be considered, and have been employed by the literature - see the review by Pesaran and Skouras (2002) for details, and the discussions by Christoffersen and Diebold (1996) and Mariano (2002).

The use of RMSE can therefore be affected by two problems: the inadequacy of its specification in terms of loss function, and its being a raw number, usuitable for statistical discrimination between forecasting performances. Both problems can be overcome employing Diebold and Mariano's (1995) test. This - with the adjustment for small sample bias proposed by Harvey, Leybourne and Newbold (1997) - is still based on a loss function without assuming a specific functional form for it, and it allows to choose between predictors. In this light, this may be viewed as a second approach to measuring forecasting performance.

It is also worth noticing that forecasting performance could be referred to something different than minimising a loss function. Granger and Pesaran (2000b) argue that a possible measure for predictive performance could be based on the capability to capture the sign of changes in the series rather than its values. A non parametric statistics that evaluates the ability to forecast changes is due to Pesaran and Timmermann (1992), which is extensively discussed also in Granger and Pesaran (2000b). An application of the Pesaran and Timmerman statistics as a descriptive measure to be associated to each predictor when ranking forecasting techniques was considered by Driver and Urga (2003). In our paper, we will consider all these three approaches (statistical measures, testing and turning point predictive capability evaluation).

The remainder of this work is as follows. First, we set out the model we will be considering for our exercise, and briefly describe the estimation techniques and the predictive performance indicators that will be employed in our experiment (Section 2). Secondly, we describe the details of the Monte

Carlo experiment (Section 3). Last, we report and analyse the simulation outcomes (Section 4); conclusions follow (Section 5).

2 Model and estimation theory

The 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 $t = 1, .., T$. The error term u_{it} is assumed to have no time specific effects, and no cross sectional dependency - i.e. $E[u_{it}u_{js}] = 0$ for all $i \neq j$. Model (1) is the classical dynamic panel data specification, as discussed extensively in Baltagi (2001). 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 known without needing forecast it.

As far as estimation is concerned, we employed both homogeneous and heterogeneous estimators, performing an exercise similar to that in Baltagi, Bresson and Pirotte (2002), Baltagi and Griffin (1997), Baltagi, Bresson, Griffin and Pirotte (2002) and Baltagi, Griffin and Xiong (2000). It is worth emphasizing that whilst heterogeneous estimators are based on model (1) and therefore take account of the parameters heterogeneity across units, homogeneous estimators assume the poolability of the data, and are based on the following different specification of the DGP:

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

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

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

where μ_i is the unobservable individual specific effect and ν_{it} is the remainder of the disturbance - see Baltagi (2001) 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). Even with homogeneous estimators, we assume no time specific effects, our aim being focused on the effect of grouping across units. We choose this specification in the light of the applied literature we mean to generalize, where the two way error component model is seldom assumed - an exception being the contribution by Baltagi, Griffin and Xiong (2000).

After setting up the model, we turn our discussion to estimation, referring to Baltagi (2001) for the details of each estimator. We also sketch the measures of forecasting performance we employ in our simulation exercise.

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 regional effects; first difference OLS to wipe out the effect of (possible) serial correlation in the error term; Within estimator, which allows for unit specific effects; Between estimator; and WLS and WLS-AR(1), where unit specific effects are assumed to be random. After the results in Pesaran and Smith (1995) and the review in Baltagi (2001), it is known that none of this estimate is either unbiased or consistent. This is due to the assumption, common to all these estimators, that regressors are exogenous. But being the model dynamic, it is well known that even if 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 renders 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 (2001).

To achieve consistency, we focus on pooled estimators based on instrumental variables. 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 regional 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 (2001) - and EC2SLS-AR(1) - see Baltagi, Griffin and Xiong (2000) - to potentially achieve more efficiency by taking account of possible serial correlation in the

[1] Such estimation technique can be applied only if $N > T$ - see Baltagi (2001).

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. Finally, we used Arellano and Bond (1991) estimation procedure, using a GMM estimation method on the specification in differences (whose outcome will be labelled as FDGMM); also, we employ the same set of instruments in first difference on a specification in levels (GMM)[3].

Last, we considered MLE estimation on the ground of the discussion in Baltagi (2003), using the iterative procedure suggested by Breusch (1987). In total, we compare 18 homogeneous estimators.

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 experiment 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.

Also, we employed a class of shrinkage/Bayesian estimators - see Maddala, Li and Srivatsava (1994) - where each individual estimate is shrunk towards the pooled estimates by weighting it with weight depending on the corresponding covariance matrix. The authors claim that shrinkage type

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

[3]It is worth noticing that such GMM estimation procedures have existence conditions depending on the sizes of N, T and k (this latter being the number of parameters to be estimated) when the two-step GMM estimation is considered (see Baltagi 2001) - this existence condition is $N > T(k-2) + (T+3)/2$. These estimators wouldn't 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.

estimator are superior to either homogeneous or to 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. In total, we compare 10 different heterogeneous estimators.

2.3 Comparing forecasting performance

As mentioned in the introduction, 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. measures of the capability of predicting turning points;
3. measure of statistical assessment of performance.

