Equal Predictive Ability Tests for Panel Data
with Applications to OECD and IMF Forecasts

Oguzhan Akgun\(^1\), Alain Pirotte\(^{1,2}\), Giovanni Urga\(^3\), and Zhenlin Yang\(^4\)

\(^1\)LEDi, University of Burgundy, France
\(^2\)CRED, Paris-Panthéon-Assas University, France
\(^3\)Bayes Business School (formerly Cass), London, United Kingdom
\(^4\)School of Economics, Singapore Management University, Singapore

February 15, 2022

Abstract

We propose two types of equal predictive ability (EPA) tests for panels to compare the predictions made by two forecasters. The first focuses on \textit{the overall EPA hypothesis} which states that the EPA holds on average over all panel units and over time, and the second on \textit{the clustered EPA hypothesis} where the EPA holds jointly for a fixed number of clusters of panel units. The asymptotic properties of proposed tests are evaluated under weak and strong cross-sectional dependence. An extensive Monte Carlo simulation exercise shows that the tests have very good finite sample properties even with little information about the cross-sectional dependence in the data. The proposed framework is applied to compare the economic growth forecasts of the OECD and the IMF, and to evaluate the performance of the consumer price inflation forecasts of the IMF.

\textbf{Keywords}: Cross-Sectional Dependence; Forecast Evaluation; Hypothesis Testing.

\textbf{JEL classification}: C12, C23.

\(^1\) Corresponding author.

Email addresses: oguzhan.akgun@u-bourgogne.fr (O. Akgun), alain.pirotte@u-paris2.fr (A. Pirotte), g.urga@city.ac.uk (G. Urga), zlyang@smu.edu.sg (Z. Yang).
1 Introduction

Formal tests of the null hypothesis of no difference in forecast accuracy using two time series of forecast errors have been widely considered in the literature: see Vuong (1989), Diebold and Mariano (1995, hereafter DM), West (1996), Clark and McCracken (2001), Clark and McCracken (2015), Giacomini and White (2006), Clark and West (2007), Mariano and Preve (2012), among others. On the contrary, the literature on such tests using panel data is scarce with two exceptions: Keane and Runkle (1990) and Davies and Lahiri (1995).

The main aim of this paper is to develop testing procedures for equal predictive ability (EPA) hypotheses based on panel data, taking into account the cross-sectional dependence (CD) and the temporal dependence in the data set. Let $e_{l, it} = y_{it} - \hat{y}_{l, it}$ be the forecast error made by agent $l = 1, 2$ at time $t = 1, 2, \ldots, T$ for unit (e.g., country, firm) $i = 1, 2, \ldots, n$, where $y_{it}$ is the target value and $\hat{y}_{l, it}$ is the forecast of agent $l$. We propose tests for comparing the predictive ability of two forecasters, using $n$ time series of loss differentials, $L(e_{1, it}) - L(e_{2, it})$, of length $T$, where $L(\cdot)$ is a generic loss function. Our setting differs from that of Keane and Runkle (1990) and Davies and Lahiri (1995) in that we consider forecasts made by two forecasters on multiple economic units over time, whereas they consider forecasts made by multiple forecasters on a single economic unit over time. Our paper fills an important gap in the literature by allowing for multiple target values for each point in time.

We develop two types of tests corresponding to two EPA hypotheses. The first type focuses on the overall EPA hypothesis which states that the EPA holds on average over all units and over time. This test is useful when a researcher is not interested in the differences of predictive ability for a specific unit but in the overall differences. The second type focuses on the clustered EPA hypothesis which states that the EPA holds jointly for a fixed number of clusters of units in the panel.

The applied literature in comparing the accuracy of two or more forecasts suggests that the forecast errors of units, such as countries, are affected by common shocks such as the global financial crisis. For instance, Pain et al. (2014) show that the economic growth projections of the OECD for the period 2007-2012 are systematically upward biased. A similar tendency exists for other
forecasters, such as the IMF. Moreover, these effects are carried into the loss differentials such that they follow a similar pattern, as we highlight later in this paper. The results of Pain et al. (2014) indicate also that the effect of these common shocks is heterogeneous across economies and some country clusters exists.

Following these insights, we built our testing framework around the loss differentials which follow an approximate factor model where some common factors affect all units in the panel with heterogeneous loadings. In addition, the errors terms are allowed to be cross-sectionally weakly correlated. We therefore simultaneously allow the loss differentials to contain weak cross-sectional dependence (WCD) arising from, e.g., spatial error correlation, and strong cross-sectional dependence (SCD) due to the existence of common factors, using the terminology of Chudik et al. (2011). To develop our tests under WCD, we use non-parametric methods of variance estimation, based on geographic or economic distances between panel units (Kelejian and Prucha, 2007; Kim and Sun, 2013). In addition, we propose a novel partial sample variance estimator for large panels which deals with the case of unknown distances while being robust to arbitrary WCD and temporal dependence. To deal with SCD, we use the principal components estimator (PCE) built for large dimensional approximate factor models (Bai, 2003; Bai and Ng, 2002). Following Driscoll and Kraay (1998), we also propose tests for the case of unknown number of common factors contrary to the ones based on the PCE.

We analyze the asymptotic properties of the proposed test statistics using joint limits. Under mild conditions, the overall EPA test statistics and the clustered EPA test statistics are shown to converge in distribution to standard normal and chi-square with $G$ degrees of freedom under the null of interest, respectively, where $G$ is the number of clusters. The finite sample properties of the tests are examined via Monte Carlo simulations. The results show that the tests have very good finite sample performance even with little prior information about the CD in the data set.

The proposed tests are used in two applications. In the first one, we compare the economic growth forecast errors of the OECD and the IMF using data for 29 countries over the period between 1998 and 2016. In the second application, reported in an appendix, the quality of the IMF consumer price inflation (CPI) forecasts is challenged with that of random walk forecasts. This data
set contains 127 countries over the period 1991-2019. In both applications, we find strong evidence of SCD in loss differentials of forecast errors. The results show little difference between the economic growth forecast errors of the OECD and the IMF. Whereas, we find significant evidence against the overall EPA hypothesis on the comparison of the IMF and random walk forecasts, especially in the pre-crisis period.

The remainder of the paper is as follows: The testing framework and the hypotheses of interest are presented in Section 2. The panel EPA test statistics and their asymptotic properties are reported in Section 3. Section 4 is on the small sample properties of the proposed tests. In Section 5 the economic growth forecast errors of the OECD and the IMF are compared using the proposed tests. Section 6 concludes. Three appendices contain additional justification of our testing framework, proofs of results and a second application comparing the IMF consumer price inflation (CPI) forecasts with random walk forecasts.

**Notation.** Let \( A = [a_{ij}] \) be an \( n \times n \) matrix. The column and row norms of \( A \) are \( \|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}| \) and \( \|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| \), respectively. The Euclidean norm of an \( n \times m \) matrix \( B \) is \( \|B\| = [\text{Tr}(B^T B)]^{1/2} \). \( M \) is a finite positive constant. \( \rightarrow_d \) and \( \rightarrow_p \) denote convergence in distribution and probability, respectively. Joint passage to infinity of \( T \) and \( n \) is denoted by \( (T,n) \rightarrow \infty \).

## 2 Testing Framework and Null Hypotheses

We are interested in comparing the errors of forecasts made by two forecasters on an economic variable observed for units \( i = 1,2,\ldots,n \) at time \( t = 1,2,\ldots,T \). The loss differential of the errors, \( \Delta L_{it} = L(e_{1,it}) - L(e_{2,it}) \), is assumed to follow the model

\[
\Delta L_{it} = \mu_i + \lambda_i^T f_t + \varepsilon_{it},
\]

where \( f_t = (f_{1t}, f_{2t}, \ldots, f_{mt})' \) is an \( m \times 1 \) vector of unobservable common factors and \( \lambda_i = (\lambda_{1i}, \lambda_{2i}, \ldots, \lambda_{mi})' \) is an \( m \times 1 \) vector of fixed factor loadings. The coefficients \( r_{ij} \) are fixed, unknown elements of an \( n \times n \) matrix \( R_n \) except they satisfy \( r_{ii} = 1 \). \( f_t \) and \( \varepsilon_{it} \) are assumed to be...
zero mean weakly stationary time series allowed to be autocorrelated through time. In addition, $\epsilon_{it}$ is cross-sectionally uncorrelated. The unit specific means satisfy $|\mu_i| < \infty$.

We assume that the error terms $\epsilon_{it}$ carry WCD, meaning that the variance of their cross-sectional average vanishes asymptotically:

$$\lim_{n \to \infty} \text{Var} \left( n^{-1} \sum_{i=1}^{n} \epsilon_{it} \right) = \lim_{n \to \infty} \left( n^{-2} \sum_{i,j=1}^{n} r_i' \gamma_{n,0} r_j \right) = 0, \quad (2)$$

where $r_i = (r_{i1}, r_{i2}, \ldots, r_{in})'$, $\gamma_{n,0} = \text{diag}(\gamma_{11,0}, \ldots, \gamma_{nn,0})$ and $\gamma_{ii,0} = \mathbb{E}(\epsilon_{it}^2)$. For this to be true, it requires the row and column sum norms of $R_n$ to be bounded; see Assumption 3 below. The common component in the process is assumed to induce SCD such that the variance of its cross-sectional average is bounded away from zero:

$$\text{Var} \left( n^{-1} \sum_{i=1}^{n} \lambda_i' f_i \right) = n^{-2} \sum_{i,j=1}^{n} \lambda_i' \Gamma_0 \lambda_j > 0, \quad (3)$$

where $\Gamma_0 = \mathbb{E}(f_i f_i')$. The conditions under which this holds true are given below, in Assumption 6. The WCD can be modelled by a spatial process, such as spatial autoregression, spatial moving average as well as their higher order versions, or a factor model with a possibly infinite number of weak common factors (see, for instance, Chudik et al., 2011).

To model the loss differentials, we follow closely the methodological implications of DM. Firstly, we adopt the view that the forecast errors may be related to forecasts made using some econometric models or simply by expert knowledge. As in DM, our approach is built on a model-free environment, meaning that it is agnostic about the process generating the forecasts. This is contrary to the extensions of DM in papers such as West (1996) and Giacomini and White (2006) which focus, among other things, on problems associated with testing using forecast errors generated by nested models. Secondly, following DM, our assumptions concern directly the loss differentials. Forecast errors can be non-stationary and induce CD. What is important is the properties which are carried into the loss differentials, not the properties of the forecast errors specifically. As an example, in Appendix A we show how the model above can be obtained for two different loss functions, namely absolute and quadratic loss, starting with a pure factor model for the forecast errors. However, our approach is not necessarily limited to these loss functions. As in DM approach, any other (sym-
metric or asymmetric) function can be used. Eventually, our methodology relies on the empirical
analysis of time series and CD properties of the loss differentials, using CD tests (Pesaran 2015)
and information criteria (Bai and Ng, 2002).

The null hypotheses of EPA. First hypothesis of interest is
\[ H_{0,1} : \bar{\mu}_n = 0, \] (4)
where \( \bar{\mu}_n = n^{-1}\sum_{i=1}^{n} \mu_i \). This hypothesis states that the forecasts generated by the two agents are
equally accurate on average over all \( i \) and \( t \). If a researcher is not interested in the difference in
predictive power for a particular unit but the average difference over units, this hypothesis should
be considered. It is particularly plausible to consider this in a micro forecasting study where the
units are random draws from a population. Throughout the paper, this hypothesis is called the
overall EPA hypothesis.

In a macro forecasting study, differences between clusters of units can have a specific economic
importance and may be of interest from a policy perspective. For instance, a question of interest is
whether the forecasts made by agents are more accurate for a particular cluster of countries in the
sample. In this case, the null hypothesis can be formulated such that the predictive equality holds
for \( G \) clusters of units:
\[ H_{0,2} : \bar{\mu}_g = 0, \quad \text{for all } g = 1, 2, \ldots, G, \] (5)
where \( \bar{\mu}_g = n^{-1}_g\sum_{i \in G_g} \mu_i \) and \( G_g \) is the set of indexes of \( n_g \) cross-sectional units which belong to
cluster \( g \). In this paper, this hypothesis is referred to as the clustered EPA hypothesis.

The difference between the two hypotheses is important and the choice depends on the specific
interest. The overall EPA may hold even if the two forecasters have different predictive ability
for different clusters. This occurs if the average loss differentials of different country clusters are
different from zero but they add up to zero when pooled. If these clusters have an economic
meaning, it is important to test the clustered EPA hypothesis.
3 The Test Statistics and Their Asymptotic Properties

3.1 Tests for overall equal predictive ability

Consider the sample mean loss differential over time and units:

$$\Delta \bar{L}_{nT} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta L_{it}.$$ 

We provide testing procedures for the overall EPA implied in (1) based on $\Delta \bar{L}_{nT}$ under different assumptions about the structure of CD in the loss differentials. Let $k_T(\cdot)$ be a kernel function and $d_T$ a sequence of non-random bandwidth parameters. In all cases, the limiting null distribution of the test statistics is obtained under Assumptions 1 and 2.

Assumption 1. $\epsilon_{it}$ follows the linear process $\epsilon_{it} = \sum_{h=0}^{\infty} c_{ih} \psi_{i,t-h}$ for each $i$, with $\psi_{it} \sim IID(0,1)$ over $i$ and $t$, $E|\psi_{it}|^4 < \infty$, $\max_{1 \leq i \leq n} \sum_{h=0}^{\infty} h|c_{ih}| < M < \infty$ and $\sum_{h=0}^{\infty} c_{ih} > 0$.

Assumption 2. (a) $k_T(x) : \mathbb{R} \rightarrow [-1,1]$ is continuous at zero, $k_T(0)=1$, $k_T(x) = k_T(-x)$ $\forall x \in \mathbb{R}$, (b) $\lim_{T \rightarrow \infty} d_T^{-1} \sum_{h=1}^{T-1} |k_T(h/d_T)| < \infty$, (c) $d_T \rightarrow \infty$ such that $d_T^2/T \rightarrow 0$ as $T \rightarrow \infty$.

Assumption 1 is sufficient to obtain a CLT for the sample mean of $\epsilon_{it}$ for each $i$. Assumption 2(a) is standard in time series literature. Assumption 2(b) is a high level assumption that is used for consistent estimation of long-run variance for each $i$. The conditions under which this assumption holds are given by Jansson (2002). Consistency also requires Assumption 2(c) which controls the expansion of $d_T$ relative to $T$.

Overall EPA tests under cross-sectional independence. We first consider the simplest case where $X'_i \mathbf{f}_t = 0$ for all $i,t$ and $r_{ij} = 0$ for all $i \neq j$. In order to test the null hypothesis $H_{0,1}$, we propose to use the following statistic:

$$S_{nT}^{(1)} = \frac{\Delta \bar{L}_{nT}}{\hat{\sigma}_{1,nT}/\sqrt{nT}},$$

where $\hat{\sigma}_{1,nT}^2 = (nT)^{-1} \sum_{i=1}^{n} \sum_{t,s=1}^{T} k_T(d_{ts}/d_T) \Delta \bar{L}_{it} \Delta \bar{L}_{is}$, with $\Delta \bar{L}_{it} = \Delta L_{it} - \Delta \bar{L}_{i,T}$, where $\Delta \bar{L}_{i,T} = T^{-1} \sum_{t=1}^{T} L_{it}$ and $d_{ts} = |t-s|$. 

7
Proposition 1. Suppose $\Delta L_{it}$ follows the model in (1) with $\chi'_i f_t = 0$ for all $i, t$, $r_{ij} = 0$ for all $i \neq j$, and Assumptions 2 and 3 hold. Then, under $H_{0,1}$ and as $(T, n) \to \infty$, $S_{nT}^{(1)} D \to N(0,1)$.

To prove this result, it is sufficient to show that
\[
\sqrt{nT} \left( \Delta \bar{L}_{nT} - \bar{\mu}_{nT} \right) / \sigma_{1,nT} \xrightarrow{d} N(0,1)
\]
where $\sigma_{1,nT}^2 = n^{-1} \sum_{i=1}^n \gamma_{i,T}$ with $\gamma_{i,T} = T^{-1} \sum_{t,s=1}^T \gamma_{i,d_{ts}}$, $\gamma_{i,d_{ts}} = E(\epsilon_{it}\epsilon_{is})$, and $\sigma_{1,nT}^2 - \sigma_{2,nT}^2 P \to 0$.

Overall EPA tests under WCD. Suppose $\chi'_i f_t = 0$ for all $i, t$ but $r_{ij} \neq 0$ for some $i \neq j$. In this case, the loss differentials $\Delta L_{it}$ are no longer independent across $i$. Define $d_{ij} = d_{ji} \geq 0$ as the distance between units $i$ and $j$. We make Assumptions 3 and 4 on the coefficients $r_{ij}$.

Assumption 3. For all $n \in \mathbb{Z}^+$, $||R_n||_1 < \infty$ and $||R_n||_{\infty} < \infty$.

Assumption 4. $\sum_{j=1}^n |r'_i r_j| |d_{ij}|^{\rho} < \infty$ for some $\rho \geq 1$.

Assumption 3 is standard in spatial econometrics literature and implies the WCD defined in (2). The role of Assumption 4 is to restrict the spatial correlation among panel units in relation to the distances between them. As noted by Kelejian and Prucha (2007), the corresponding condition in the time series context is fading memory over time. Under this assumption, as the distance between two panel units increase, the correlation between them decreases. This in turn sets a basis for using the spatial kernel function $k_S(\cdot)$ which gives smaller weights to the covariance between units which are more distant from each other.

To deal with the WCD when $T = 1$, KP proposed a spatial heteroskedasticity and autocorrelation consistent estimator of the variance-covariance matrix. This estimator can be seen as a spatial version of the kernel estimators for time series such that it uses a spatial kernel based on the distance between units. The estimator is generalized by Kim and Sun (2013) to panel data regression, by combining the spatial and time kernels. Define $k_q = \lim_{x \to 0} [1 - k_S(x)]/|x|^q$ and let $\rho_s = \max\{q : k_q < \infty\}$ be the Parzen exponent of $k_S(\cdot)$ (Andrews, 1991). Assumption 5 is placed on the spatial kernel function.

Assumption 5. (a) $k_S(x) : \mathbb{R} \to [-1, 1]$ is continuous at zero with $\rho_s \geq 1$, $k_S(0) = 1$, $k_S(x) = k_S(-x)$ $\forall x \in \mathbb{R}$, (b) $\max_{1 \leq i \leq n} \lim_{n \to \infty} d_n^{-1} \sum_{j=1}^n |k_S(d_{ij}/d_n)| < \infty$, (c) $d_n \to \infty$ such that $d_n/n \to 0$ as $n \to \infty$. 

8
The condition in [4(a)] is satisfied by all kernel functions used in practice, such as Bartlett, Parzen, Tukey–Hanning and quadratic spectral (see Andrews, 1991). Conditions similar to that in [4(b)] are used by Kelejian and Prucha (2007), Moscone and Tosetti (2012) and Kim and Sun (2013). All these studies allow solely for kernels which truncate, i.e. those which equals zero after a certain value of the bandwidth parameter. The first two papers place assumptions on the relative expansion of \( l_n = \max_{1 \leq i \leq n} l_{i,n} \) where \( l_{i,n} = \sum_{j=1}^n 1\{d_{ij} \leq d_n\} \). Assumption [4(c)] controls the expansion of \( d_n \) relative to \( n \). This condition, as well as [4(b)] is not necessary for the consistent estimation of the variance-covariance matrices in panels using a time series kernel. Nevertheless, these conditions are retained to compare our results with the existing literature.

We propose the following test statistic in this case of WCD:

\[
S_{nT}^{(2)} = \frac{\Delta \tilde{L}_{nT}}{\hat{\sigma}_{2,nT}/\sqrt{nT}},
\]

where

\[
\hat{\sigma}_{2,nT}^2 = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T k_S \left( \frac{d_{ij}}{d_n} \right) k_T \left( \frac{d_{ts}}{d_T} \right) \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}.
\]

Theoretical properties of the estimator \( \hat{\sigma}_{2,nT}^2 \) are explored by Kim and Sun (2013). Moscone and Tosetti (2012) use a similar estimator with the difference being that they set \( k_T(\cdot) = 1 \).

The disadvantage of this variance estimator is that for its implementation the distances between all pairs of units, \( d_{ij} \) have to be known to the researcher. Furthermore, the cut-off distance \( d_n \) has to be chosen. In practice there may be many possible distance metrics and economic theory does not always help to choose between them. When a distance metric is not available we can use a partial sample estimator given by

\[
\hat{\sigma}_{2,nT}^2 = \frac{1}{n^2} \sum_{i,j=1}^n \sum_{t,s=1}^T k_T \left( \frac{d_{is}}{d_T} \right) \Delta \tilde{L}_{it} \Delta \tilde{L}_{js},
\]

where \( n \), an increasing function of \( n \), is the number of observations used to calculate the variance. It is strictly smaller than \( n \). Similar variance estimators are used by Bai and Ng (2006) and Moscone and Tosetti (2015). The first study focuses on the factor models whereas the second one deals with panel regression models with small \( T \). Our variance estimator generalizes that of the Moscone and Tosetti (2015) by allowing for a large \( T \) with the help of a time kernel. The
Proposition 2. Suppose \( \Delta L_{it} \) follows the model in (1) with \( \lambda_i^t f_i = 0 \) for all \( i, t \) but \( r_{ij} \neq 0 \) for some \( i \neq j \), and Assumptions 7 and 8 hold. Then, under \( H_{0,1} \) and as \( (T, n) \to \infty \), (i) \( S_{nT}^{(2)} \overset{d}{\to} N(0, 1) \) if \( d_n \to \infty \), (ii) and \( S_{nT}^{(2)} \overset{d}{\to} N(0, 1) \) if \( n \to \infty \) such that \( n/\kappa \to \in \{0, 1\} \).

Under Assumptions 6-8 on the common factors, their loadings and the error terms.

Overall EPA tests under SCD. This is the most general case with no specific restriction imposed on the CD of the loss differentials. To obtain a CLT for the means, we make Assumptions 6-8 and 8 on the common factors, their loadings and the error terms.

Assumption 6. (a) \( f_t \) is independent of \( \varepsilon_{it} \) and follows the linear process \( f_t = \sum_{h=0}^{\infty} C_h \Psi_{t-h} \) with \( \Psi_t \sim IID(0, I_m) \), \( E[\|\Psi_t\|^d] < \infty \), \( \sum_{h=0}^{\infty} h \|C_h\| < M < \infty \) and \( \sum_{h=0}^{\infty} C_h \) is full rank. (b) Factor loadings \( \lambda_i \) are fixed parameters such that \( ||\lambda_i|| < \infty \), \( n^{-1} \sum_{i=1}^{n} \lambda_i \lambda_i' \rightarrow \Sigma_{\lambda} > 0 \) for an \( m \times m \) matrix \( \Sigma_{\lambda} \). (c) There exists at least one common factor \( f_{kt}, k \in \{1, \ldots, m\} \), for which the loadings satisfy \( n^{-1} \sum_{i=1}^{n} \lambda_{ki} > 0 \) for all \( n \in \mathbb{Z}^+ \).

Assumption 7. \( E[\psi_t^8] < \infty \) and \( E[n^{-1/2} \sum_{i=1}^{n} (\varepsilon_{it} \varepsilon_{is} - E(\varepsilon_{it} \varepsilon_{is}))^4] < \infty \) for all \( i, t, s \).

Assumption 8. (a) For each \( t \), \( E[||(nT)^{-1/2} \sum_{i=1}^{n} f_{it} (\varepsilon_{it} \varepsilon_{is} - E(\varepsilon_{it} \varepsilon_{is}))||^2] < \infty \). (b) The \( m \times m \) matrix satisfies \( E[||(nT)^{-1/2} \sum_{i=1}^{n} f_i \lambda_i' \varepsilon_{it}||^2] < \infty \).

It is easy to see that, \( \lambda_i^t f_i \) satisfying Assumption 6 lead to SCD as defined in (3). We require Assumption 6(a) for the consistent estimation of the long-run variance of the common factors. Assumption 6(b) is standard for factor models. Assumption 6(c) ensures that at least one of the factors contribute to the asymptotic variance of the cross-sectional averages. Although nonstandard, this assumption is not restrictive as it does not affect the validity of our testing procedures. Assumptions 7 and 8 are also standard in the literature on approximate factor models [Bai, 2004].
In this case of SCD, the variance estimator given in \((8)\) can be modified by setting \(k_S(\cdot) = 1\). This variance estimator does not require any knowledge of a distance measure between the units. Moreover, it assigns weights equal to one for all covariances from the same time period, hence robust to SCD as well as WCD. The test statistic takes the form:

\[
S_{nT}^{(3)} = \frac{\bar{\Delta} L_{nT}}{\hat{\sigma}_{3,nT}/\sqrt{T}} \quad (10)
\]

where

\[
\hat{\sigma}_{3,nT}^2 = \frac{1}{n^2 T} \sum_{i,j=1}^n \sum_{t,s=1}^T k_T \left( \frac{d_{ts}}{d_T} \right) \bar{\Delta} L_{it} \bar{\Delta} L_{js} \quad (11)
\]

The variance estimator \((11)\) was proposed by DK, which is valid when \(T\) is large, regardless of \(n\) being finite or infinite.

An alternative way to estimate the covariance matrix is to exploit the factor structure of the DGP. The PCE of large panels is investigated by Stock and Watson (2002), Bai and Ng (2002), Bai (2003), among others. This is a flexible estimator which is robust to WCD and autocorrelation in the error terms. It is obtained by minimizing the average squared residuals computed for \(m\) common factors:

\[
V(m) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\bar{\Delta} L_{it} - \lambda_i f_t)^2 \quad (12)
\]

subject to \(\text{Var}(f_t) = I_m\) and \(\Lambda_n = (\lambda_1, \lambda_2, \ldots, \lambda_n)'\). Then the solution for the estimates of the common factors, \(\hat{f}_t\), are given by \(\sqrt{T}\) times the first \(m\) eigenvectors of the matrix \(\sum_{i=1}^n \bar{\Delta} L_{it} \bar{\Delta} L_{it}'\) with \(\bar{\Delta} L_{it} = (\Delta \tilde{L}_{i1}, \Delta \tilde{L}_{i2}, \ldots, \Delta \tilde{L}_{iT})'\) and the factor loadings can be estimated as \(\hat{\lambda}_i = \frac{1}{T} \sum_{t=1}^T \hat{f}_t \Delta \tilde{L}_{it}\). Then the overall EPA hypothesis can be tested using

\[
S_{nT}^{(3)} = \frac{\Delta \tilde{L}_{nT}}{\hat{\sigma}_{3,nT}/\sqrt{T}} \quad (13)
\]

where

\[
\hat{\sigma}_{3,nT}^2 = \frac{1}{n^2 T} \sum_{i,j=1}^n \sum_{t,s=1}^T k_T \left( \frac{d_{ts}}{d_T} \right) \hat{\lambda}_i \hat{f}_t \hat{\lambda}_j + \frac{1}{n^2 T} \sum_{i=1}^n \sum_{t,s=1}^T k_T \left( \frac{d_{ts}}{d_T} \right) \hat{\varepsilon}_{it} \hat{\varepsilon}_{is} \quad (14)
\]

with \(\hat{\varepsilon}_{it} = \Delta \tilde{L}_{it} - \hat{\lambda}_i \hat{f}_t\).

**Proposition 3.** Suppose \(\Delta L_{it}\) follows the model in \((1)\), Assumptions \((1) - (8)\) hold. Then, under \(H_{0,1}\) and as \((T,n) \to \infty\), (i) \(S_{nT}^{(3)} \overset{d}{\to} N(0,1)\), (ii) and \(\hat{\sigma}_{nT}^{(3)} \overset{d}{\to} N(0,1)\).
Under Assumptions 1-3, 6 and 7, \( \sqrt{T}(\Delta \bar{L}_{nT} - \bar{\mu}_n)/\sigma_{nT} \overset{d}{\rightarrow} N(0,1) \), where
\[
\sigma_{3,nT}^2 = \frac{1}{n^2} \sum_{i,j=1}^{n} (\lambda_i' \bar{\Gamma}_T \lambda_j + r_i' \bar{\gamma}_{nT} r_j). 
\] (15)
with \( \bar{\Gamma}_T = T^{-1} \sum_{t,s=1}^{T} \Gamma_{d_{ts}} \). The rate of convergence in the CLT is \( T^{1/2} \) instead of the usual rate of \( (nT)^{1/2} \) in the cases of no CD and WCD. This follows from the SCD characterized in (3).

In (15), the first term in parentheses dominate the second one. This is because, under Assumptions 1, 3 and 6, \( n^{-2} \sum_{i,j=1}^{n} \lambda_i' \bar{\Gamma}_T \lambda_j = O(1) \) but \( n^{-2} \sum_{i,j=1}^{n} r_i' \bar{\gamma}_{nT} r_j = O(1/n) \). Hence, the latter is asymptotically negligible. This means that under SCD, one can use a simpler variance estimator without the second term in (14).

### 3.2 Tests for clustered equal predictive ability

Define \( G_g, g = 1, \ldots, G \), as the set of indexes of \( n_g \) cross-sectional units which belong to cluster \( g \) such that \( G_g \cap G_{g'} = \emptyset, \forall g \neq g' \). In this subsection, we are interested in testing the null hypothesis \( H_{0,2} \) which can be written as
\[
H_{0,2} : \bar{\mu}_n = 0, \tag{16}
\]
where \( \bar{\mu}_n = (\bar{\mu}_{1,n}, \bar{\mu}_{2,n}, \ldots, \bar{\mu}_{G,n_G})' \), \( \bar{\mu}_{g,n_g} = n_g^{-1} \sum_{i \in G_g} \mu_i \). Our tests are based on the empirical counterpart of this quantity: \( \Delta \bar{L}_{nT} = (\Delta \bar{L}_{1,n_1,T}, \Delta \bar{L}_{2,n_2,T}, \ldots, \Delta \bar{L}_{G,n_G,T})' \) where \( \Delta \bar{L}_{g,n_g,T} = (n_g T)^{-1} \sum_{i \in G_g} \sum_{t=1}^{T} \Delta \bar{L}_{it} \). We assume that the sets of indexes \( G_g, g = 1, \ldots, G \) are known. Assumption 9 is placed to control the asymptotic number of units per cluster.

**Assumption 9.** As \( n \to \infty \), \( n_g/n \to \tau_g \in (0,1) \) for each \( g = 1, \ldots, G \).

With this assumption we do not rule out the possibility of having \( G = 1 \) in which case we have \( n_1 = n \). This particular case corresponds to the overall EPA tests of the previous subsection. In what follows, these test statistics are generalized for \( G > 1 \) for each case of CD.

**Clustered EPA tests under cross-sectional independence.** In this case \( \lambda_i' \ell_t = 0 \) for all \( i, t \) and \( r_{ij} = 0 \) for all \( i \neq j \). We propose the following statistic to test the hypothesis in (16):
\[
C_{nT}^{(1)} = nT \Delta \bar{L}_{nT} \hat{\bar{\Omega}}_{1,nT}^{-1} \Delta \bar{L}_{nT}, \tag{17}
\]
where

\[ \hat{\Omega}_{1,nT} = \frac{1}{T} \sum_{i=1}^{n} \frac{n}{n_i} \sum_{t,s=1}^{T} \sum_{i,j=1}^{G} \left( \frac{d_{is}}{d_T} \right) h_{gi} h'_{gj} \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}, \]

with \( g_i \in \{1, 2, \ldots, G\} \) being a variable which states the cluster which \( i \)th unit belongs to and \( h_{gi} \) being the \( g_i \)th column of \( I_G \). Notice that \( \hat{\Omega}_{1,nT} \) is a diagonal matrix which contains an estimate of the average long-run variances of each cluster as diagonal elements up to a factor which is asymptotically equal to \( \tau^{-1} \).

**Proposition 4.** Suppose \( \Delta L_{it} \) follows the model in (1) with \( X_i f_i = 0 \) for all \( i, t \), \( r_{ij} = 0 \) for all \( i \neq j \), and Assumptions 1, 2 and 9 hold. Then, under \( H_{0,2} \) and as \((T, n) \rightarrow \infty \), \( C_{nT}^{(1)} \overset{\text{d}}{\rightarrow} \chi^2_G \).

Under Assumptions 1, 2 and 9 we have \( \sqrt{nT} \hat{\Omega}_{1,nT}^{-1/2}(\Delta \tilde{L}_{nT} - \mu_n) \overset{\text{d}}{\rightarrow} N(0, I_G) \) as \((T, n) \rightarrow \infty \), where \( \Omega_{1,nT} = \sum_{i=1}^{n} \frac{n}{n_i} h_{gi} h'_{gj} \tilde{\gamma}_{i,T} \). This is a generalization of the CLT following Proposition 1.

The latter is obtained when \( G = 1 \).