The indicators we chose are, for each class:

1. MAE, RMSE and Theil's U statistics, whose expressions are respectively

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

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

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

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}. To obtain a single overall measure of performance, we considered the average of each indicator across units, similarly to Baltagi, Bresson and Pirotte (2002), Baltagi and Griffin (1997), Baltagi, Bresson, Griffin and Pirotte (2002) and Baltagi, Griffin and Xiong (2000). These indicators are all based on the residuals from forecast, and widely employed in the realm of forecasting. Particularly, the literature that has inspired our contribution bases its assessment of the estimators predictive ability on RMSE - see the applied works by Baltagi, Bresson and Pirotte (2002), Baltagi and Griffin (1997), Baltagi, Bresson, Griffin and Pirotte (2002) and Baltagi, Griffin and Xiong (2000), and also the works by Granger and Giacomini (2003) and Maddala, Li and Srivastava (1994). We report these three "classical" measures, even though the position we take is different from that of the cited literature. Indeed, most of our attention will be devoted to Theil's U statistics, given its nature of relative measure which doesn't have the scaling problem of both RMSE and MAE;

2. as a measure of the capability of predicting turning points, we employ Pesaran and Timmerman's (1992) statistics, defined as

$$PT_j = \frac{\hat{P}_j - \hat{P}_j^*}{\sqrt{\hat{V}\left(\hat{P}_j\right) - \hat{V}\left(\hat{P}_j^*\right)}}$$

where

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

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

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

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

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;

3. the statistics we use in this group is Diebold and Mariano's (1995) test. This statistics can be used for any h and doesn't require gaussianity, zero-mean, serial or contemporaneous incorrelation of the forecast errors. 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. We like to emphasize that under the null hypothesis of there being no difference between forecast performances, the Diebold and Mariano statistics is distributed as a standard normal. Even in this case, we compute the test statistics for every unit of the panel and then take the average across units.

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 experiment

We generate a sample of N units with length $T+T0$, where $T0$ 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 iteration 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:

$$\alpha_i = \bar{\alpha} + \alpha^H N_i^{\alpha},$$

$$\beta_i = \bar{\beta} + \beta^H N_i^{\beta},$$

$$\gamma_i = \bar{\gamma} + \gamma^H N_i^{\gamma},$$

[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 $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ are the mean values of the parameters, $N_i^{\cdot}$ denotes an independent (across i) extraction from a random variable and α^H, β^H and γ^H control the parameters heterogeneity across units, which will be useful throughout the simulation 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 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;

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

$$x_{it} = \alpha_i + \beta_i + \delta x_{it-1} + \eta_{it}, \tag{3}$$

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

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

  This correlation is such that $E(x_{it}\nu_{it}) = 0$ for any i - and hence x_{it} endogeneity is ruled out - and $E(x_{it}\mu_i) \neq 0$. This two results make x_{it} a strictly exogenous variable and a valid instrument for GMM estimation *a la* Arellano and Bond (1992) thanks to its correlation with the unit specific effect - see Baltagi (2001) for discussion.

We considered the following values for the parameters of our simulation exercise:

- 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 $(0.9, 0.9)$[5]. These two couples are aimed at taking into account the case of non autocorrelation and of near integration;
- the number of initial observations to be discarded was set equal to $T0 = 100$;
- the forecasting horizon is set equal to $h = 10$[6].

4 Simulation results

In this section we report and comment the results of the various Monte Carlo experiments. First, we present the predictive performance of the alternative estimators measured according to the Theil's U statistics[7]; second, we report Pesaran and Timmerman's (1992) statistics; third, we report some comments on the outcomes of Diebold and Mariano (1995) test.

4.1 Statistical measures of accuracy

The results we report here are contained in Tables 1.a)-1.e), which are provided in Appendix I. They refer to two different degrees of heterogeneity - obtained by assuming $\alpha^H = \beta^H = \gamma^H \equiv H = 0.1$ and $\alpha^H = \beta^H = \gamma^H \equiv H = 0.9$ respectively -, and two different specifications for dynamics - the couple (ρ, ϑ) was set equal to $(0, 0)$ and $(0.9, 0.9)$. The couples (T, N) we consider are: $(5, 10)$, $(5, 20)$, $(10, 20)$, $(10, 50)$, and $(20, 50)$.

The main results can be summarised as follows:

- the degree of heterogeneity has indeed a strong impact on the forecasting performance, and, as expected, this is particularly true for homogeneous estimators. For instance in Table 1.a), the performance of OLS estimator when $(N, T) = (10, 5)$ becomes four times as weak (the value of the statistics moves from around 0.4-0.5 to around 1.8-1.9)

[5] Further developments of this work will consider the following spectrum of values for (ρ, ϑ): $\{-0.9, -0.3, 0, 0.3, 0.9\} \times \{-0.9, -0.3, 0, 0.3, 0.9\}$

[6] We plan to extend this to the cases $h = 1$ and $h = 5$, as in Baltagi, Bresson and Pirotte (2002), Baltagi and Griffin (1997), Baltagi, Bresson, Griffin and Pirotte (2002) and Baltagi, Griffin and Xiong (2000)

[7] We only report results for the Theil's U since this statistics is not affected by scale problems. We also computed RMSEs and the MAEs for each simulation confirming the same findings. The results are available upon request.

when heterogeneity is included. Table 1.e) shows that this behavior is virtually unchanged when the couple (N, T) grows large;

- on the other hand, the Hierarchical Bayes estimator (last row in all Tables) has an almost constant prediction outcome, being thus almost independent of T (this result confirms previous findings reported in Hsiao, Pesaran and Tahmiscioglu, 1999), of the degree of heterogeneity (H) and value of autoregressive coeffcient ρ. Theil's U statistics is never larger than 0.7-0.8, and in most cases it is lower;

- the small sample problem arises when using Individual OLS and 2SLS, whose performance is instead comparable with that of the other estimators when the sample size is equal to 10 or higher. For $T = 5$ (see Tables 1.a)-1.b)), the Theil's statistics is never lower than 10^4. 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);

- first difference estimators show an average performance, which improves when the error dynamics is nearly integrated and heterogeneity remains low, as shown in the third column of Tables;