**Clustered EPA tests under WCD.** Suppose \( X_i f_i = 0 \) for all \( i, t \) but \( r_{ij} \neq 0 \) for some \( i \neq j \). We can use the following statistic in order to test \( H_{0,2} \):

\[ C_{nT}^{(2)} = nT \Delta \tilde{L}'_{nT} \hat{\Omega}_{2,nT}^{-1} \Delta \tilde{L}_{nT}, \tag{18} \]

where

\[ \hat{\Omega}_{2,nT} = \frac{1}{T} \sum_{i,j=1}^{n} \frac{n}{n_g} \frac{1}{n_{gj}} \sum_{t,s=1}^{T} \sum_{i,j=1}^{G} \left( \frac{d_{is}}{d_T} \right) h_{gi} h'_{gj} \Delta \tilde{L}_{it} \Delta \tilde{L}_{js} \].

As discussed in the previous section, the estimator has the disadvantage of relying on a known distance between each unit in the panel. An alternative variance estimator for this case of cross-sectional clusters can be constructed as in (9). Such an estimator is

\[ \hat{\Omega}_{2,nT} = \frac{1}{T} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{n}{n_g} \frac{1}{n_{gj}} \sum_{i,j=1}^{T} \sum_{t,s=1}^{T} \left( \frac{d_{is}}{d_T} \right) h_{gi} h'_{gj} \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}, \tag{19} \]

where \( n = \sum_{g=1}^{G} n_g \), with \( n_g \) being the number of observations taken into account in the calculation of variance in cluster \( g \). The corresponding test statistic is

\[ C_{nT}^{(2)} = nT \Delta \tilde{L}'_{nT} \hat{\Omega}_{2,nT}^{-1} \Delta \tilde{L}_{nT}. \tag{20} \]

**Theorem 1.** Suppose \( \Delta L_{it} \) follows the model in (1) with \( X_i f_i = 0 \) for all \( i, t \) but \( r_{ij} \neq 0 \) for some
i \neq j, \text{ and Assumptions 1-5 and 9 hold. Then, under } H_{0.2} \text{ and as } (T, n) \to \infty, (i) \frac{C_{nT}^{(2)}}{d_n} \to \chi^2_G \text{ if } d_n \to \infty, (ii) \text{ and } \frac{C_{nT}^{(2)}}{\sqrt{n}} \to \chi^2_G \text{ if as } n \to \infty, \forall g = 1, \ldots, G \text{ such that } \frac{n_g}{n} \to \kappa \in \{0, 1\}. \]

Under the assumptions of the theorem, we have \( \sqrt{n} \frac{\Omega_{2,nT}^{-1/2}}{d_n} (\Delta \bar{L}_{nT} - \bar{\mu}_n) \overset{d}{\to} N(0, I_G) \) as \( (T, n) \to \infty \), where \( \Omega_{2,nT} = \sum_{i,j=1}^{n} \frac{1}{n} h_{gi} h'_{gj} r_i r_j \). Then the stated results follow from the consistency of the variance estimators. This theorem nests Propositions 1, 2 and 4. The first is obtained when \( R_n = I_n \) and \( G = 1 \), the second is obtained with only \( G = 1 \) and the last is when only \( R_n = I_n \).

**Clustered EPA tests under SCD.** We now consider the case of no specific restrictions on the CD in the loss differentials. With SCD, the cluster based EPA hypothesis \( H_{0,2} \) can be tested using:

\[
C_{nT}^{(3)} = T \Delta \bar{L}_{nT} \Omega_{3,nT}^{-1} \Delta \bar{L}_{nT};
\]

\[
\hat{\Omega}_{3,nT} = \frac{1}{T} \sum_{i,j=1}^{n} \frac{1}{n} \sum_{t,s=1}^{T} k_T \left( \frac{d_{ts}}{d_T} \right) h_{gi} h'_{gj} \Delta \bar{L}_{it} \Delta \bar{L}_{js}.
\]

As discussed previously, this type of variance estimator is robust to arbitrary CD and it has the advantage of not requiring known distances between units. However, its performance may be poor in cases of \( n \) large relative to \( T \). Hence, once more the factor structure of the loss differentials can be used to form variance estimators. A test statistic with such an estimate is

\[
C_{nT}^{(3)} = T \Delta \bar{L}_{nT} \hat{\Omega}_{3,nT}^{-1} \Delta \bar{L}_{nT};
\]

\[
\hat{\Omega}_{3,nT} = \frac{1}{T} \sum_{i,j=1}^{n} \frac{1}{n} \sum_{t,s=1}^{T} k_T \left( \frac{d_{ts}}{d_T} \right) h_{gi} h'_{gj} \hat{\lambda}_i \hat{\lambda}'_j + \frac{1}{T} \sum_{i=1}^{n} \frac{1}{n} \sum_{t,s=1}^{T} k_T \left( \frac{d_{ts}}{d_T} \right) h_{gi} h'_{gj} \hat{\epsilon}_i \hat{\epsilon}'_j.
\]

**Theorem 2.** Suppose \( \Delta L_{it} \) follows the model in (1), Assumptions 1-3, 6-9 hold. Then, under \( H_{0,2} \) and as \( (T, n) \to \infty \), (i) \( \frac{C_{nT}^{(3)}}{d_n} \to \chi^2_G \); (ii) and \( \frac{C_{nT}^{(3)}}{\sqrt{n}} \to \chi^2_G \).

Under the assumptions of the theorem, we have \( \sqrt{T} \frac{\Omega_{3,nT}^{-1/2}}{d_n} (\Delta \bar{L}_{nT} - \bar{\mu}_n) \overset{d}{\to} N(0, I_G) \), where \( \Omega_{3,nT} = \sum_{i,j=1}^{n} \frac{1}{n} h_{gi} h'_{gj} (\lambda_i \Gamma_T \lambda_j + r_i r_j \hat{\gamma}_{nT} r_j) \). Proposition 3 is a special case of this theorem with \( G = 1 \).
3.3 Special cases and extensions

Tests for joint equal predictive ability. In macroeconomic applications, the differences in the predictive ability for each cross-sectional unit can have a specific economic importance. When this is the case, one may be interested in the following hypothesis:

\[ H_{0,3} : \mathbb{E}(\Delta L_t) = \mu_i = 0, \text{ for all } i = 1, 2, \ldots, n. \]

This hypothesis, namely the joint EPA hypothesis, states that the EPA hypothesis holds for each unit in the sample. This hypothesis can be seen as a special case of the clustered EPA hypothesis \( H_{0,2} \) with the number of clusters being equal to the number of units in the panel, that is \( G = n \). However, in this case Assumption 9 is violated as for each \( g = 1, \ldots, G, \) \( n_g = 1 \) and therefore \( n_g/n \to 0 \). Nevertheless, the hypothesis can still be tested using the above test statistics if \( n \) is fixed, after suitable modification of the convergence rates. For instance, a test statistic which is robust to arbitrary CD is

\[ J_{nT} = T \Delta \tilde{L}_{nT} \hat{\Omega}_{nT}^{-1} \Delta \tilde{L}_{nT} \xrightarrow{D} \chi^2_n, \]

where \( \hat{\Omega}_{nT} = \frac{1}{T} \sum_{i,j=1}^n \sum_{t,s=1}^T k_T \left( \frac{d_{ts}}{d_T} \right) h_i h'_j \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}, \)

with \( \Delta \tilde{L}_{nT} = (\Delta \tilde{L}_{1T}, \Delta \tilde{L}_{2T}, \ldots, \Delta \tilde{L}_{nT})' \), and \( h_i \) being the \( i \)th column of \( \mathbb{I}_n \). Different cases of CD can be covered by modifications on the variance estimator. These cases are discussed at length in a previous version of this paper (available at https://arxiv.org/abs/2003.02803), where we also consider the case of large \( n \).

Tests for individual cross-sections. The overall test statistics \( S_{nT}^{(1)}, S_{nT}^{(2)} \) and \( S_{nT}^{(2)} \), and clustered test statistics \( C_{nT}^{(1)}, C_{nT}^{(2)} \) and \( C_{nT}^{(2)} \) are directly applicable to a single cross-section which is a special case where \( T = 1 \). For a single cross-section, a CLT can be obtained as a basis for these statistics by replacing the long-run variances with the variance of the loss differentials. Then, we can apply the CLT for independent but heterogeneous sequence (see, e.g. [White, 2001], Theorem 5.10) to show the normality of tests with no CD, and the CLT for spatially correlated triangular arrays of [Kelejian and Prucha, 1998] to show the normality of tests with WCD, respectively.
4 Monte Carlo Study

To investigate the finite sample properties of the test statistics given above, a set of Monte Carlo experiments are conducted. 2000 samples are generated from each DGP described below for the dimensions of $T \in \{10, 20, 30, 50, 100\}$ and $n \in \{10, 20, 30, 50, 100\}$. All tests are applied for the nominal size of 5%.

4.1 Design

Two different DGPs are considered to explore the effect of WCD and SCD on the performance of the tests. DGP1 contains only WCD. In this case, for each unit $i$, two independent forecast error series $(e_{1,it}, e_{2,it})$ are generated using spatial AR(1) processes defined as

$$
\zeta_{l,it} = \rho \sum_{j=1}^{n} w_{ij} \zeta_{l,j} + u_{l,it}, \quad \text{with,} \quad u_{l,it} \sim N(0,1), \quad l = 1, 2,$$

where $w_{ij}$ is the element of the spatial matrix $W_n$ in row $i$ and column $j$. To make the power results across different experiments comparable, the average unconditional variance of the forecast error series $e_{l,it}$, $l = 1, 2$, is held fixed. Such series are generated as

$$
e_{l,n,t} = \frac{1}{\bar{s}_2} S_n u_{l,n,t}, \quad (24)$$

where $u_{l,n,t} = (u_{l,1,t}, u_{l,2,t}, \ldots, u_{l,nt})'$, $S_n = (I_n - \rho W_n)^{-1}$ and $\bar{s}_2 = n^{-1} tr(S_n S_n')$. It can now be shown that the average of the diagonal elements of the variance-covariance matrix of this process is equal to one. In this DGP a quadratic loss function is used.

DGP2 contains SCD and WCD. In this case, following Giacomini and White (2006), we generate the loss differential directly, so the tests do not rely on a specific loss function. This is given by

$$
\Delta L_{it} = \phi (\mu_i + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + \varepsilon_{it}).
$$

To investigate the size properties we set $\mu_i = 0$ for each $i = 1, 2, \ldots, n$ and generate factor loadings as $\lambda_{1i}, \lambda_{2i} \sim N(1, 0.2)$. The common factors are formed by $f_{1t}, f_{2t} \sim N(0, 1)$, The error series $\varepsilon_{it}$ are generated in the same spirit as in (24). We finally set $\phi = 1/3.4$ to control for the variance of the loss differential series.
We explore the power properties of various tests under two different alternative hypotheses. The first one is the homogeneous alternative and the second one is the heterogeneous alternative.

For DGP1 with homogeneous alternative, we generate a third set of forecast error series as $e_{3i,t} = \sqrt{1.2} e_{2i,t}$ and report the results from testing the equality of forecast accuracy of $e_{1i,t}$ and $e_{3i,t}$. In the heterogeneous scenario, we generate the third series according to $e_{3i,t} = \sqrt{\theta_i} e_{2i,t}$ where $\theta_i = 0.8$ for $i = 1, \ldots, n/2$ and $\theta_i = 1.2$ for $i = n/2+1, \ldots, n$. Similarly, in the case of DGP2, we set $\mu_i = 1.2$ for each $i$ in the case of homogeneous alternative and $\mu_i = -0.2$ for $i = 1, \ldots, n/2$ and $\mu_i = 0.2$ for $i = n/2 + 1, \ldots, n$ in the case of heterogeneous alternative. It is important to note that in the case of heterogeneous alternative, the unconditional expectations of the loss differentials are equal to zero in all DGPs. Hence, the overall EPA hypothesis holds. On the other hand, for each unit, the expected value of the loss differential is different from zero. Therefore, the joint EPA hypothesis does not hold. As a consequence, we expect the overall EPA tests not to have increasing power against the heterogeneous alternative whereas joint EPA tests to be consistent.

Three different spatial AR(1) parameters are considered for both DGPs: $\rho = 0, 0.5$ and 0.9. To save space we report results for only $\rho = 0.5$. As error series and common factors are generated for each unit as white noise, it is implicitly assumed that these are one-step ahead forecasts.

As we generate one-step ahead forecasts, the time series kernel $k_T(\cdot) = 1$ if $t = s$ and $k_T(\cdot) = 0$ otherwise. Spatial interactions between units are created with a rook-type weight matrix where two units in the panel are neighbors if their Euclidean distance is less than or equal to one. In the computation of the spatial kernel $k_S(\cdot)$, we used these distances. In addition, we use distances based on the wrong assumption that the units are located on a line. We use Bartlett kernel for all experiments and following KP, we set the spatial kernel bandwidth to $d_n = \lceil n^{1/4} \rceil$ where $\lceil \cdot \rceil$ stands for the smallest integer bigger than its argument. Similarly, the clustered test statistics using partial sample variance estimators are calculated by setting $n_1 = n_2 = \lceil d_n/2 \rceil$. For these tests, we reorder the cross-sectional units such that the units which are included in the calculation of the partial sample estimator appear at the beginning of the full sample. Hence, the variance estimates are calculated with $n = n_1 + n_2$.

For the tests using common factors, we consider three possibilities for the selection of number
of factors. First, we calculate them assuming \( m = 2 \), meaning that for DGP2 it is assumed that the number of common factors is correctly specified. Second, we assume that the number is under-specified, e.g. we set \( m = 1 \). Third, we select the number of common factors using an information criterion (IC) proposed by Bai and Ng (2002). The authors consider several information criteria. In their simulations \( IC_{p1} \) appears to be the best performing criterion under WCD. This criterion is given by

\[
IC_{p1} = V(m) + m \left( \frac{n + T}{nT} \right) \ln \left( \frac{nT}{n + T} \right)
\]

where \( V(m) \) is defined by (12). We select the number of common factors which minimizes this IC. The case of over-specification of the number of common factors is not separately considered because \( IC_{p1} \) almost always over-estimates the number of common factors in small samples (see Table 11 below).

Before the discussion of the size and power properties of the robust tests, as a benchmark we refer to the results on the non-robust tests \( S_{nT}^{(1)} \) and \( C_{nT}^{(1)} \). The size and power of these tests are reported in Table 1. As is expected, all tests are incorrectly sized.

### 4.2 Size properties

The results on the size properties of CD-robust overall EPA tests with DGP1 are given in Table 2. The size of the kernel robust test \( S_{nT}^{(2)} \) of the overall EPA hypothesis improves with either \( T \) or \( n \). First, we focus on the results when the distance metric is correctly specified. In the smallest samples with \( T = 10 \) and \( n = 10 \), this particular setting provides an empirical size of 9.9%. For \( T = 100 \) with \( n = 10 \) corresponding value equals 8.65% whereas for \( T = 10 \) with \( n = 100 \) it is 8.35%. In the largest sample its size is 6.25% which is close to the nominal value of 5%. When the distance between the panel units is incorrectly specified, the size of the test still improves with either dimension. However, as expected the size distortions are slightly larger in this case. In the largest smallest and largest sample sizes its size equals 12.5% and 8.3%, respectively. The test which uses the partial sample estimator of the variance has similar size values. However, its size improves only with \( T \). When \( T = n = 10 \) its size is slightly larger than that of the kernel robust test with a misspecified distance. When \( T, n = (10, 100) \) the size distortion increases (13.4%).
the case of large $T$, however, it performs better than the kernel robust test with either correct or incorrect distance. For instance, when $T = 100$ and $n = 10$ its size equals 6.25%.

The test $S^{(3)}_{nT}$ performs very well especially when $T$ is large and $n$ is small. In most of the combinations of $T$ and $n$ it shows better properties than $S^{(2)}_{nT}$. When $T$ is greater than 50, it is correctly sized except when $n = 100$. However, even when $T = 30$ and $n = 100$ its size is 6.95% which makes it the preferred test over any version of $S^{(2)}_{nT}$.

The test $S^{(3)}_{nT}$ shows good properties even though it wrongly assumes that the loss series contain common factors. The test with $m = 2$ performs similarly to $S^{(3)}_{nT}$ when $n$ is small but its performance is less good as $n$ gets large. In general size distortion of the test is bigger when $m = 1$.

The case where the number of common factors is chosen by IC requires some special attention. The test performs well for small $n$ and $T$ but its performance drops as $n$ or $T$ gets large. To understand this behavior of the test, we check the small sample properties of $IC_{p1}$. The average number of common factors chosen by this criterion over simulations is reported in Table 1. It is seen that, when either of the dimension of the panel grows, the performance of the IC increases. However, when one of the dimensions is small, this improvement is very slow. In DGP1, only when one of the dimensions is greater than 50, the performance is in acceptable levels. Once either $T$ or $n$ is greater than 50, the average number of factors selected over replications is either zero or very close to zero. This means, in fact in these cases the test converges to the non-robust test. Hence, the empirical size and power of the test equals the size and power of the non-robust test given in Table 1. Of course, in practice the number of common factors is rarely known to the researcher and the most realistic application of this test is this case which is based on IC. However, it should not be understood that our testing procedure is best described by the performance in this case. In practice, if the no CD hypothesis is rejected by a suitable test, and the researcher decides with the help of an IC that the loss differentials do not contain common factors, an EPA test which is robust to WCD has to be used; for instance $S^{(2)}_{nT}$ or $S^{(3)}_{nT}$. As in the case of zero common factors $S^{(3)}_{nT}$ is identical to $S^{(1)}_{nT}$, these tests have to be used only if the no CD hypothesis cannot be rejected before the application of EPA tests.

The results for the cluster based EPA tests are reported in the right block of Table 2. The kernel
robust test \( C_{nT}^{(2)} \) performs slightly worse than the overall test \( S_{nT}^{(2)} \) and its performance improves rapidly with increases in the number of observations in either dimensions of the panel. In the case of large samples the empirical sizes of the two tests are comparable. Similar to the findings on the overall test, when the distance is misspecified, the size distortion in the test is only slightly higher and its performance gets better with increases of number of observations. The performance of the test based on the partial sample estimator, namely \( C_{nT}^{(2)} \) is unsatisfactory in small samples, especially for small \( n \). For \( n = 10 \), even for the largest \( T \) we have, the empirical size of the test equals 19.35%. However its performance improves with \( T \), and it reaches 5.5% in the largest sample considered.

The test \( C_{nT}^{(3)} \) is a very viable alternative to \( C_{nT}^{(2)} \). For small \( T \), its size distortion is superior to that of \( C_{nT}^{(2)} \) but even for \( n = 20 \) it has good size properties. However, contrary to \( C_{nT}^{(2)} \), its performance improves only with \( T \). For instance, when \( n = 10 \) and \( T = 20 \), its size equals 9.85% which is better than that of the kernel robust test (10.65%). When \( n = 100 \) and \( T = 20 \) the performance of the latter improves dramatically and reaches to 6.25% whereas that of \( C_{nT}^{(3)} \) is 8.55%.

Finally we focus on the size of the test using estimated common factors, \( C_{nT}^{(3)} \). Irrespective of the choice of the number of common factors, when \( T \) is small, the test suffers from more size distortions compared to \( C_{nT}^{(2)} \) and \( C_{nT}^{(3)} \). However for large \( T \) and small \( n \) the test with \( m = 2 \) has a similar performance to the others. In fact for the smallest \( T \) and largest \( n \) that we consider it has an empirical size of 5.95% which is better than that of \( C_{nT}^{(2)} \) (9.65%) and \( C_{nT}^{(3)} \) (6.85%). When \( m = 2 \) the improvement of the size of the test is much slower with the increase in \( T \) and when the number of common factors is chosen by IC, it approaches the nonrobust test, as expected.

The size results for DGP2 are reported in Table 3. As expected, for this DGP the overall tests \( S_{nT}^{(2)} \) and \( S_{nT}^{(3)} \) are grossly over-sized and their performance does not improve with increases in the sample size in any dimension. The test \( S_{nT}^{(3)} \) shows very good properties except when \( T \) is very small, in particular when \( T > 30 \). Conclusions are similar for the factor-robust tests \( S_{nT}^{(3)} \). In fact, this test performs better than \( S_{nT}^{(3)} \) for all samples sizes considered. A very important finding is that, even when the number of common factors is underspecified, the test performs very well. We
also see that the three versions of the test are equivalent in large samples. For $T > 10$ and $n > 30$ these three tests have equal empirical size.

The findings concerning the cluster based joint EPA tests robust to SCD are similar to those in the case of DGP1 with a few points worth mentioning. The test $C^{(3)}_{nT}$ behaves in line with theoretical expectations such that it has lower size distortions for large $T$ and small $n$ and it performs slightly worse than the overall test $S^{(3)}_{nT}$. The tests based on estimated common factors are found to be oversized in small samples when the number of common factors are chosen by the researcher. The IC based version shows less size distortions. For small $T$ it outperforms $C^{(3)}_{nT}$ overall. Finally, in largest samples the IC based test and the test with $m = 2$ have identical size properties. This is expected because the information criterion $IC_{p1}$ consistently chooses the number of common factors when $n$ and $T$ are large, as seen in Table 11.

To summarize, in the case of both DGPs overall EPA hypothesis can be tested with a size close to the nominal value for almost all sample sizes. In particular, it is found that the test $S^{(3)}_{nT}$ has very good properties in both DGPs. For DGP1, for small $T$ and large $n$, the kernel robust test is preferred over the test based on the partial sample estimator given that the distance metric is correctly specified. Finally, the test $C^{(3)}_{nT}$ is preferred over others, however, $C^{(3)}_{nT}$ has also good properties when the number of common factors is overspecified.

4.3 Power properties

The power results of the tests for DGP1 under the homogeneous alternative hypothesis are given in Table 4. In the previous subsection, we have seen that the size of the overall EPA tests $S^{(2)}_{nT}$ and $S^{(3)}_{nT}$ approach to the nominal level for this DGP. Here, it is seen that the power of the tests $S^{(2)}_{nT}$ and $S^{(2)}_{nT}$ converge to 100%, for the latter the distance metric being unimportant. Hence the test is consistent in all cases. For moderate to large $T$, the test $S^{(3)}_{nT}$ is correctly sized. We observe that its power is only slightly lower compared to that of $S^{(2)}_{nT}$ in these sample sizes. Even though they wrongly assume that there are common factors in the DGP, the power of the factor-robust tests $C^{(3)}_{nT}$ are very close to that of $S^{(3)}_{nT}$.

The previous results of the cluster based joint EPA tests showed that in general they are
correctly sized only for large $T$. Here we see that their power is only slightly lower than the overall
tests. For instance, the power of the asymptotically correctly sized tests $S_{nT}^{(2)}$ and $C_{nT}^{(2)}$ are 19.05% and 17.30%, respectively and their power reaches 100% in the larges sample.

Table 5 reports the power results of the tests for DGP2 under the homogeneous alternative hypothesis. For this DGP, we have seen that the tests $S_{nT}^{(2)}$ are over-sized even asymptotically. Hence, we focus on the tests $S_{nT}^{(3)}$ and $S_{nT}^{(3)}$. It is seen that the power of both tests are very similar for all sample sizes. For instance, for most of the sample sizes that we consider, the power of $C_{nT}^{(3)}$ equals the power of $C_{nT}^{(3)}$ when the number of common factors is chosen by the IC.

To save space, we do not report the power results under the heterogeneous alternative hypothesis. The main finding is summarized in Figure 1 where the power of $S_{nT}^{(3)}$ and $C_{nT}^{(3)}$ are shown. It can be seen that under the homogeneous alternative, the power of both overall and cluster based joint EPA test approach to 100%. Whereas, under the heterogeneous alternative, the power of the overall test approaches to the nominal size. This is because under this alternative hypothesis the expected value of the loss differential equals zero. However, they are different from zero for clusters of panel units, hence the joint EPA test has power against this alternative.

## 5 Empirical Applications

The application of the panel EPA tests require some preliminary information on the cross-sectional and temporal dependence properties of the loss differentials. First, the researcher has to determine whether the loss series contain CD, to choose between the non-robust tests $S_{nT}^{(1)}$ and $C_{nT}^{(1)}$, and the other tests which deal with CD. If the data displays CD, one has to have some information on the type of the CD in loss differentials as WCD and SCD may require the use of different tests. Exceptions to this are $S_{nT}^{(3)}$ and $C_{nT}^{(3)}$ which are shown to be performing very well under any type of CD in our simulations. Second, to determine the time series kernel bandwidth parameter, one has to determine whether the loss differentials are autocorrelated or not. This panel EPA testing approach is based on the following three steps.

**Step A – Analysis of CD:** Test the no CD hypothesis using a test such as that of Breusch and Pagan (1980, BP hereafter) or the modified and standardized version of it Pesaran et al. (2008, Modified...
BP hereafter). If the no CD hypothesis is not rejected, proceed to Step B.1, otherwise calculate $IC_{p1}$ and proceed to Step B.2.

**Step B – Testing for no autocorrelation:** Consider the following empirical model for loss differentials:

$$
\Delta L_{it} = \pi_1 \Delta L_{i,t-1} + \pi_2 \Delta L_{i,t-2} + \cdots + \pi_p \Delta L_{i,t-p} + a_i + \zeta_{it}, 
$$

and the hypothesis $H_{0}^{ac}: \pi_1 = \pi_2 = \cdots = \pi_p$.

1. Run a fixed effects regression on (26) and test $H_{0}^{ac}$ using a Wald test with a variance robust to panel level heteroskedasticity.

2. Run a fixed effects regression on (26) and test $H_{0}^{ac}$ using a Wald test with a variance calculated by clustering on time index.

If the no autocorrelation hypothesis is not rejected, set $d_T = 1$, otherwise set $d_T > 1$. If the no CD hypothesis is not rejected in Step A, proceed to Step C.1, otherwise proceed to Step C.2.

**Step C – Testing for EPA:**

1. To test the overall EPA hypothesis $H_{0,1}$ use $S_{nT}^{(1)}$, and for the clustered EPA hypothesis $H_{0,2}$ use $C_{nT}^{(1)}$.

2. If $IC_{p1}$ indicates $m = 0$, use $S_{nT}^{(2)}$, $S_{nT}^{(2)}$, or $S_{nT}^{(3)}$ to test $H_{0,1}$, and $C_{nT}^{(2)}$, $C_{nT}^{(2)}$, or $C_{nT}^{(3)}$ to test $H_{0,2}$.

   If $IC_{p1}$ indicates $m > 0$, use $S_{nT}^{(3)}$ or $S_{nT}^{(3)}$ to test $H_{0,1}$, and $C_{nT}^{(3)}$ or $C_{nT}^{(3)}$ to test $H_{0,2}$.

In Step A, we suggest two tests of CD. The null hypothesis of BP test is the joint absence of CD between all pairs in the panel. The statistic is distributed as $\chi^2_q$ with $q = n(n-1)/2$ under the null. Hence, the test is more suitable for the cases of fixed and small $n$. The Modified BP statistic is a bias corrected and standardized version of the LM test. It is asymptotically normal under the null of no CD as $n \to \infty$ and is more suitable for large panels. These CD tests can be used to test the hypothesis of no CD in a data set but they do not help to identify the type of the CD. To see if the loss differentials contain common factors, we suggest the information criterion $IC_{p1}$ in (23). As seen in our simulations (Table 11), this IC performs quite well to choose between WCD and SCD.
In Step B, we analyze the serial correlation in loss differentials. The test of the no autocorrelation hypothesis follows the analysis of the CD properties of the loss differentials because the variance computed for the Wald test of the no autocorrelation hypothesis depends on whether the loss series contain CD or not. We suggest using a variance estimator calculated by clustering on the time index if the no CD hypothesis is not rejected in Step A. This estimation corresponds to the Driscoll and Kraay (1998) variance estimator with a time series kernel bandwidth chosen such that the error terms in (26) are assumed to be serially uncorrelated. For simplicity, here we do not distinguish between WCD and SCD, although this is possible by considering other variance estimators such as that of Kim and Sun (2013). However, as noted by Driscoll and Kraay (1998), their variance estimator is valid under WCD. One important point in Step B is the determination of the lag length \( p \). Han et al. (2017) found that the BIC is inconsistent in panel autoregressions even in the absence of fixed effects. We suggest to use the general to specific (GS) methodology that they propose. The method starts with \( p_{\text{max}} \) chosen by the researcher, and it continues by eliminating the biggest insignificant lag until a significant lag is found. If the significance level of these tests is fixed by the researcher, the probability of overestimation of the lag length is nonzero. To avoid this, we determine the significance level as \( \alpha_{nT} = \exp\{\ln(0.25)\sqrt{nT}/10\} \), as suggested by Han et al. (2017).

Finally, in Step C we test the EPA hypotheses \( H_{0,1} \) or \( H_{0,2} \) on the basis of the outcome of the two previous steps. In the next subsection, we follow this methodology.

5.1 Comparing OECD and IMF economic growth forecasts

We compare the OECD and IMF GDP growth forecasts using the EPA methodology described above. The data for the IMF forecasts come from their Historical WEO Forecasts Database. The database includes historical \( h \)-steps ahead forecast values, \( h = 1, 2, \ldots, 5 \), for GDP growth rate and covers up to 192 countries and starts from early 1990’s. We collected similar data from the past vintages of the Economic Outlook of the OECD. The Economic Outlook contains only 1-step ahead forecasts. Both organizations publish their forecasts twice a year. In our application we focus on their summer forecasts made for the following year. These forecasts are published in June every
year by the OECD whereas IMF forecasts are published in July. Publishing dates are close, hence the forecast errors are comparable. To compute the forecast errors, we used the respective GDP growth outturns published by each organization. Eventually we constructed a balanced panel data set of GDP growth forecast errors of 29 OECD countries from the two organization between 1998 and 2016.

Two different loss functions are used: absolute loss and quadratic loss. The absolute error loss differential is created as
\[ \Delta L_{it}^{(1)} = |e_{1,it}| - |e_{2,it}|, \]
where, as is throughout this application, first organization is the OECD. This loss function is important when we compare the magnitude of the absolute bias made by the two organizations. The quadratic loss is generated as
\[ \Delta L_{it}^{(2)} = e_{1,it}^2 - e_{2,it}^2. \]

This loss function is arguably the most frequently used one. If the forecasts of the both organizations are unbiased for each country, the expectation of absolute error loss is zero and quadratic loss permits to compare the variances.

We begin the analysis by the DM tests applied to each country. In the computations of the DM test statistics, we use a bandwidth parameter \( d_T = 1 \) because we have 1-step ahead forecasts. Note that below an autocorrelation test is used to confirm that the loss differentials are serially uncorrelated. The results of the DM tests are given in Table 6 where we report average loss differentials, DM test statistics and their \( p \)-values for each country over the period 1998-2016.

First, in terms of the sign of the average loss differentials, a considerable amount of heterogeneity can be observed in the sample. For both types of loss functions roughly half of the statistics are negative. Second, most of these statistics are statistically insignificant with exceptions being BEL, CAN, ESP, HUN and NZL. For BEL which is a country where the predictive ability of the IMF is superior, the EPA hypothesis can be rejected at 5% and 10% levels with absolute and quadratic losses, respectively. In the case of CAN, we can reject the EPA hypothesis with absolute loss at 10% significance level. For CAN too, the IMF predicts the economic growth rate better than the
OECD. In the case of ESP and HUN, the differences in predictive ability are significant with both absolute and quadratic losses. For ESP OECD predictions, for HUN IMF predictions outperform the other. For NZL we can reject the EPA hypothesis with absolute loss at 5% level.

Finally, at the bottom right of Table 6, we report average loss differentials over clusters of countries and the period 1998-2016. The average absolute loss differential over all 29 OECD countries in the sample is 0.003 whilst the average quadratic loss is equal to -0.049 for the same sample. This shows that, when we average over all countries, the difference between the forecast biases of the two institutions is positive but very small. Whereas, the average quadratic loss is negative.

An interesting question is whether there are clusters of countries for which the forecast performance of the two institutions differ dramatically. Dreher et al. (2008) test the hypothesis that the forecast performance of the IMF differs with respect to the direct influence of a country on the institution. They use GDP of a country as a proxy of the political influence and find evidence that the forecast bias of the institution declines with GDP. To see if we can find similar evidence in terms of the differences between the bias and efficiency of the forecasts made by the two institutions, we divide our country sample into the G7 and non-G7 countries. The GDP of the G7 countries account for almost 70% of the total GDP of all 29 countries in our sample in 2016, the last year of our data set. Hence, if the OECD’s forecast performance does not vary with a country’s GDP but the performance of the IMF does, as found by Dreher et al. (2008), we expect to have heterogeneity in average losses of the G7 and non-G7 countries.

The table shows that the average absolute loss and quadratic loss for G7 countries are 0.047 and 0.054, respectively. For non-G7 countries these averages are -0.011 and -0.081, respectively. This shows that for G7 countries the IMF has a superior performance whereas for non-G7 countries the OECD does better in terms of their growth forecasts. Hence, the forecast ability of the two institutions indeed varies with country clusters. Below, we test if these averages are statistically different from zero using our tests.