- when the size of T is close to that of N, the IV estimators based on Arellano and Bond's (1992) instruments performs very poorly, whether it is based on the model in level or on the first differenced one;

- on average, when T is small, homogeneous estimators prove to be better than heterogenous estimators, even in the presence of large heterogeneity, as reported in Table 1.a)-1.b) illustrating the poorness of Individual OLS and 2SLS estimates. These are unreliable for $T = 5$, irrespectively of other parameters in the simulation. OLS and WLS performance is often the best when there is no dynamics in the idiosyncratic component, as reported in the first column of Tables 1.a)-1.b). This is no longer valid when either the time size increases, or the error term has a nearly integrated dynamics;

- the forecasting performance of Pesaran and Smith's (1995) average estimators is quite good, despite that in the small sample case ($T = 5$) the individual estimates performe poorly. Tables 1.a)-1.b) show that the predictive performance of the heterogeneous estimators is comparable with that of OLS especially when the value of ρ and ϑ increases.

This result is consistent with the theory that insures the goodness of these estimates (in terms of bias reduction and consistency) when the number of units N increases. We notice that for $N = 10$ we get quite good outcomes;

- shrinkage estimators show good predictive ability for T larger than 5; in this case their predictive performance is very close to the one of Hierarchical Bayes estimator, as Tables 1.c)-1.e) show, and they sometimes outperform it[8].

4.2 Measures of capability to forecast turning points

In this section, we describe the results of our Monte Carlo for the Pesaran and Timmerman's (1992) statistics, reported in Table 2.a)-2.e). Since Pesaran and Timmerman's test is asymptotically distributed as a standard normal under the null hypothesis of no capability to detect turning points, the data in our Tables can be interpreted in two ways:

- as raw number to rank estimators (the larger the value of the statistics, the higher the turning points detection capability);
- comparing them with quantiles of the normal distribution to test whether each estimator predicting capability is significant or not.

The main results can be summarised as follows:

- heterogeneity has the same impact across experiments. Heterogeneity makes, *ceteris paribus*, the turning points predictive ability of homogeneous estimators lower than in the nearly homogeneous case ($H = 0.1$). The performance of homogeneous estimators becomes very poor when $H = 0.9$. This is the case of the OLS, whose capability of detecting turning points is shrunk towards zero whenever H is set equal to 0.9. It can be noticed that when the degree of heterogeneity is low, it is always an homogeneous estimator to achieve the largest value, thereby outperforming both Shrinkage and heterogeneous estimators. This happens independently of the dynamics, or the size of the couple (N, T);

[8] The improvement coming from considering the iterative version of shrinkage estimators is usually rather marginal, whilst the amount of calculations needed makes them much more cumbersome. It is also worth noticing that in our simulation exercises the computation of Hierarchical Bayes estimator was found to be much slower than the one of iterative shrinkage estimators.

- the performance of GMM based estimators shows an interesting pattern. Even though the instruments for these estimators are chosen with respect to a specification in levels for the model, the performance of estimates based on the specification in first differences is always better than the one based on specification (1). The difference between the two is found to be very large when heterogeneity is large;

- the predictive ability of Individual estimators is often significant and very close to be the best among all estimators if one ranks them on the ground of the statistics. This also happens for small T (i.e. in the case of $T = 5$), when these estimators are computed for each unit with a degree of freedom equal to 2. This should lead to conclude that predictive performance measured with Theil's U statistics (and with other statistical indicators such as RMSE and MAE, too) is different and unrelated with this aspect of forecasting performance. When dynamics is close to the case of integration, such estimates show a constant improvement, which is found to be not sensitive to other parameters;

- the presence of heterogeneity always improves the predictive ability of heterogeneous and shrinkage estimators. These latter ones particularly are always the best when heterogeneity is high, and their performance is always significantly different from zero;

- the presence of a nearly integrated dynamics makes homogeneous estimators based on the first differenced model the best, as shown by the third column of all Tables.

4.3 Diebold and Mariano's (1995) test

The outcome of Diebold and Mariano's test is represented by a lower triangular matrix of dimensions 28×28. Since the amount of output generated by this part of the exercise exceeds a reasonable number of pages, we decide not to report it.[9]

The main results can be summarised as follows:

- when $T = 5$ no one of the estimators has a significantly better performance than the others. However, there is statistical evidence that Hierarchical Bayes is marginally better when heterogeneity increases;

[9] All results on Diebold and Mariano's test are available upon request.

- when T increases, the difference between homogeneous and heterogeneous estimators becomes significant, the latter group performing better than the former. This holds especially, as already seen, when heterogeneity increases. When we have a small amount of heterogeneity there is virtually no difference across estimators, Hierarchical Bayes included. Such finding illustrates that as long as heterogeneity is limited across units the choice of estimators is not crucial for the forecasting performance. This holds for $T = 10$ or greater. Finally, the presence of dynamics in the error term doesn't affect these findings;

- these conclusions are reinforced when the number of units is large (i.e. $N = 50$). Here too the presence of heterogeneity is crucial in marking the difference between pooled and heterogeneous estimators. Once again the latter perform better than the former;

- the gain from considering an iterative shrinkage estimator rather than a non iterative one has been assessed as poor when interpreting the results concerning Theil's U statistics; Diebold and Mariano test reinforces this conlusion showing that it is not significant.

5 Conclusions

In this paper, we compare the predictive performance of several homogeneous, heterogeneous and shrinkage estimators applied to a heterogeneous model. We analyze the forecasting performance of the various estimators by varying the degree of heterogeneity in the panel and using alternative specifications for error dynamics.