**Cross-sectional and temporal dependence in loss differentials.** As found in our Monte Carlo simulations, the increase in the number of cross-sections increases the power of EPA tests.
However, the gain from the usage of panels depends on the degree and the nature of CD. Furthermore, the application of the correct EPA tests require some information on the cross-sectional and temporal dependence of the loss differentials, as we summarized in the previous subsection. Hence, before proceeding to panel tests of EPA, we analyze the CD and autocorrelation properties of our dataset.

The results of BP and Modified BP tests are given in Table 7. As can be seen, the null hypothesis of no CD is rejected using any test on both loss functions in conventional significance levels. This means that the tests which allow for CD are more reliable for our data set. Next, $IC_{p_1}$ indicates existence of 6 common factors in both loss differential series. Hence, we conclude that both series display SCD. To determine the time series kernel bandwidth parameter, we estimate (26) with $p = 1$ which was chosen by the GS methodology and check the significance of the common autoregressive parameter by clustering in the time index. This autocorrelation test confirms that the loss differentials are serially uncorrelated, hence, we set $d_T = 1$.

To see the time series profile of the common factors in the loss differential series, we report the plot of the first 6 PCs in both loss differential series in panels (a) and (c) of Figure 2. The PCs are numbered in decreasing order with respect to their eigenvalues. For better interpretation of the common factor estimates, we report a focus on the first PC in the absolute loss differential in panel (b) of the figure. Similarly, panel (d) provides a focus on the first three PCs of the quadratic loss differential. Associated factor loadings estimates are reported in Table 8. To save space, estimated loadings for other PCs are dropped but they are available from the authors upon request.

The estimates of the common factors in loss differentials show the effect of financial crisis. The first common factor in absolute loss drops dramatically in 2008 and reaches a peak in 2010. As can be seen in Table 8, there are 13 countries in the sample for which the estimated loadings are positive. This shows that the first organization, the OECD, has systematically higher predictive ability compared to the IMF in 2008 for these countries whereas IMF has systematically better prediction performance for the recovery period.

A similar picture emerges for the quadratic loss differential. The first and second common factors in quadratic loss fall in 2009. For 18 countries in the sample, the factor loading estimates
are negative for the first common factor. Hence, the OECD, had a superior predictive ability compared to the IMF in this year. The second PC shows a similar pattern whereas the third PC has a movement in the opposite direction in the recovery period. This PC therefore compensates the effect of the first two common factors for some countries.

**Panel tests for the EPA hypotheses.** Here we apply the panel EPA tests to compare the performance of the two institutions. Following the insights of the Monte Carlo results and the CD analysis, we apply the factor-robust tests $S_{nT}^{(3)}$, $C_{nT}^{(3)}$, $S_{nT}^{(3)}$ and $C_{nT}^{(3)}$. As a benchmark, we also report the results from the tests assuming no CD, namely $S_{nT}^{(1)}$ and $C_{nT}^{(1)}$.

The results are given in Table 9. The first part of the table reports the tests of the overall EPA hypothesis. Hence, they are the significance tests of the overall differences reported in Table 6, which are 0.003 and -0.049 for absolute and quadratic loss differentials, respectively. The test statistics for the non-robust test $S_{nT}^{(1)}$ are 0.13 and -0.32, both of which are insignificant. The conclusion does not change using the robust tests for which the test statistics are slightly weaker.

In the second part of the table the results for the cluster-based tests are reported. The first set of clusters we consider is G7 countries and the non-G7 countries for which the average loss differentials are reported in Table 6. It is seen that for these clusters, the clustered EPA hypothesis cannot be rejected. Hence, the predictive ability differences of the two institutions are insignificant.

As a last exercise, we check if some clusters can be identified for which the differences are economically and statistically significant. To this end, we determine 2 clusters for each loss differentials using k-means algorithm (see Steinley, 2006, for instance). The number of clusters is chosen using the Bayesian Information Criterion of Bonhomme and Manresa (2015). The clusters of countries estimated by the algorithm and their associated average loss differentials are given in Table 10. As can be seen in Table 9, now for both loss functions we can reject the cluster based EPA hypothesis using any test.

In Appendix C, we report an alternative interesting empirical application to evaluate the consumer price inflation forecasts made by the IMF by comparing them with no change, i.e. random walk forecasts. We found significant evidence against the overall EPA hypothesis while comparing the IMF and random walk forecast, particularly strong in the pre-crisis period.
6 Conclusions

In this paper, we proposed novel predictive ability test for panels, corresponding to two different equal predictive ability hypotheses. Several overall EPA tests were developed to evaluate the hypothesis that the predictive ability of two forecasters is equal on average over all periods and units. Clustered EPA tests were also built which are able to test the hypothesis of whether two forecasters have equal predictive power for all clusters of units. Our proposed tests are robust to different forms of cross-sectional dependence in the loss differentials, arising from weak and strong cross-sectional dependence. The proposed tests are found to have appropriate size and power in a set of Monte Carlo simulations. In particular, the overall EPA tests robust to arbitrary cross-sectional dependence are correctly sized. Finally, we provided some useful three-step guideline on how to run the tests, which is then implemented in two applications. In the first application, we compare the prediction performance of two major organizations, the OECD and the IMF, on their historical economic growth forecasts. We found evidence of strong cross-sectional dependence in loss differentials of forecast errors. The results showed that there is minor differences between the predictive ability of the OECD and the IMF in terms of their economic growth forecasts.

Acknowledgements

We thank participants in the 18th International Workshop on Spatial Econometrics and Statistics (AgroParisTech, Paris, France, 23-24 May 2019), in particular the discussant Paul Elhorst and Davide Fiaschi, the 39th International Symposium on Forecasting (Thessaloniki, Greece, 16-19 June 2019), the 25th International Panel Data Conference (Vilnius University, Vilnius, Lithuania, 4-5 July 2019), and the 22nd Dynamic Econometrics Conference (Nuffield College, University of Oxford, Oxford, United Kingdom, 9-10 September 2019). The usual disclaimer applies. This paper is a part of Oguzhan Akgun’s PhD research project and it was developed while Giovanni Urga and Zhenlin Yang were visiting professors at University Paris II Panthéon-Assas - CRED (France) in 2019 and during Oguzhan Akgun and Alain Pirotte’s several visits at the Centre for Econometric Analysis (CEA) of Bayes Business School. We thank both institutions for financial support.
References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Breusch, T. S. and Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, 47(1):239–253.

Chudik, A., Pesaran, M. H., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *Econometrics Journal*, 14(1):45–90.

Clark, T. E. and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105(1):85–110.

Clark, T. E. and McCracken, M. W. (2015). Nested forecast model comparisons: A new approach to testing equal accuracy. *Journal of Econometrics*, 186(1):160–177.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Davies, A. and Lahiri, K. (1995). A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics*, 68(1):205–228.

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

Dreher, A., Marchesi, S., and Vreeland, J. R. (2008). The political economy of imf forecasts. *Public Choice*, 137(1-2):145–171.
Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially
dependent panel data. *Review of Economics and Statistics*, 80(4):549–560.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*,
74(6):1545–1578.

Han, C., Phillips, P. C., and Sul, D. (2017). Lag length selection in panel autoregression. *Econometric Reviews*, 36(1-3):225–240.

Hannan, E. J. (1970). *Multiple time series*. John Wiley & Sons.

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory*, 18(6):1449–1459.

Keane, M. P. and Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence
from panel data. *The American Economic Review*, pages 714–735.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure
for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of
Real Estate Finance and Economics*, 17(1):99–121.

Kelejian, H. H. and Prucha, I. R. (2007). HAC estimation in a spatial framework. *Journal of
Econometrics*, 140(1):131–154.

Kim, M. S. and Sun, Y. (2013). Heteroskedasticity and spatiotemporal dependence robust inference
for linear panel models with fixed effects. *Journal of Econometrics*, 177(1):85–108.

Mariano, R. S. and Preve, D. (2012). Statistical tests for multiple forecast comparison. *Journal of
econometrics*, 169(1):123–130.

Moscone, F. and Tosetti, E. (2012). HAC estimation in spatial panels. *Economics Letters*,
117(1):60–65.

Moscone, F. and Tosetti, E. (2015). Robust estimation under error cross section dependence. *Economics Letters*, 133:100–104.

Pain, N., Lewis, C., Dang, T.-T., Jin, Y., and Richardson, P. (2014). OECD forecasts during and
after the financial crisis. Working Paper 1107, OECD Economics Department.

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, 34(6-10):1089–1117.
Pesaran, M. H., Ullah, A., and Yamagata, T. (2008). A bias-adjusted LM test of error cross-section independence. *The Econometrics Journal*, 11(1):105–127.

Phillips, P. C. and Magdalinos, T. (2009). Unit root and cointegrating limit theory when initialization is in the infinite past. *Econometric Theory*, 25(6):1682–1715.

Phillips, P. C. and Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, pages 971–1001.

Pinelis, I. (1994). Optimum bounds for the distributions of martingales in banach spaces. *The Annals of Probability*, pages 1679–1706.

Steinley, D. (2006). K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1):1–34.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–1084.

White, H. (2001). *Asymptotic theory for econometricians*. Academic Press, Revised Edition.
Table 1: Small Sample Properties of the Non-Robust Tests $S_{n,T}^{(1)}$ and $C_{n,T}^{(1)}$

|                | Overall EPA Tests: $S_{n,T}^{(1)}$ | Cluster Based EPA Tests: $C_{n,T}^{(1)}$ |
|----------------|-----------------------------------|---------------------------------------|
| $n \backslash T$ | 10 13.55 13.05 12.10 12.70 13.40 | 10 15.40 16.00 14.50 14.75 14.80 |
|                | 20 13.05 12.50 12.15 10.45 11.45 | 20 14.75 14.95 14.05 12.50 13.20 |
|                | 30 11.40 11.70 11.85 11.40 10.50 | 30 14.05 12.45 12.60 11.95 11.05 |
|                | 50 14.30 11.60 10.35 11.10 10.10 | 50 16.00 12.90 11.35 12.05 12.20 |
|                | 100 13.50 11.05 12.75 12.40 11.40 | 100 12.90 10.60 11.60 10.85 10.15 |

DGP1: Size

|                | 10 48.00 47.30 44.85 42.90 48.45 | 10 41.90 40.60 38.35 34.80 42.10 |
|                | 20 62.15 62.20 60.40 58.40 56.80 | 20 56.30 55.70 55.35 52.65 51.30 |
|                | 30 67.65 64.25 65.05 65.55 65.35 | 30 63.20 59.90 60.65 61.15 59.95 |
|                | 50 72.20 72.35 72.00 74.15 73.00 | 50 69.25 68.90 67.60 69.95 68.55 |
|                | 100 79.90 80.75 81.90 79.20 80.00 | 100 76.40 75.85 77.35 75.55 76.50 |

DGP2: Size

|                | 10 24.15 33.05 38.15 54.80 76.25 | 10 23.60 29.75 34.20 49.30 71.70 |
|                | 20 32.00 49.20 61.35 78.30 95.90 | 20 28.70 44.75 57.30 73.20 93.45 |
|                | 30 40.95 61.10 74.65 89.50 99.35 | 30 38.20 53.95 67.70 85.80 98.65 |
|                | 50 55.50 78.25 91.05 98.50 100.00 | 50 51.65 73.05 87.25 97.55 100.00 |
|                | 100 79.85 96.35 99.30 100.00 100.00 | 100 73.00 94.25 98.65 99.95 100.00 |

DGP1: Power

|                | 10 95.40 99.60 100.00 100.00 100.00 | 10 95.55 99.60 100.00 100.00 100.00 |
|                | 20 98.15 99.80 100.00 100.00 100.00 | 20 98.45 99.90 100.00 100.00 100.00 |
|                | 30 98.75 99.95 100.00 100.00 100.00 | 30 99.05 99.95 100.00 100.00 100.00 |
|                | 50 99.05 100.00 100.00 100.00 100.00 | 50 99.55 100.00 100.00 100.00 100.00 |
|                | 100 99.50 100.00 100.00 100.00 100.00 | 100 100.00 100.00 100.00 100.00 |

DGP2: Power

Note: Overall EPA Tests are introduced in Section 3.1 and Cluster Based EPA Tests are in Section 3.2. The nominal size is 5%. Power is calculated under homogeneous alternative hypothesis.
Table 2: Size - DGP1: No Common Factors, Spatial Dependence

|                | Overall EPA Tests | Cluster Based EPA Tests |
|----------------|------------------|------------------------|
|                | \(n \setminus T\) | \(n \setminus T\)      |
|                | 10               | 10                     |
|                | 20               | 20                     |
|                | 30               | 30                     |
|                | 50               | 50                     |
|                | 100              | 100                    |

\[S_{n,T}^{(2)}\]

|                | \(S_{n,T}^{(2)}[ms]\) | \(S_{n,T}^{(2)}[ms]\) |
|----------------|------------------------|------------------------|
|                | \(S_{n,T}^{(2)}[m = 2]\) | \(S_{n,T}^{(2)}[m = 1]\) |
|                | \(S_{n,T}^{(3)}[ic]\) | \(S_{n,T}^{(3)}[ic]\) |

|                | \(C_{n,T}^{(2)}\) | \(C_{n,T}^{(2)}\) |
|----------------|------------------|------------------|
|                | \(C_{n,T}^{(3)}[m = 2]\) | \(C_{n,T}^{(3)}[m = 1]\) |
|                | \(C_{n,T}^{(3)}[ic]\) | \(C_{n,T}^{(3)}[ic]\) |

Note: See the note of Table 1. [\(ms\)] indicates that the test uses a misspecified distance metric. For \(S_{n,T}^{(3)}\), the number of common factors is shown in brackets. [\(ic\)] means that the number of common factors is chosen by IC.
## Table 3: Size - DGP2: Common Factors, Spatial Dependence

| Overall EPA Tests | Cluster Based EPA Tests |
|-------------------|-------------------------|
| $n \backslash T$   | $n \backslash T$        |
|                   | 10  | 20  | 30  | 50  | 100 |      | 10  | 20  | 30  | 50  | 100 |      |
|                   |     |     |     |     |     |      |     |     |     |     |     |      |
| $S_{n,T}^{(2)}$   |     |     |     |     |     |      |     |     |     |     |     |      |
| 10                | 30.70 | 30.50 | 26.60 | 24.70 | 29.60 |     |     |     |     |     |     |
| 20                | 29.95 | 29.35 | 28.05 | 25.90 | 25.00 |     |     |     |     |     |     |
| 30                | 37.25 | 34.65 | 35.55 | 34.75 | 32.60 |     |     |     |     |     |     |
| 50                | 47.35 | 45.35 | 42.90 | 44.35 | 42.70 |     |     |     |     |     |     |
| 100               | 59.25 | 59.10 | 57.95 | 56.25 | 57.95 |     |     |     |     |     |     |
| $S_{n,T}^{(2)}[ms]$ |     |     |     |     |     |      |     |     |     |     |     |      |
| 10                | 37.65 | 36.05 | 33.50 | 31.30 | 38.05 |     |     |     |     |     |     |
| 20                | 45.50 | 44.85 | 43.45 | 41.35 | 40.25 |     |     |     |     |     |     |
| 30                | 52.75 | 49.50 | 51.15 | 50.80 | 50.00 |     |     |     |     |     |     |
| 50                | 60.85 | 59.30 | 58.90 | 60.20 | 59.80 |     |     |     |     |     |     |
| 100               | 68.15 | 67.25 | 67.55 | 66.40 | 67.35 |     |     |     |     |     |     |
| $S_{n,T}^{(3)}$   |     |     |     |     |     |      |     |     |     |     |     |      |
| 10                | 10.15 | 8.45  | 6.00  | 6.30  | 5.65  |     |     |     |     |     |     |
| 20                | 9.35  | 7.35  | 6.00  | 5.20  | 5.85  |     |     |     |     |     |     |
| 30                | 10.25 | 6.50  | 7.70  | 5.75  | 5.40  |     |     |     |     |     |     |
| 50                | 8.65  | 6.85  | 6.50  | 5.45  | 4.95  |     |     |     |     |     |     |
| 100               | 8.45  | 7.05  | 5.25  | 5.65  | 4.70  |     |     |     |     |     |     |
| $S_{n,T}^{(3)}[m = 2]$ |     |     |     |     |     |      |     |     |     |     |     |      |
| 10                | 9.90  | 8.00  | 5.60  | 6.00  | 5.25  |     |     |     |     |     |     |
| 20                | 9.25  | 7.25  | 6.30  | 5.00  | 5.35  |     |     |     |     |     |     |
| 30                | 10.10 | 6.45  | 7.65  | 5.50  | 5.35  |     |     |     |     |     |     |
| 50                | 8.55  | 6.80  | 6.35  | 5.40  | 4.85  |     |     |     |     |     |     |
| 100               | 8.50  | 7.05  | 5.20  | 5.65  | 4.60  |     |     |     |     |     |     |
| $S_{n,T}^{(3)}[m = 1]$ |     |     |     |     |     |      |     |     |     |     |     |      |
| 10                | 9.75  | 7.65  | 5.55  | 5.75  | 5.15  |     |     |     |     |     |     |
| 20                | 9.25  | 7.20  | 6.20  | 4.90  | 5.30  |     |     |     |     |     |     |
| 30                | 10.10 | 6.40  | 7.60  | 5.50  | 5.30  |     |     |     |     |     |     |
| 50                | 8.60  | 6.65  | 6.35  | 5.40  | 4.85  |     |     |     |     |     |     |
| 100               | 8.55  | 7.05  | 5.20  | 5.65  | 4.60  |     |     |     |     |     |     |
| $S_{n,T}^{(3)}[ic]$ |     |     |     |     |     |      |     |     |     |     |     |      |
| 10                | 10.10 | 8.30  | 5.95  | 6.20  | 5.50  |     |     |     |     |     |     |
| 20                | 9.25  | 7.25  | 6.40  | 5.05  | 5.50  |     |     |     |     |     |     |
| 30                | 10.15 | 6.45  | 7.65  | 5.50  | 5.35  |     |     |     |     |     |     |
| 50                | 8.80  | 6.80  | 6.35  | 5.40  | 4.85  |     |     |     |     |     |     |
| 100               | 8.45  | 7.05  | 5.20  | 5.65  | 4.60  |     |     |     |     |     |     |

$S_{n,T}^{(2)}$, $S_{n,T}^{(3)}$, $C_{n,T}^{(2)}$, $C_{n,T}^{(3)}$, $C_{n,T}^{(3)}[m = 2]$, $C_{n,T}^{(3)}[m = 1]$, $C_{n,T}^{(3)}[ic]$.