Our main results are that for short panels with a limited degree of heterogeneity, homogeneous estimators are preferable to the heterogeneous ones. This confirms the findings of Baltagi, Bresson and Pirotte (2002), Baltagi and Griffin (1997), Baltagi, Bresson, Griffin and Pirotte (2002) and Baltagi, Griffin and Xiong (2000). This outcome is not affected by the the dynamics in the error term specification. For the case of near integration, the homogeneous first difference estimators performs best. Homogeneous estimators perform also well when T increases and heterogeneity is small. On the other hand, the performance of heterogeneous estimators show some sign of improvements with respect to the caes of small T mainly due to the higher degree of freedom in the individual regressions. The good performance of the homogenous estimators is better than that of the Hierarchical Bayes estimator and of shrinkage estimators, though the Hierarchical Bayes has in general a better performance across all experiments, regardless of heterogeneity or error dynamics.

Heterogeneity greatly affects performance of the various estimators. Homogeneous estimators show a very poor predictive performance, and in panels with long T heterogeneous estimators are preferrable. Anyway, under the presence of heterogeneity the shrinkage estimators and the Hierarchical Bayes estimator show the best performance. This conclusion is consistent with Hsiao, Pesaran and Tahmiscioglu (1999) analysis, that shows how Bayesian estimation gives the best results in terms of bias reduction. Diebold and Mariano's test shows that shrinkage estimators performance is also *significantly* better than the one of the other estimators.

If the performance to be looked at is the capability of detecting turning points, results are quite similar, even if the performance of Individual OLS and 2SLS is always good across all possible cases. Finally, homogeneous estimators sistematically fail to predict turning points when heterogeneity is high.

Our findings provide some guidelines when we use panel data to forecast. When we have panels with reasonably long T ($T = 10$ or larger) and heterogeneity, the Hierarchical Bayes estimator is the best technique as far as both traditional statistical measures of predicitve accuracy and capability of detecting turning points are concerned.

References

[1] Anderson, T.W., Hsiao, C. (1982), "Formulation and Estimation of Dynamic Models with an Error Component", *Journal of Econometrics*, Vol. 18, pp. 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, pp. 277-97.

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

[6] Baltagi, B.H. (2001), "Econometric Analysis of Panel Data", Wiley, 2nd edition.

[7] Baltagi, B.H., Bresson, G., Griffin, J.M., Pirotte, A. (2002), "Homogeneous, Heterogeneous or Shrinkage Estimators? Some Empirical Evidence from French Regional Gasoline Consumption", mimeo, Texas A&M University.

[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, pp. 375-82.

[9] 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, pp. 303-27.

[10] 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), pp. 117-26.

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

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

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

[14] Clements, M.P., Hendry, D. (2002), "A Companion to Economic Forecasting", Blackwell Publishers, Oxford.

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

[16] Cornwell, C. and P. Rupert (1988), "Efficient Estimation with Panel Data: an Empirical Comparison of Instrumental Variables Estimators", *Journal of Applied Econometrics*, Vol. 3, pp. 149-55.

[17] Diebold, F.X., Mariano, R. (1995), "Comparing Predictive Accuracy", *Journal of Business Economics and Statistics*, 13, pp. 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, pp. 241-268.

[19] Driver, C., Urga, G. (2003), "Interpreting Business Survey Data. An Empirical Study Using the CBI Industrial Trends Survey", mimeo, Cass Business School.

[20] 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, pp. 1-9.

[21] Granger, C.W.J., Giacomini, R. (2003), "Aggregation of Space-Time Processes", forthcoming in *Journal of Econometrics*.

[22] 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..

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

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

[25] 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.

[26] Leitch, G., Tanner, J.E. (1991), "Economic Forecast Evaluation: Profits versus the Conventional Error Measures", *American Economic Review*, Vol. 81, No.3, pp. 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*, Vol. 2, pp. 1-30.

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

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

[30] Pesaran, M.H. (2003), "A Simple Panel Unit Root Test in the Presence of Cross Section Dependence", draft, Cambridge University.

[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*, Vol. 68, pp. 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, pp. 241-67.

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

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

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

Appendix I

Tables 1.a)-1.f) provide Theil's U statistics for each set of experiments. This statistics attains its minimum value (zero) when the prediction is perfect. The larger the statistics, the poorer the forecasting performance. We emphasized in bold the greatest value of the statistics per each experiment.

(T, N)	$(5, 10)$	$(5, 10)$	$(5, 10)$	$(5, 10)$
(ρ, ϑ, H)	$(0, 0, 0.1)$	$(0, 0, 0.9)$	$(0.9, 0.9, 0.1)$	$(0.9, 0.9, 0.9)$
OLS	**0.4252**	1.8201	0.4752	1.8828
Within	0.4296	1.8741	0.4404	1.9273
Between	0.6698	1.7434	0.7928	1.8152
FD-OLS	1.1656	1.5509	0.9997	1.9771
WLS	0.4344	1.8911	0.4708	1.9667
WLS-AR(1)	3.1086	20.8932	0.7649	15.5999
2SLS	0.4279	6.3207	0.4526	10.4835
FD-2SLS	0.7748	3.1532	**0.3033**	17.7187
Within-2SLS	**0.4252**	1.821	0.4752	1.8828
Between-2SLS	0.4300	36.1206	0.4428	3.0079
MLE	0.4268	1.7379	0.4411	1.7786
EC2SLS	0.4718	1.6463	0.4693	1.6850
EC2SLS-AR(1)	0.4757	1.5808	0.4473	1.6328
G2SLS	0.4672	$5.4{\cdot}10^5$	0.4638	528.1
2SLS-KR	0.4302	$11.4{\cdot}10^5$	0.4673	$\sim 10^{12}$
FD-2SLS-KR	0.7753	$2.9{\cdot}10^5$	**0.3039**	1364.9
FDGMM	0.7644	1.4142	0.3116	1.8925
GMM	0.4507	1.4370	0.5426	1.4656
Ind. OLS	$2.74{\cdot}10^5$	$8.6{\cdot}10^7$	$3.3{\cdot}10^5$	$2{\cdot}10^4$
Ind. 2SLS	$2.74{\cdot}10^5$	$8.6{\cdot}10^7$	$3.3{\cdot}10^5$	$2{\cdot}10^4$
Average OLS	0.4514	2.3245	0.4443	1.5325
Average 2SLS	0.4514	2.3245	0.4443	1.5325
Swamy	0.4297	1.8245	0.4531	1.8822
Bayes OLS	$1.2{\cdot}10^5$	$2{\cdot}10^5$	$2{\cdot}10^4$	192.2
It. Bayes OLS	$0.9{\cdot}10^5$	741.5	160.4	52.58
Bayes 2SLS	$1.2{\cdot}10^5$	$2{\cdot}10^5$	$2{\cdot}10^4$	192.2
It. Bayes 2SLS	$0.9{\cdot}10^5$	741.5	160.4	52.58
It. Bayes	0.4780	**0.5075**	0.3805	**0.4947**