Note: See the note of Table 2.
|                  | Overall EPA Tests                  | Cluster Based EPA Tests                  |
|------------------|------------------------------------|------------------------------------------|
|                  | $S_{n,T}^{(2)}$                     | $C_{n,T}^{(2)}$                           |
|                  | $n \backslash T$ | 10 | 20 | 30 | 50 | 100 |                      | $n \backslash T$ | 10 | 20 | 30 | 50 | 100 |                      |
| $S_{n,T}^{(2)}$  | 10 | 19.05 | 26.60 | 30.80 | 47.60 | 70.80 |                      | 10 | 17.30 | 22.25 | 26.80 | 40.30 | 64.05 |
|                  | 20 | 24.55 | 40.05 | 53.05 | 71.15 | 93.20 |                      | 20 | 21.70 | 34.30 | 45.65 | 63.70 | 88.40 |
|                  | 30 | 33.55 | 51.85 | 67.20 | 85.15 | 98.85 |                      | 30 | 29.50 | 44.60 | 57.75 | 78.35 | 97.30 |
|                  | 50 | 48.55 | 71.60 | 87.80 | 97.50 | 100.00 |                      | 50 | 42.80 | 63.50 | 81.10 | 94.80 | 100.00 |
|                  | 100 | 71.15 | 94.05 | 98.40 | 99.95 | 100.00 |                      | 100 | 63.40 | 89.80 | 96.75 | 99.95 | 100.00 |
| $S_{n,T}^{(2)}[ms]$ | 10 | 21.80 | 29.85 | 34.90 | 51.60 | 73.80 |                      | 10 | 20.00 | 26.10 | 30.55 | 45.10 | 68.25 |
|                  | 20 | 28.65 | 45.30 | 57.30 | 75.10 | 95.00 |                      | 20 | 24.85 | 39.55 | 51.90 | 68.50 | 91.05 |
|                  | 30 | 37.60 | 56.25 | 71.15 | 87.85 | 99.25 |                      | 30 | 33.90 | 49.60 | 62.80 | 82.35 | 98.30 |
|                  | 50 | 48.55 | 71.60 | 87.80 | 97.50 | 100.00 |                      | 50 | 47.00 | 67.95 | 84.55 | 96.15 | 100.00 |
|                  | 100 | 71.15 | 94.05 | 98.40 | 99.95 | 100.00 |                      | 100 | 67.35 | 91.75 | 97.75 | 99.95 | 100.00 |
| $S_{n,T}^{(3)}$  | 10 | 16.45 | 21.70 | 25.35 | 39.10 | 63.50 |                      | 10 | 18.45 | 19.90 | 22.30 | 31.95 | 53.50 |
|                  | 20 | 23.75 | 37.25 | 48.85 | 66.20 | 90.40 |                      | 20 | 26.00 | 32.45 | 42.25 | 57.80 | 84.90 |
|                  | 30 | 31.80 | 48.80 | 63.00 | 82.15 | 98.25 |                      | 30 | 33.05 | 42.90 | 54.25 | 71.10 | 92.00 |
|                  | 50 | 47.00 | 67.65 | 84.25 | 96.25 | 100.00 |                      | 50 | 45.45 | 60.70 | 78.40 | 92.80 | 99.95 |
|                  | 100 | 68.50 | 92.20 | 97.80 | 99.90 | 100.00 |                      | 100 | 66.60 | 88.40 | 95.75 | 99.80 | 100.00 |
| $S_{n,T}^{(3)}[m = 2]$ | 10 | 17.40 | 20.75 | 25.10 | 35.95 | 57.15 |                      | 10 | 21.00 | 21.80 | 23.50 | 31.00 | 47.40 |
|                  | 20 | 25.50 | 38.55 | 50.40 | 67.10 | 90.00 |                      | 20 | 28.00 | 37.75 | 46.30 | 61.40 | 84.85 |
|                  | 30 | 34.25 | 52.25 | 66.85 | 84.05 | 98.05 |                      | 30 | 34.35 | 47.80 | 58.80 | 79.00 | 96.85 |
|                  | 50 | 50.55 | 72.45 | 85.80 | 97.10 | 99.95 |                      | 50 | 48.70 | 67.60 | 81.90 | 95.00 | 100.00 |
|                  | 100 | 73.25 | 94.45 | 98.45 | 99.95 | 100.00 |                      | 100 | 66.60 | 88.40 | 95.75 | 99.80 | 100.00 |
| $S_{n,T}^{(3)}[m = 1]$ | 10 | 16.25 | 21.00 | 23.80 | 36.75 | 64.85 |                      | 10 | 19.05 | 19.70 | 20.90 | 30.60 | 56.35 |
|                  | 20 | 23.80 | 38.65 | 56.00 | 76.30 | 95.80 |                      | 20 | 26.55 | 35.30 | 51.55 | 71.10 | 93.30 |
|                  | 30 | 31.55 | 56.00 | 73.10 | 89.00 | 99.35 |                      | 30 | 33.50 | 49.85 | 65.70 | 85.85 | 98.65 |
|                  | 50 | 48.20 | 71.75 | 87.10 | 97.55 | 100.00 |                      | 50 | 47.00 | 71.75 | 87.10 | 97.55 | 100.00 |
|                  | 100 | 71.75 | 96.35 | 99.30 | 100.00 | 100.00 |                      | 100 | 68.00 | 94.20 | 98.60 | 99.95 | 100.00 |

Note: See the note of Table 2.
Table 5: Power Under Homogeneous Alternative – DGP 2: Common Factors, Spatial Dependence

|                | Overall EPA Tests |                               | Cluster Based EPA Tests |                               |
|----------------|-------------------|--------------------------------|-------------------------|--------------------------------|
|                |                   |                                |                         |                                |
| $n \backslash T$ |                   |                                |                         |                                |
| 10             |                   |                                |                         |                                |
| 20             |                   |                                |                         |                                |
| 30             |                   |                                |                         |                                |
| 50             |                   |                                |                         |                                |
| 100            |                   |                                |                         |                                |
| $S^{(2)}_{n,T}$|                   |                                |                         |                                |
| 10             | 91.80             | 98.90                          | 99.85                   | 100.00                         | 100.00|
| 20             | 92.75             | 99.20                          | 99.95                   | 100.00                         | 100.00|
| 30             | 95.60             | 99.50                          | 100.00                  | 100.00                         | 100.00|
| 50             | 97.10             | 99.85                          | 100.00                  | 100.00                         | 100.00|
| 100            | 98.45             | 99.95                          | 100.00                  | 100.00                         | 100.00|
| $S^{(2)}_{n,T}[m]$|                |                                |                         |                                |
| 10             | 93.80             | 99.30                          | 99.95                   | 100.00                         | 100.00|
| 20             | 96.30             | 99.55                          | 100.00                  | 100.00                         | 100.00|
| 30             | 97.60             | 99.75                          | 100.00                  | 100.00                         | 100.00|
| 50             | 98.25             | 99.95                          | 100.00                  | 100.00                         | 100.00|
| 100            | 99.15             | 100.00                         | 100.00                  | 100.00                         | 100.00|
| $S^{(3)}_{n,T}$|                   |                                |                         |                                |
| 10             | 74.45             | 93.05                          | 98.95                   | 99.95                          | 100.00|
| 20             | 74.90             | 94.50                          | 99.35                   | 100.00                         | 100.00|
| 30             | 77.65             | 95.60                          | 99.25                   | 100.00                         | 100.00|
| 50             | 78.10             | 95.35                          | 99.45                   | 99.95                          | 100.00|
| 100            | 78.95             | 96.60                          | 99.65                   | 100.00                         | 100.00|
| $S^{(3)}_{n,T}[m = 2]$|         |                                |                         |                                |
| 10             | 73.75             | 92.80                          | 98.90                   | 99.95                          | 100.00|
| 20             | 74.55             | 94.55                          | 99.30                   | 100.00                         | 100.00|
| 30             | 77.40             | 95.55                          | 99.20                   | 100.00                         | 100.00|
| 50             | 77.90             | 95.35                          | 99.45                   | 99.95                          | 100.00|
| 100            | 78.95             | 96.65                          | 99.65                   | 100.00                         | 100.00|
| $S^{(3)}_{n,T}[m = 1]$|         |                                |                         |                                |
| 10             | 73.60             | 92.55                          | 98.90                   | 99.95                          | 100.00|
| 20             | 74.20             | 94.50                          | 99.25                   | 100.00                         | 100.00|
| 30             | 77.35             | 95.55                          | 99.20                   | 100.00                         | 100.00|
| 50             | 77.75             | 95.35                          | 99.45                   | 99.95                          | 100.00|
| 100            | 78.95             | 96.55                          | 99.65                   | 100.00                         | 100.00|
| $S^{(3)}_{n,T}[ic]$|               |                                |                         |                                |
| 10             | 74.30             | 93.00                          | 98.95                   | 99.95                          | 100.00|
| 20             | 74.85             | 94.65                          | 99.35                   | 100.00                         | 100.00|
| 30             | 77.55             | 95.55                          | 99.20                   | 100.00                         | 100.00|
| 50             | 78.05             | 95.35                          | 99.45                   | 99.95                          | 100.00|
| 100            | 78.95             | 96.65                          | 99.65                   | 100.00                         | 100.00|

Note: See the note of Table 2.
Figure 1: Power of Selected Tests Under Different Alternative Hypotheses for DGP2 (5% Nominal Size)
Table 6: Average Loss Differentials for the Economic Growth Forecasts, DM Test Statistics and Their p-values, 1998-2016 (OECD vs. IMF)

| Country | Absolute Loss | Quadratic Loss | Country | Absolute Loss | Quadratic Loss |
|---------|---------------|----------------|---------|---------------|----------------|
| AUS     | -0.035        | -0.047         | ISL     | -0.092        | -0.461         |
|         | (-0.616)      | (-0.405)       |         | (-0.533)      | (-0.471)       |
|         | [0.538]       | [0.686]        |         | [0.594]       | [0.638]        |
| AUT     | 0.041         | 0.021          | ITA†    | 0.099         | 0.602          |
|         | (0.689)       | (0.148)        |         | (1.070)       | (1.171)        |
|         | [0.491]       | [0.882]        |         | [0.285]       | [0.242]        |
| BEL     | **0.137**     | **0.544**      | JPN†    | 0.097         | 0.190          |
|         | (2.014)       | (1.663)        |         | (1.423)       | (1.035)        |
|         | [0.044]       | [0.096]        |         | [0.155]       | [0.301]        |
| CAN†    | **0.096**     | 0.273          | KOR     | 0.087         | 0.563          |
|         | (1.783)       | (1.501)        |         | (0.598)       | (0.556)        |
|         | [0.075]       | [0.133]        | LUX     | 0.155         | 1.358          |
|         | (1.046)       | (1.098)        |         | (0.925)       | (1.014)        |
|         | [0.295]       | [0.272]        |         | [0.355]       | [0.311]        |
| CZE     | -0.229        | -1.280         | MEX     | -0.078        | -0.386         |
|         | (-1.062)      | (-1.000)       |         | (-0.520)      | (-0.382)       |
|         | [0.288]       | [0.317]        |         | [0.603]       | [0.703]        |
| DEU†    | -0.035        | -0.633         | NLD     | 0.006         | 0.203          |
|         | (-0.369)      | (-1.311)       |         | (0.081)       | (0.871)        |
|         | [0.712]       | [0.258]        |         | [0.935]       | [0.384]        |
| DNK     | 0.003         | -0.138         | NOR     | 0.001         | -0.169         |
|         | (0.045)       | (-0.703)       |         | (0.008)       | (-0.628)       |
|         | [0.965]       | [0.482]        |         | [0.993]       | [0.530]        |
| ESP     | **-0.155**    | **-0.630**     | NZL     | **-0.166**    | **-0.481**     |
|         | (-1.696)      | (-1.692)       |         | (-2.073)      | (-1.535)       |
|         | [0.090]       | [0.091]        |         | [0.038]       | [0.125]        |
| FIN     | 0.062         | 0.099          | POL     | -0.055        | -0.523         |
|         | (0.424)       | (0.125)        |         | (-0.447)      | (-0.960)       |
|         | [0.672]       | [0.900]        |         | [0.655]       | [0.337]        |
| FRA†    | 0.064         | 0.140          | PRT     | -0.007        | 0.062          |
|         | (1.421)       | (1.451)        |         | (-0.068)      | (0.127)        |
|         | [0.156]       | [0.147]        |         | [0.946]       | [0.899]        |
| GBR†    | -0.014        | -0.196         | SWE     | -0.083        | -0.105         |
|         | (-0.244)      | (-1.123)       |         | (-0.661)      | (-0.164)       |
|         | [0.808]       | [0.261]        |         | [0.509]       | [0.870]        |
| GRC     | -0.207        | -2.090         | TUR     | -0.017        | -0.707         |
|         | (-1.071)      | (-1.451)       |         | (-0.074)      | (-0.302)       |
|         | [0.284]       | [0.147]        |         | [0.941]       | [0.763]        |
| HUN     | **0.235**     | **1.216**      | USA†    | 0.018         | 0.003          |
|         | (2.387)       | (1.874)        |         | (0.201)       | (0.008)        |
|         | [0.017]       | [0.061]        |         | [0.841]       | [0.994]        |
| IRL     | 0.084         | 0.890          |         | Average: All  | 0.003          |
|         | (0.472)       | (0.656)        |         | Average: G7   | 0.047          |
|         | [0.637]       | [0.512]        |         | Average: Non-G7| -0.011         |

Note: † G7 countries. DM statistics in parentheses are calculated as $S_{i,T}^{(0)} = \sqrt{T}(\Delta \bar{L}_{i,T}/\hat{\sigma}_{i,T}) \overset{d}{\rightarrow} N(0,1)$ where $\hat{\sigma}_{i,T}^{2} = \frac{1}{T} \sum_{t=1}^{T} \Delta \tilde{L}_{it}^{2}$. Differences significant at 10% are shown in bold. p-values are in brackets.
Table 7: CD Tests of the Economic Growth Forecasts (OECD vs. IMF)

|                  | Absolute Loss | Quadratic Loss |
|------------------|---------------|---------------|
| BP Test          | 478.103       | 856.275       |
|                  | (0.008)       | (0.000)       |
| Modified BP Test | 2.745         | 17.139        |
|                  | (0.006)       | (0.000)       |

Note: The values shown in parentheses are \( p \)-values.