Table 1.a). Results are reported for the case of no cross dependence.

(T,N)	$(5,20)$	$(5,20)$	$(5,20)$	$(5,20)$
(ρ,ϑ,H)	$(0,0,0.1)$	$(0,0,0.9)$	$(0.9,0.9,0.1)$	$(0.9,0.9,0.9)$
OLS	**0.4374**	1.2731	0.5463	1.4238
Within	0.4430	4.7539	0.5291	5.4049
Between	0.5461	1.1963	0.7895	1.3203
FD-OLS	1.1.740	1.7872	1.1027	2.2819
WLS	0.4426	1.5104	0.5209	1.8063
WLS-AR(1)	37.17	$1.6{\cdot}10^4$	4.6772	127.46
2SLS	0.4388	1.3913	0.5328	1.5570
FD-2SLS	0.7641	1.2963	0.5298	1.6177
Within-2SLS	**0.4374**	1.2731	0.5463	1.4238
Between-2SLS	0.4402	8.1617	0.5286	409.66
MLE	**0.4379**	4.0258	0.5282	4.5677
EC2SLS	0.4796	1.3017	0.5605	1.4480
EC2SLS-AR(1)	0.4823	1.4137	0.5556	1.6057
G2SLS	0.4753	1.4357	0.5537	1.6052
2SLS-KR	0.4396	1.1960	0.5285	1.3930
FD-2SLS-KR	0.7637	1.8935	0.5239	2515
FDGMM	0.7603	1.3410	0.5259	1.6999
GMM	0.4582	1.4986	0.5359	1.5559
Ind. OLS	$6{\cdot}10^8$	$3.4{\cdot}10^7$	$5.92{\cdot}10^5$	$1.2{\cdot}10^{13}$
Ind. 2SLS	$6{\cdot}10^8$	$3.4{\cdot}10^7$	$5.92{\cdot}10^5$	$1.2{\cdot}10^{13}$
Average OLS	0.4571	1.6222	0.5348	80.851
Average 2SLS	0.4571	1.6222	0.5348	80.851
Swamy	0.4394	1.4283	0.5449	1.6131
Bayes OLS	$9{\cdot}10^7$	$8.5{\cdot}10^5$	$1.35{\cdot}10^4$	$4.14{\cdot}10^5$
It. Bayes OLS	$3{\cdot}10^7$	$7.9{\cdot}10^5$	1717	$1.09{\cdot}10^5$
Bayes 2SLS	$9{\cdot}10^7$	$8.5{\cdot}10^5$	$1.35{\cdot}10^4$	$4.14{\cdot}10^5$
It. Bayes 2SLS	$3{\cdot}10^7$	$7.9{\cdot}10^5$	1717	$1.09{\cdot}10^5$
It. Bayes	0.4711	**0.4979**	**0.4852**	**0.5070**

Table 1.b). Results are reported for the case of no cross dependence.

(T,N)	$(10,20)$	$(10,20)$	$(10,20)$	$(10,20)$
(ρ,ϑ,H)	$(0,0,0.1)$	$(0,0,0.9)$	$(0.9,0.9,0.1)$	$(0.9,0.9,0.9)$
OLS	**0.4359**	1.1802	0.4942	1.3399
Within	0.4380	4.3263	0.4618	5.4408
Between	0.6192	0.8990	0.7096	0.8755
FD-OLS	1.1721	1.7838	0.9870	3.0327
WLS	**0.4352**	1.1956	0.4969	1.3984
WLS-AR(1)	$5.7{\cdot}10^5$	1.5227	561.2	2854
2SLS	0.4365	1.2437	0.4703	1.4150
FD-2SLS	0.7595	1.2928	**0.2643**	2.1386
Within-2SLS	**0.4359**	1.1802	0.4943	1.3399
Between-2SLS	0.4370	8.4888	0.4640	9.1411
MLE	0.4362	3.8103	0.4636	4.7385
EC2SLS	0.4833	0.9121	0.4967	0.9142
EC2SLS-AR(1)	0.4885	0.8758	0.4726	0.8370
G2SLS	0.4791	0.9097	0.4918	0.8996
2SLS-KR	0.4372	3.3284	0.4677	118.21
FD-2SLS-KR	0.7603	1.2402	**0.2645**	1.8856
FDGMM	0.7571	1.3901	0.2956	2.4016
GMM	0.4866	0.8715	0.5780	0.8470
Ind. OLS	0.4917	0.4862	0.4445	0.4760
Ind. 2SLS	0.5274	0.6794	0.4450	0.5071
Average OLS	0.4383	1.1130	0.4583	0.9885
Average 2SLS	0.4381	1.1130	0.4569	0.9852
Swamy	0.4408	0.9841	0.4896	1.0387
Bayes OLS	0.4583	0.4415	0.4309	0.4076
It. Bayes OLS	0.4412	0.4703	0.4278	0.3948
Bayes 2SLS	0.4601	0.4498	0.4293	0.4046
It. Bayes 2SLS	0.4416	0.4693	0.4261	0.3905
It. Bayes	0.4645	**0.4322**	0.4196	**0.3897**

Table 1.c). Results are reported for the case of no cross dependence.

(T, N) (ρ, ϑ, H)	$(10, 50)$ $(0, 0, 0.1)$	$(10, 50)$ $(0, 0, 0.9)$	$(10, 50)$ $(0.9, 0.9, 0.1)$	$(10, 50)$ $(0.9, 0.9, 0.9)$
OLS	**0.4231**	1.8514	0.4812	2.1556
Within	0.4242	3.5802	0.4446	4.2831
Between	0.5844	0.9458	0.6932	0.9101
FD-OLS	1.1634	1.8818	0.9956	3.0532
WLS	**0.4230**	1.7449	0.4833	2.0158
WLS-AR(1)	21.325	3.2576	1.4555	6.2860
2SLS	0.4233	1.8226	0.4524	2.0629
FD-2SLS	0.7339	1.4010	**0.2497**	2.6155
Within-2SLS	**0.4231**	1.8514	0.4812	2.1556
Between-2SLS	0.4234	4.1183	0.4449	5.1049
MLE	0.4232	3.1429	0.4470	3.7594
EC2SLS	0.4730	1.3635	0.4835	1.5265
EC2SLS-AR(1)	0.4772	0.9544	0.4489	0.9255
G2SLS	0.4686	1.3519	0.4770	1.4358
2SLS-KR	0.4234	1.2984	0.4457	1.4143
FD-2SLS-KR	0.734	1.1097	**0.2497**	1.8192
FDGMM	0.7331	1.5830	0.2620	2.9752
GMM	0.4508	1.2371	0.5507	1.3424
Ind. OLS	0.4894	0.4433	0.4393	0.4289
Ind. 2SLS	0.5053	0.6018	0.4370	0.4183
Average OLS	0.4244	0.9037	0.4432	0.9005
Average 2SLS	0.4242	0.9045	0.4414	0.8974
Swamy	0.4258	1.6072	0.4726	1.8437
Bayes OLS	0.4479	0.4190	0.4208	0.3681
It. Bayes OLS	0.4251	**0.4172**	0.4171	**0.3610**
Bayes 2SLS	0.4494	0.4225	0.4192	0.3636
It. Bayes 2SLS	0.4249	0.4201	0.4157	0.3566
It. Bayes	0.4482	0.4305	0.4099	0.4197

Table 1.d). Results are reported for the case of no cross dependence.

(T, N) (ρ, ϑ, H)	$(20, 50)$ $(0, 0, 0.1)$	$(20, 50)$ $(0, 0, 0.9)$	$(20, 50)$ $(0.9, 0.9, 0.1)$	$(20, 50)$ $(0.9, 0.9, 0.9)$
OLS	0.4218	1.8698	0.4803	2.1673
Within	0.4222	3.3117	0.4500	3.9204
Between	0.6619	0.9303	0.7312	0.8814
FD-OLS	1.1637	1.8951	0.9958	3.0845
WLS	**0.4207**	1.7100	0.4883	1.9626
WLS-AR(1)	1384	5.0995	1.4142	2.8458
2SLS	0.4219	1.7104	0.4516	1.9216
FD-2SLS	0.7333	1.4204	**0.2495**	2.6679
Within-2SLS	0.4218	1.8698	0.4803	2.1673
Between-2SLS	0.4219	4.0141	0.4434	5.0345
MLE	0.4219	3.0826	0.4528	3.6505
EC2SLS	0.4760	1.1623	0.4940	1.2160
EC2SLS-AR(1)	0.4822	0.9247	0.4778	0.8737
G2SLS	0.4723	1.0433	0.4886	1.0248
2SLS-KR	0.4220	5.7204	0.4445	75.30
FD-2SLS-KR	0.7336	1.0551	0.2495	1.7034
FDGMM	0.7330	1.6784	0.3041	3.2109
GMM	0.5119	0.9327	0.6196	0.9004
Ind. OLS	0.4428	0.3811	0.4483	0.3423
Ind. 2SLS	0.4446	0.3827	0.4425	0.3334
Average OLS	0.4221	0.8246	0.4472	0.8726
Average 2SLS	0.4220	0.8248	0.4418	0.8687
Swamy	0.4269	1.3887	0.5026	1.5204
Bayes OLS	0.4270	0.3772	0.4398	0.3380
It. Bayes OLS	0.4219	**0.3769**	0.4383	0.3363
Bayes 2SLS	0.4277	0.3783	0.4346	0.3295
It. Bayes 2SLS	0.4218	0.3780	0.4336	**0.3280**
It. Bayes	0.4390	0.3823	0.4384	0.3552

Table 1.e). Results are reported for the case of no cross dependence.

Appendix II

Tables 2.a)-2.f) provide Pesaran and Timmermann's (1992) measure for turning points prediction ability. This statistics is close to zero when the capability to detect turning points is poor; the larger the modulus of the statistics, the better the turning points forecasting ability.