Table 8: PC Estimates of the Factor Loadings in the Loss Differentials of the Economic Growth Forecasts (OECD vs. IMF)

| Country | \( PC1 \) | \( PC1 \) | \( PC2 \) | \( PC3 \) |
|---------|---------|---------|---------|---------|
| AUS     | -0.08   | 0.07    | -0.10   | 0.32    |
| AUT     | 0.25    | 0.04    | 0.32    | -0.20   |
| BEL     | 0.30    | 0.13    | 0.17    | -1.01   |
| CAN     | 0.23    | 0.20    | 0.17    | -0.28   |
| CHE     | -0.14   | -0.13   | -0.45   | -0.11   |
| CZE     | -2.84   | 0.27    | -3.22   | -2.12   |
| DEU     | -0.94   | -0.79   | -1.51   | 1.19    |
| DNK     | 0.23    | 0.26    | 0.21    | 0.15    |
| ESP     | 0.78    | 0.41    | 0.93    | 0.52    |
| FIN     | -1.72   | -0.85   | -2.55   | 1.51    |
| FRA     | 0.19    | -0.05   | 0.03    | -0.01   |
| GBR     | 0.19    | -0.02   | 0.21    | 0.52    |
| GRC     | 0.90    | -2.19   | 2.99    | -3.72   |
| HUN     | 0.62    | -0.47   | -0.31   | -1.78   |
| IRL     | -2.41   | -2.26   | -4.35   | -1.07   |
| ISL     | -1.61   | 0.51    | -1.39   | 0.15    |
| ITA     | -0.75   | -0.80   | -1.31   | -0.18   |
| JPN     | -0.43   | 0.07    | -0.41   | 0.01    |
| KOR     | -1.27   | -2.23   | -2.09   | 2.46    |
| LUX     | -1.15   | -1.54   | -2.66   | -3.20   |
| MEX     | -1.25   | -2.19   | -2.73   | 0.62    |
| NLD     | -0.37   | -0.05   | -0.56   | -0.01   |
| NOR     | -0.12   | 0.52    | 0.35    | 0.49    |
| NZL     | 0.50    | 0.35    | 0.34    | 0.78    |
| POL     | 0.21    | -0.98   | 0.64    | -0.77   |
| PRT     | 0.12    | -0.42   | -0.18   | -0.14   |
| SWE     | -0.55   | -0.60   | -1.54   | -0.73   |
| TUR     | 0.39    | -9.77   | 2.29    | 0.88    |
| USA     | -0.60   | -0.25   | -0.52   | 0.81    |
Table 9: Panel Tests of EPA for the Economic Growth Forecasts (OECD vs. IMF)

| Test  | Overall Tests | Clustered Tests |
|-------|---------------|-----------------|
|       | Absolute Loss | Quadratic Loss  | Absolute Loss | Quadratic Loss |
| $S_{n,T}^{(1)}$ | 0.13          | -0.32           | $C_{n,T}^{(1)}$ | 2.90          |
|        | (0.89)        | (0.75)          |                  | (0.23)        |
| $S_{n,T}^{(3)}$ | 0.11          | -0.21           | $C_{n,T}^{(3)}$ | 2.48          |
|        | (0.92)        | (0.84)          |                  | (0.29)        |
| $C_{n,T}^{(3)}$ | 0.11          | -0.21           |                  | 2.04          |
|        | (0.92)        | (0.84)          |                  | (0.36)        |

Note: The values shown in parentheses are p-values.

Table 10: Country Clusters Selected by $k$-means for the Economic Growth Forecasts (OECD vs. IMF)

| Clusters | Average Loss Differential | Countries |
|----------|----------------------------|-----------|
|          | Absolute Loss              |           |
| Cluster 1| 0.06                       | AUT, BEL, CAN, CHE, DNK, FIN, FRA, GBR, HUN, IRL, ITA, JPN, KOR, LUX, NLD, NOR, PRT, TUR, USA |
| Cluster 2| -0.11                      | AUS, CZE, DEU, ESP, GRC, ISL, MEX, NZL, POL, SWE |

|          | Quadratic Loss             |           |
| Cluster 1| 0.29                       | AUS, AUT, BEL, CAN, CHE, DNK, FIN, FRA, GBR, HUN, IRL, ITA, JPN, KOR, LUX, NLD, NOR, PRT, SWE, USA |
| Cluster 2| -0.80                      | CZE, DEU, ESP, GRC, ISL, MEX, NZL, POL, TUR |

Table 11: Small Sample Properties of the Information Criterion $IC_{p1}$

| $n \setminus T$ | DGP1          | DGP2          |
|-----------------|---------------|---------------|
|                 | 10 20 30 50 100 | 10 20 30 50 100 |
| 10              | 5.00 5.00 4.95 4.57 2.57 | 5.00 5.00 5.00 5.00 5.00 |
| 20              | 5.00 3.88 1.09 0.20 0.01 | 5.00 5.52 4.48 3.73 3.28 |
| 30              | 4.99 1.12 0.25 0.06 0.00 | 5.00 4.07 2.80 2.45 2.28 |
| 50              | 4.94 0.15 0.05 0.01 0.00 | 5.00 4.25 2.08 2.04 2.01 |
| 100             | 3.62 0.01 0.00 0.00 0.00 | 4.98 1.94 1.99 2.00 2.00 |
Figure 2: PC Estimates of the Common Factors in the Loss Differentials of the Economic Growth Forecasts (OECD vs. IMF)
Appendances

Appendix A Loss Differentials Specification: A Justification

In this appendix, we present the derivation of the model for the loss differentials starting from a pure common factor model for the forecast errors assuming two most commonly used loss functions, namely absolute loss and quadratic loss. The forecast error $e_{l,it}$ of the forecaster $l = 1, 2$, is assumed to be given by

$$e_{l,it} = c_{l_i} + \theta'_{l_i} g_{lt} + u_{l,it}, \quad i = 1, 2, \ldots, n, \quad t = 1, 2, \ldots, T,$$

where $g_{lt} = (g_{l1,t}, \ldots, g_{lm_l,t})'$ is an $m_l \times 1$ vector of common factors, $\theta_{l_i} = (\theta_{l_{1,i}}, \ldots, \theta_{l_{m_l,i}})'$ is their respective factor loadings vector, and $u_{l,it}$ is an error term which can in general be serially and cross-sectionally weakly correlated. Since our objective in this section is to demonstrate the validity of the common factor model for the loss differential, we will not focus on the weak-dependence in the errors. We assume that $E(u_{l,it} | g_t) = 0$ and $E(u_{l,it}^2 | g_t) = \sigma^2_{l_i}$ for $l = 1, 2$ where $g_t = (g'_{1t}, g'_{2t})'$.

The absolute loss differential is given by $d_{1,it} = |e_{1,it}| - |e_{2,it}|$. Then, we have $E(d_{1,it} | g_t) = \mu_i + \lambda' f_t$ where $\lambda_i = (\theta'_{1i}, \theta'_{2i})'$, $\mu_i = E[\text{sign}(e_{1,it})]c_{1i} - E[\text{sign}(e_{2,it})]c_{2i}$ and

$$f_t' = \begin{cases} 
(g'_{1t}, -g'_{2t})', & \text{if } e_{1,it} \geq 0 \text{ and } e_{2,it} \geq 0, \\
(g'_{1t}, g'_{2t})', & \text{if } e_{1,it} \geq 0 \text{ and } e_{2,it} < 0, \\
(-g'_{1t}, g'_{2t})', & \text{if } e_{1,it} < 0 \text{ and } e_{2,it} \geq 0, \\
(-g'_{1t}, -g'_{2t})', & \text{if } e_{1,it} < 0 \text{ and } e_{2,it} < 0.
\end{cases}$$

Hence, the model for the loss differentials is obtained with $m = m_1 + m_2$ in the case of absolute loss function. The quadratic loss differential is defined as $d_{2,it} = e_{1,it}^2 - e_{2,it}^2$. To simplify the notation, let
us assume that \( c_{li} = 0 \) for each \( l = 1, 2 \). The squared forecast error then satisfies

\[
e_{t,it}^2 = \theta'_{li} g_i g_{it}' \theta_{li} + 2 \theta'_{li} g_i u_{i,t} + u_{i,t}^2
\]

\[
= \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_1} \theta_{li,k_1} \theta_{li,k_2} g_{k_1,t} g_{k_2,t} + 2 \theta'_{li} g_i u_{i,t} + u_{i,t}^2
\]

\[
= \sum_{k=1}^{m_1^2} \gamma_{li,k} h_{i,k,t} + 2 \theta'_{li} g_i u_{i,t} + u_{i,t}^2
\]

\[
= \gamma'_{li} h_{i,t} + 2 \theta'_{li} g_i u_{i,t} + u_{i,t}^2,
\]

with straightforward definitions of \( \gamma_{li,k} \) and \( h_{i,k,t} \) which are the \( k \)th elements of \( m_1^2 \times 1 \) vectors \( \gamma_{li} \) and \( h_{i,t} \), respectively. Then, we obtain the model for the loss differentials as \( E(d_{2,it}|g_t) = \mu_i + \lambda_i' f_t \) with \( \mu_i = \sigma_1^2 - \sigma_2^2 \), \( \lambda_i = (\gamma_{li,1}, \gamma_{li,2})' \), \( f_t = (h_{i,t}, h_{2t})' \) and \( m = m_1^2 + m_2^2 \).

### Appendix B  Proofs

In this appendix, we present the proofs of Theorems 1 and 2. Propositions 1-4 are special cases of these theorems. Let \( H_n \) be an \( n \times G \) matrix which has \( h_{g,i}' \), the \( g \)th row with \( g_i \in \{1, 2, \ldots, G\} \) being a variable which states the cluster which \( i \)th unit belongs to.

Define also \( \epsilon_{n,t} = (\epsilon_{1t}, \epsilon_{2t}, \ldots, \epsilon_{nt})' \), \( c_n = \text{diag}(c_1, c_2, \ldots, c_n)' \), \( c_i = \sum_{h=0}^{\infty} c_{ih} \), \( A_n = (\lambda_1, \lambda_2, \ldots, \lambda_n)' \), \( C = \sum_{h=0}^{\infty} C_h \), \( \tilde{f}_t = f_t - \frac{1}{T} \sum_{t=1}^{T} f_t \), \( \tilde{\epsilon}_{n,t} = \epsilon_{n,t} - \frac{1}{T} \sum_{t=1}^{T} \epsilon_{n,t} \), \( D_n = \text{diag}(n_1, n_2, \ldots, n_G) \) and \( \gamma_{ij,d_t} = E(\epsilon_{it}\tilde{\epsilon}_{js}) \).

Define

\[
V_{1,nT} = \frac{1}{n} \sum_{i=1}^{n} h_{g,i}' \tilde{h}_{g,i} \tilde{\epsilon}_{i,T},
\]

\[
V_{2,nT} = \frac{1}{n} \sum_{i,j=1}^{n} h_{g,i}' h_{g,j}' \tilde{r}_{ij} \tilde{\gamma}_{nT} \tilde{r}_{ij},
\]

\[
V_{3,nT} = \frac{1}{n^2} \sum_{i,j=1}^{n} h_{g,i}' h_{g,j}' \tilde{\gamma}_{ij} \tilde{\gamma}_{nT} \tilde{r}_{ij} + \frac{1}{n^2} \sum_{i,j=1}^{n} h_{g,i}' h_{g,j}' \tilde{r}_{ij} \tilde{\gamma}_{nT} \tilde{r}_{ij}.
\]

To prove Theorems 1 and 2, we need the following lemmas.

**Lemma 1.** Suppose Assumptions 1, 2 and 4 hold. Then, as \( T, n \to \infty \),

\[(i) \lim_{n \to \infty} \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} H_{n}' \epsilon_{n,t} \overset{d}{\to} N(0, V_1), \text{ where } V_1 = \lim_{(n,T) \to \infty} V_{1,nT}, \]
Lemma 2. Suppose Assumptions 1-3 and 6 hold. Then, for any fixed $n$, as $T \to \infty$

(i) $\frac{1}{\sqrt{nT}} \sum_{t=1}^{T} H_n^T R_n \epsilon_{n,t} \overset{d}{\to} N(0, V_2)$, where $V_2 = \lim_{(n,T) \to \infty} V_{2,nT}$;

(ii) $\frac{1}{nT} \sum_{t,s=1}^{T} H_n^T A_n f_{s,t} \overset{d}{\to} N(0, V_3)$, where $V_3 = \lim_{(n,T) \to \infty} V_{3,nT}$.

Lemma 3. Suppose Assumptions 1-8 hold. Then

$$\delta_{nT}(\hat{\lambda}'^i f_t - \lambda_j f_s) = O_p(1)$$

where $\delta_{nT} = \min(\sqrt{T}, \sqrt{n})$.

Lemma 4. Suppose Assumptions 1-8 hold. Then,

(i) $\frac{1}{\sqrt{nT}} \sum_{t=1}^{n} \sum_{s=1}^{T} k_T (d_{ts}/dT) \hat{\epsilon}_{is} = o_p(1)$,

(ii) $\frac{1}{nT} \sum_{i,j=1}^{n} \left[ \frac{1}{T} \sum_{t,s=1}^{T} k_T (d_{ts}/dT) \left( \hat{\lambda}'^i f_t \hat{\lambda}'_j f_s - \lambda_i f_t \lambda_j f_s \right) \right] = o_p(1)$.

Proof of Lemma 1 (i) is a special case of (ii) with $R_n = I_n$, hence we focus on the two last results.

We start by applying the Beveridge-Nelson (BN) decomposition (see, for instance, Phillips and Solo, 1992) to each component of $\epsilon_{n,t}$. We have $\epsilon_{it} = c_i \psi_{it} + \tilde{\psi}_{i,t-1} + \tilde{\psi}_{it}$, where $\tilde{\psi}_{it} = \sum_{h=0}^{\infty} \tilde{c}_{ih} \psi_{i,t-h}$, $\tilde{c}_{ih} = \sum_{j=h+1}^{\infty} c_{ij}$. Then, we can write

$$\frac{1}{\sqrt{nT}} \sum_{t=1}^{T} H_n^T R_n \epsilon_{n,t} = \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} H_n^T R_n c_n \psi_{n,t} + \frac{1}{\sqrt{nT}} H_n^T R_n \tilde{\psi}_{n,t-1} - \frac{1}{\sqrt{nT}} H_n^T R_n \tilde{\psi}_{n,T}$$

$$= A_1 + A_2 + A_3,$$

where $\psi_{n,t} = (\psi_{1t}, \psi_{2t}, \ldots, \psi_{nt})'$ and $\tilde{\psi}_{n,t} = (\tilde{\psi}_{1t}, \tilde{\psi}_{2t}, \ldots, \tilde{\psi}_{nt})'$. The variance of the first term is
Var \( (A_1) = n^{-1} \sum_{i,j=1}^{n} h_{gi} r_i' c_i c_j' r_j h_{gj} \). A typical element of this matrix, say \( c_{ih} \), satisfies

\[
\left| \frac{1}{n} \sum_{i \in G_t} \sum_{j \in G_w} r_i' c_i c_j' r_j \right| = \left| \frac{1}{n} \sum_{i \in G_t} \sum_{j \in G_w} \sum_{k=1}^{n} r_{ik} T_{jk} \sum_{h=0}^{\infty} \sum_{h'=0}^{\infty} c_{kh} c_{kh'} \right|
\leq \frac{1}{n} \sum_{i \in G_t} \sum_{j \in G_w} \sum_{k=1}^{n} |r_{ik}| |r_{jk}| \sum_{h=0}^{\infty} \sum_{h'=0}^{\infty} |c_{kh}| |c_{kh'}| \leq \frac{1}{n} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |r_{ik}| \sum_{j \in G_w} |r_{jk}| \sum_{h=0}^{\infty} \sum_{h'=0}^{\infty} |c_{kh}| |c_{kh'}| = O(1),
\]

where we used the fact that absolute summability is implied by the summability condition we impose on coefficients \( c_{ih} \) in Assumption \[\text{[ii]}\]. Hence we showed that \( A_1 \) is \( O_p(1) \). The variances of the second and the third terms are Var \( (A_2) = (nT)^{-1} \sum_{i,j=1}^{n} h_{gi} r_i' \mathbb{E}(\tilde{\psi}_{n1} \tilde{\psi}_{n1}' r_j h_{gj} \text{ and Var } (A_3) = (nT)^{-1} \sum_{i,j=1}^{n} h_{gi} r_i' \mathbb{E}(\tilde{\psi}_{nT} \tilde{\psi}_{nT}' r_j h_{gj} \text{, respectively. For } [\text{iii}] \text{ to hold, we need to show that these}
\]

are asymptotically negligible. This holds if elements of \( \mathbb{E}(\tilde{\psi}_{n1} \tilde{\psi}_{n1}') \) and \( \mathbb{E}(\tilde{\psi}_{nT} \tilde{\psi}_{nT}') \) are finite which holds under Assumption \[\text{[i]}\] as shown by the BN Lemma in Phillips and Solo (1992). Then the \( l \)th elements of Var \( (A_2) \) and Var \( (A_3) \) satisfy \( (nT)^{-1} \sum_{i \in G_t} \sum_{j \in G_w} r_i' r_j O(1) \) which is \( O(T^{-1}) \) under Assumption \[\text{[iii]}\]. It follows that the second and the third terms are dominated by the first one.