(T,N) (ρ,ϑ,H)	$(5,10)$ $(0,0,0.1)$	$(5,10)$ $(0,0,0.9)$	$(5,10)$ $(0.9,0.9,0.1)$	$(5,10)$ $(0.9,0.9,0.9)$
OLS	1.5157	0.0442	1.7367	0.0801
Within	1.4154	1.7681	1.7640	1.9171
Between	0.4659	-0.0368	0.3644	0.0268
FD-OLS	-1.4445	-0.1922	0.0235	-0.3090
WLS	1.5289	0.0177	1.8817	0.0578
WLS-AR(1)	1.2191	0.4393	1.6620	0.4970
2SLS	1.5113	-0.0895	1.7460	-0.0431
FD-2SLS	1.6674	0.9027	**2.8754**	0.9953
Within-2SLS	1.5157	0.0442	1.7367	0.0801
Between-2SLS	1.4251	1.6882	1.7343	1.8183
MLE	1.5128	0.8167	1.7928	0.8575
EC2SLS	1.4989	0.7126	1.9884	0.7989
EC2SLS-AR(1)	1.4217	0.9230	1.9967	1.0616
G2SLS	1.5855	0.5091	2.0413	0.6094
2SLS-KR	1.5058	-0.0353	1.7585	0.0198
FD-2SLS-KR	1.6705	0.8814	2.8738	0.9733
FDGMM	**1.6902**	0.9184	2.8551	1.0020
GMM	1.1285	-0.0131	1.0057	0.0694
Ind. OLS	1.3796	2.2323	2.0988	2.7793
Ind. 2SLS	1.3796	2.2323	2.0988	2.7793
Average OLS	1.3926	1.1333	1.8179	1.2663
Average 2SLS	1.3926	1.1333	1.8179	1.2663
Swamy	1.4775	0.5824	1.8351	0.6441
Bayes OLS	1.4545	2.2506	2.1198	**2.7796**
It. Bayes OLS	1.5011	2.2563	2.1210	2.7791
Bayes 2SLS	1.4545	2.2506	2.1198	**2.7796**
It. Bayes 2SLS	1.5011	2.2563	2.1210	2.7791
It. Bayes	1.5662	**2.3351**	2.0884	2.5650

Table 2.a). Results are reported for the case of no cross dependence.

(T, N) (ρ, ϑ, H)	$(5, 20)$ $(0, 0, 0.1)$	$(5, 20)$ $(0, 0, 0.9)$	$(5, 20)$ $(0.9, 0.9, 0.1)$	$(5, 20)$ $(0.9, 0.9, 0.9)$
OLS	1.8793	0.1455	1.7930	0.2660
Within	1.7967	0.6254	1.8693	0.6274
Between	0.6650	-0.0249	0.4923	0.0854
FD-OLS	-1.4296	-0.3223	-1.3068	-0.4898
WLS	1.8711	0.1136	1.8012	0.2359
WLS-AR(1)	1.3780	0.0465	1.6473	0.1282
2SLS	1.8647	0.0259	1.8586	0.1547
FD-2SLS	1.6426	0.9149	2.2677	1.0461
Within-2SLS	1.8793	0.1455	1.7930	0.2660
Between-2SLS	1.8308	1.1563	1.8602	1.1929
MLE	**1.8819**	0.0577	1.8867	0.0696
EC2SLS	1.6032	0.2483	1.7830	0.3720
EC2SLS-AR(1)	1.5578	0.2268	1.8360	0.3706
G2SLS	1.6511	0.1057	1.8655	0.2277
2SLS-KR	1.8576	-0.0179	1.9285	0.0883
FD-2SLS-KR	1.6460	0.9007	**2.2734**	1.0356
FDGMM	1.6575	0.9135	2.2582	1.0401
GMM	1.5765	0.0027	1.2667	0.1655
Ind. OLS	1.3522	1.7363	1.7468	2.1593
Ind. 2SLS	1.3522	1.7363	1.7468	2.1593
Average OLS	1.7116	0.5292	1.8625	0.6771
Average 2SLS	1.7116	0.5292	1.8625	0.6771
Swamy	1.8584	0.3008	1.8122	0.4118
Bayes OLS	1.5336	1.7755	1.8445	2.1886
It. Bayes OLS	1.6288	1.7802	1.8942	**2.1979**
Bayes 2SLS	1.5336	1.7755	1.8445	2.1886
It. Bayes 2SLS	1.6288	1.7802	1.8942	**2.1979**
It. Bayes	1.7578	**1.8790**	1.9319	2.0729

Table 2.b). Results are reported for the case of no cross dependence.

(T, N) (ρ, ϑ, H)	$(10, 20)$ $(0, 0, 0.1)$	$(10, 20)$ $(0, 0, 0.9)$	$(10, 20)$ $(0.9, 0.9, 0.1)$	$(10, 20)$ $(0.9, 0.9, 0.9)$
OLS	1.8940	0.1601	2.0616	0.3281
Within	1.8686	0.3896	2.2651	0.3815
Between	0.3627	0.0068	0.2872	0.0865
FD-OLS	-1.4311	-0.3281	1.4110	-0.6745
WLS	1.8828	0.1553	1.9632	0.3236
WLS-AR(1)	1.3390	0.0771	1.8783	0.2362
2SLS	1.8889	0.0294	2.1953	0.2063
FD-2SLS	1.6553	0.9261	2.8753	1.2263
Within-2SLS	1.8940	0.1601	2.0616	0.3281
Between-2SLS	1.8820	1.1544	2.2421	1.1896
MLE	**1.8964**	-0.0204	2.2538	-0.0181
EC2SLS	1.5538	0.0835	2.0894	0.2589
EC2SLS-AR(1)	1.4707	0.0096	2.3043	0.1104
G2SLS	1.6221	0.0185	2.1540	0.2056
2SLS-KR	1.8743	-0.0006	2.2547	0.1484
FD-2SLS-KR	1.6540	0.9234	**2.8736**	1.2236
FDGMM	1.6621	0.9267	2.8078	1.2454
GMM	1.1041	0.0002	0.8994	0.1612
Ind. OLS	1.7322	1.9982	2.2394	2.6322
Ind. 2SLS	1.7126	1.9867	2.2528	2.6351
Average OLS	1.8923	0.7826	2.2668	0.8977
Average 2SLS	1.8814	0.7761	2.2820	0.9334
Swamy	1.8308	0.0743	2.0919	0.3067
Bayes OLS	1.8115	2.0170	2.2650	2.6306
It. Bayes OLS	1.8693	1.9927	2.2674	2.6276
Bayes 2SLS	1.7991	2.0098	2.2816	**2.6363**
It. Bayes 2SLS	1.8656	1.9839	2.2838	2.6347
It. Bayes	1.7850	**2.0265**	2.3314	2.4804

Table 2.c). Results are reported for the case of no cross dependence.

(T,N) (ρ,ϑ,H)	$(10,50)$ $(0,0,0.1)$	$(10,50)$ $(0,0,0.9)$	$(10,50)$ $(0.9,0.9,0.1)$	$(10,50)$ $(0.9,0.9,0.9)$
OLS	2.0308	0.8432	2.2123	1.0375
Within	2.0539	1.5085	2.4786	1.6013
Between	0.5198	0.1057	0.5726	0.2373
FD-OLS	-1.5026	-0.5947	0.9225	-0.8751
WLS	2.0250	0.8588	2.0549	1.0629
WLS-AR(1)	1.5595	0.7704	2.0565	0.9583
2SLS	2.0277	0.4112	2.4099	0.5692
FD-2SLS	1.7300	1.3878	2.9010	1.7071
Within-2SLS	2.0308	0.8432	2.2123	1.0375
Between-2SLS	**2.0476**	1.4494	2.5203	1.5778
MLE	2.0305	0.2240	2.4546	0.2593
EC2SLS	1.6810	0.6490	2.2157	0.8360
EC2SLS-AR(1)	1.6420	0.1872	2.4554	0.3323
G2SLS	1.7228	0.2638	2.2751	0.4135
2SLS-KR	2.0323	0.1848	2.5071	0.3174
FD-2SLS-KR	1.7305	1.4761	**2.9012**	1.8628
FDGMM	1.7340	1.3003	2.8754	1.5795
GMM	1.8347	0.4513	1.4645	0.6129
Ind. OLS	1.8101	2.2141	2.2979	2.7084
Ind. 2SLS	1.7924	2.2055	2.3083	2.7102
Average OLS	2.0366	1.3731	2.4837	1.6344
Average 2SLS	2.0365	1.3696	2.4978	1.6505
Swamy	1.9931	0.6985	2.2691	0.8822
Bayes OLS	1.9065	**2.2362**	2.3250	2.7070
It. Bayes OLS	2.0221	2.2227	2.3379	2.7072
Bayes 2SLS	1.9059	2.2262	2.3346	2.7120
It. Bayes 2SLS	2.0208	2.2153	2.3449	**2.7125**
It. Bayes	1.9243	2.1690	2.3871	2.5308

Table 2.d). Results are reported for the case of no cross dependence.

(T, N) (ρ, ϑ, H)	$(20, 50)$ $(0, 0, 0.1)$	$(20, 50)$ $(0, 0, 0.9)$	$(20, 50)$ $(0.9, 0.9, 0.1)$	$(20, 50)$ $(0.9, 0.9, 0.9)$
OLS	2.0335	0.8566	2.1985	1.0595
Within	2.0495	1.4905	2.4189	1.5643
Between	0.3158	0.0735	0.2982	0.1267
FD-OLS	-1.4960	-0.5708	1.1885	-0.8350
WLS	2.0196	0.8836	2.0812	1.0971
WLS-AR(1)	1.5571	0.7759	2.0171	0.9728
2SLS	2.0309	0.4329	2.4087	0.6015
FD-2SLS	1.7322	1.3700	**2.9002**	1.6907
Within-2SLS	2.0335	0.8566	2.1985	1.0595
Between-2SLS	**2.0468**	1.4064	2.5203	1.5013
MLE	2.0322	0.7182	2.3968	0.7876
EC2SLS	1.6503	0.5200	2.0895	0.6683
EC2SLS-AR(1)	1.5859	0.0957	2.2769	0.1810
G2SLS	1.7010	0.2231	2.1507	0.3032
2SLS-KR	2.0335	0.2938	2.5132	0.4516
FD-2SLS-KR	1.7315	1.4774	**2.9002**	1.8634
FDGMM	1.7343	1.3827	2.7851	1.7262
GMM	1.0420	0.2808	0.9318	0.4051
Ind. OLS	1.9102	2.3173	2.2822	2.6933
Ind. 2SLS	1.9000	2.3108	2.3185	2.7046
Average OLS	2.0445	1.4157	2.4316	1.6296
Average 2SLS	2.0457	1.4139	2.4797	1.6595
Swamy	1.9674	0.5810	2.0490	0.7598
Bayes OLS	1.9761	**2.3206**	2.3077	2.6911
It. Bayes OLS	2.0415	2.3153	2.3204	2.6904
Bayes 2SLS	1.9724	2.3175	2.3419	2.7043
It. Bayes 2SLS	2.0433	2.3124	2.3565	**2.7051**
It. Bayes	1.9243	2.3031	2.3404	2.6673

Table 2.e). Results are reported for the case of no cross dependence.