Since the variance of the first term is \( O(1) \), it satisfies a central limit theorem for triangular arrays (Kelejian and Prucha, 1998) and \[\text{[ii]}\] follows.

To prove \[\text{[iii]}\], we apply the multivariate generalization of the BN decomposition (see, for instance, Phillips and Solo, 1992, p. 985) to \( \bar{f} \). We have

\[
\frac{1}{n \sqrt{T}} \sum_{t=1}^{T} H_n' A_n \bar{f}_t = \frac{1}{n \sqrt{T}} \sum_{t=1}^{T} H_n' A_n C \Psi_t + \frac{1}{n \sqrt{T}} H_n' A_n \bar{\Psi}_t - \frac{1}{n \sqrt{T}} H_n' A_n \bar{\Psi}_t
\]

\[= B_1 + B_2 + B_3, \]

where \( \bar{\Psi}_t = \sum_{h=0}^{\infty} \bar{C}_h \Psi_{t-h} \), \( \bar{C}_h = \sum_{j=h+1}^{\infty} C_j \). The variance of the first term is given by Var \( (B_1) = \frac{1}{n^2} \sum_{i,j=1}^{n} h_{gi} \lambda_i' C C' \lambda_j h_{gj} \). The \( l \)th element of this matrix satisfies

\[
\left| \frac{1}{n^2} \sum_{i \in G_t} \sum_{j \in G_w} \lambda_i' C C' \lambda_j \right| \leq \frac{1}{n^2} \sum_{i \in G_t} \sum_{j \in G_w} |\lambda_i' C C' \lambda_j| \leq \frac{1}{n^2} \sum_{i \in G_t} \sum_{j \in G_w} ||C||^2 ||\lambda_i|| ||\lambda_j|| = O(1),
\]

where we used the fact that \( ||C||^2 \leq \sum_{h=0}^{\infty} ||C_h||^2 < \infty \). It follows that \( B_1 \) is \( O_p(1) \). The variances of the second and the third terms are Var \( (B_2) = \frac{1}{n \sqrt{T}} H_n' A_n E(\tilde{\Psi}_1 \tilde{\Psi}_1') A_n' H_n \) and Var \( (B_3) = \frac{1}{n \sqrt{T}} H_n' A_n E(\tilde{\Psi}_T \tilde{\Psi}_T') A_n' H_n \), respectively. Similar to the reasoning above, we have \( E(\tilde{\Psi}_1 \tilde{\Psi}_1') < \infty \)
and \( \text{E}(\Psi_T \Psi_T') < \infty \), under the summability condition we impose on the coefficients of the process \( f_T \) (see, for instance, Phillips and Magdalinos, 2003). Furthermore, the \( lwth \) element of the matrix 
\[
\frac{1}{n} \mathbf{H}'_n \mathbf{A}_n \mathbf{A}'_n \mathbf{H}_n
\]

satisfies 
\[
|n^{-2} \sum_{i \in G_L} \sum_{j \in G_o} \sum_{k,k'=1}^n \lambda_{ik} \lambda_{jk'}| = O(1).
\]
Hence, \( B_2 \) and \( B_3 \) are \( o_p(1) \). Since the first term is \( O_p(1) \), it dominates the second and the third terms and a central limit theorem for triangular arrays (Kelejian and Prucha, 1998) applies to the first term. Now, since the second term in \( \mathbf{V}_{3,nT} \) is \( O(n^{-1}) \) from [ii], the asymptotic variance of \( \mathbf{V}_{3,nT} \) equals the asymptotic variance of \( B_1 \) which leads to [iii]

**Proof of Lemma 2** [i] is proved in Theorem 2 of Jansson (2002). [ii] follows immediately from their proof noting that \( T^{-1} \sum_{t=1}^T f_t = o_p(1) \) under Assumption [a]. [iii] holds again by their proof, as by Assumption [a] \( f_t \) and \( \epsilon_{n,s} \) are independent for every \( t, s \), and noting that \( T^{-1} \sum_{t=1}^T \epsilon_{n,t} = o_p(1) \). Similarly, [iv] is a result of their proof under Assumption [a].

**Proof of Lemma 3**. The proof amounts to verify that Assumptions A-F of Bai (2003) hold. The model considered by Bai (2003) does not contain a constant term. On the contrary, our common factor model is \( \Delta L_{it} = \mu_i + \lambda'_i f_t + \epsilon_{it} \) where \( \mu_i \) is not necessarily zero. But we have \( \Delta L_{it} = (\mu_i + \Delta \tilde{L}_{i,T}) + \lambda'_i f_t + \epsilon_{it} \). Since by Assumptions [a] and [b] \( \mu_i - \Delta \tilde{L}_{i,T} = o_p(1) \), we have \( \Delta L_{it} = \lambda'_i f_t + \epsilon_{it} + o_p(1) \) where the error term is uniform in \( i \). We can therefore use \( \Delta \tilde{L}_{it} \) to estimate \( \lambda_i \) and \( f_t \) using PCE.

For simplicity, we will rely on this approximation, and show that the statement holds for the model \( \Delta \tilde{L}_{it} = \lambda'_i f_t + \epsilon_{it} \).

Bai’s Assumption A requires \( \text{E}[|f_t|^4] < \infty \) and \( T^{-1} \sum_{t=1}^T f_t f_t' \overset{p}{\to} \Sigma_f \) where \( \Sigma_f \) is positive definite. First statement follows from the moment condition on the innovations in Assumption [a] together with the summability condition on the coefficients. This was proved for the scalar case in the proof of their Theorem 3.3 by Phillips and Solo (1992). For the proof they use Burkholder’s inequality. Application of a vector-generalization of this inequality (Pinelis, 1994) leads to the statement. Second statement is implicitly proved in the proof of Lemma [a][iii].

Assumption B is identical to our Assumption [c]. Assumption C1 follows from our Assumption [c] by the same lines of the proof of Theorem 3.3 of Phillips and Solo (1992). For the first part of Assumption C2, write \( \text{E}(n^{-1} \sum_{i=1}^n \epsilon_{it} \epsilon_{is}) = \text{E}(n^{-1} \sum_{i=1}^n r^*_i, \gamma_n, d_n, r_i) = n^{-1} \sum_{i=1}^n r^*_i, \gamma_n, d_n, r_i \).

Then, 
\[
|n^{-1} \sum_{i=1}^n r^*_i, \gamma_n, d_n, r_i| = |n^{-1} \sum_{i,j=1}^n r^2_{ij} \gamma_{ij,0}| = n^{-1} \sum_{j=1}^n \gamma_{j,0} \sum_{i=1}^n r^2_{ij} < \infty
\]

by absolute summability of coefficients \( r_{ij} \) which gives the desired result. We have, 
\[
T^{-1} \sum_{t,s=1}^T |n^{-1} \sum_{i=1}^n r^*_i, \gamma_n, d_n, r_i| \leq
\]
\( T^{-1} \sum_{t,s=1}^{T} n^{-1} \sum_{i,j=1}^{n} r_{ij}^2 \gamma_{j,t,s} = n^{-1} \sum_{i,j=1}^{n} r_{ij}^2 T^{-1} \sum_{t,s=1}^{T} |\gamma_{j,t,s}| = O(1) n^{-1} \sum_{i,j=1}^{n} r_{ij}^2 < \infty \). This proves the second part of Assumption C2. To show Assumption C3, write \( E(\varepsilon_{it}\varepsilon_{jt}) = r_{ij}^t \gamma_{n,0} r_{ij} \).

We have the desired result by

\[
\left| n^{-1} \sum_{i,j,k=1}^{n} r_{ik} r_{jk} \gamma_{k,0} \right| \leq n^{-1} \sum_{i,j,k=1}^{n} |r_{ik}| |r_{jk}| \gamma_{k,0} = n^{-1} \sum_{k=1}^{n} \gamma_{k,0} \sum_{i=1}^{n} |r_{ik}| \sum_{j=1}^{n} |r_{jk}| < \infty.
\]

Assumption C4 follows by \( (nT)^{-1} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} E(\varepsilon_{it} \varepsilon_{js}) = n^{-1} \sum_{i,j=1}^{n} r_{ij}^t \gamma_{nT} r_{ij} \) where we have \( \gamma_{nT} = O(1) \) and

\[
\left| n^{-1} \sum_{i,j=1}^{n} r_{ij}^t r_{ij} \right| = \left| n^{-1} \sum_{i,j,k=1}^{n} r_{ik} r_{jk} \right| \leq n^{-1} \sum_{i,j,k=1}^{n} |r_{ik}| |r_{jk}| < \infty.
\]

Assumption C5 is identical to the second part of our Assumption 7.

Assumption D is implied by Assumption A and C of Bai under the the assumption that \( f_i \) and \( \varepsilon_{it} \) are independent. Assumption E1 states, \( \sum_{s=1}^{T} |E(n^{-1} \sum_{i=1}^{n} \varepsilon_{it} \varepsilon_{is})| < \infty \). This expression satisfies

\[
\sum_{s=1}^{T} |r_{ij}^t \gamma_{n,ds} r_i| \leq n^{-1} \sum_{k=1}^{n} \sum_{s=1}^{T} |\gamma_{ks,d}| \sum_{i=1}^{n} r_{ik}^2 < \infty
\]

and

\[
\sum_{s=1}^{T} |\gamma_{k,ds}| < \sum_{s=1}^{T} \sum_{h=0}^{\infty} |c_{ih}| |c_{i,ds}| < \infty.
\]

Assumption E2 follows as

\[
\sum_{k=1}^{n} |E(\varepsilon_{it} \varepsilon_{kt})| \leq \sum_{k=1}^{n} \sum_{h=1}^{n} |r_{ih}| |r_{kh}| \sum_{h=0}^{\infty} |c_{ih}| |c_{kh}| < \infty.
\]

Assumptions F1 and F2 are identical to our Assumptions 4(a) and 4(b). Assumptions F3 and F4 are CLTs which follow immediately from our Assumptions 1 and 6. Then the stated result follows from Theorem 3 of Bai (2003).

**Proof of Lemma 2.** We have \( \varepsilon_{it} = \Delta L_{it} - \Delta \hat{L}_{i,T} - \hat{\lambda}^t f_i = (\mu_i - \Delta \hat{L}_{i,T}) + (\lambda^t f_i - \hat{\lambda}^t f_i) + \varepsilon_{it} \). By Assumptions 1 and 6, \( \mu_i - \Delta \hat{L}_{i,T} = O_p(T^{-1/2}), \varepsilon_{it} = O_p(1) \) and by Lemma 3, \( \lambda^t f_i - \hat{\lambda}^t f_i = O_p(\delta_{n_T}^{-1}) \). After

\[
(2)\quad (n^2 T)^{-1} \sum_{i=1}^{n} \sum_{t,s=1}^{T} k_t (d_{ts}/dt) \varepsilon_{it} \varepsilon_{is} = O_p(\delta_{n_T}^{-1}) + (n^2 T)^{-1} \sum_{i=1}^{n} \sum_{t,s=1}^{T} k_t (d_{ts}/dt) \varepsilon_{it} \varepsilon_{is} = A_1 + A_2,
\]

we have

\[
A_1 = -n^{-2} \sum_{i=1}^{n} O_p(\delta_{n_T}^{-1}) \left[ T^{-1} \sum_{h=-T+1}^{T-1} k_t (h/dT) \right]
\]

where the term in brackets is \( O(d_T/T) \) by Assumption 2 which in turn gives \( A_1 = O(n^{-1}) O_p(\delta_{n_T}^{-1}) O(d_T/T) = o_p(1) \). For the second term, we find

\[
A_2 = \frac{1}{n^2 T} \sum_{i=1}^{n} \sum_{t,s=1}^{T} k_t (d_{ts}/dt) r_{ij}^t \gamma_{n,ds} r_i = n^{-2} \sum_{i=1}^{n} r_{ik}^2 \left[ T^{-1} \sum_{h=-T+1}^{T-1} k_t (h/dT) \right]
\]

By Assumption 1, the spectral density function of \( \varepsilon_{it}, f_i(\cdot) \), exists and is bounded. By Theorem 2 of Jansson (2002), under Assumption 2,

\[
\lim_{T \to \infty} T^{-1} \sum_{h=-T+1}^{T-1} k_t (h/dT) \gamma_i,h = T^{-1} 2\pi f_i(0) = O(T^{-1})
\]


for each $i$. Then we have $E(A_2) = O(T^{-1})n^{-2} \sum_{i,k=1}^n r_{ik}^2 = O[(nT)^{-1}]$. For the variance, we find

$$
\text{Var}(A_2) = E \left[ \frac{1}{n^4 T^2} \sum_{i,j=1}^n \sum_{t_1, t_2, t_3, t_4=1}^T k_T \left( d_{1,i} \frac{d_{1,s_1}}{d_T} \right) k_T \left( d_{2,s_2} \frac{d_{2,s_2}}{d_T} \right) E(\epsilon_{it_1} \epsilon_{is_1} \epsilon_{jt_2} \epsilon_{js_2}) \right] 
$$

$$
= \frac{1}{n^4 T^2} \sum_{i,j=1}^n \sum_{t_1, t_2, t_3, t_4=1}^T k_T \left( d_{1,i} \frac{d_{1,s_1}}{d_T} \right) k_T \left( d_{2,s_2} \frac{d_{2,s_2}}{d_T} \right) E(\epsilon_{it_1} \epsilon_{is_1} \epsilon_{jt_2} \epsilon_{js_2}) 
$$

$$
= \frac{1}{n^4 T^2} \sum_{i,j=1}^n \sum_{t_1, t_2, t_3, t_4=1}^T k_T \left( d_{1,i} \frac{d_{1,s_1}}{d_T} \right) k_T \left( d_{2,s_2} \frac{d_{2,s_2}}{d_T} \right) E(r_{it_1} \epsilon_{n,t_1} \epsilon_{n,s_1} r_{it_2} \epsilon_{n,t_2} r_{is_2} \epsilon_{n,s_2} r_{js_2}) 
$$

$$
= \frac{1}{n^4 T^2} \sum_{i,j=1}^n \sum_{t_1, t_2, t_3, t_4=1}^T k_T \left( d_{1,i} \frac{d_{1,s_1}}{d_T} \right) k_T \left( d_{2,s_2} \frac{d_{2,s_2}}{d_T} \right) E \left( \sum_{l_1, l_2=1}^n r_{il_1} r_{il_2} \epsilon_{l_1,t_1} \epsilon_{l_2,s_1} \sum_{l_3, l_4=1}^n r_{jl_3} r_{jl_4} \epsilon_{l_3,t_2} \epsilon_{l_4,s_2} \right) 
$$

$$
= \frac{1}{n^4} \sum_{i,j,l_1, l_2, l_3, l_4=1}^n r_{il_1} r_{il_2} r_{jl_3} r_{jl_4} \left[ \frac{1}{T^2} \sum_{t_1, t_2, t_3, t_4=1}^T k_T \left( d_{1,i} \frac{d_{1,s_1}}{d_T} \right) k_T \left( d_{2,s_2} \frac{d_{2,s_2}}{d_T} \right) E(\epsilon_{l_1,t_1} \epsilon_{l_2,s_1} \epsilon_{l_3,t_2} \epsilon_{l_4,s_2}) \right]. 
$$

(29)

In general, the expectation in the last line can be written as

$$
E(\epsilon_{l_1,t_1} \epsilon_{l_2,s_1} \epsilon_{l_3,t_2} \epsilon_{l_4,s_2}) = \gamma_{l_1,t_2} \gamma_{l_3,t_2} \gamma_{l_2,l_4} \gamma_{l_4,l_1} + \gamma_{l_1,t_2} \gamma_{l_3,t_2} \gamma_{l_2,l_4} \gamma_{l_4,l_1} + \gamma_{l_1,t_2} \gamma_{l_3,t_2} \gamma_{l_2,l_4} \gamma_{l_4,l_1} + \gamma_{l_1,t_2} \gamma_{l_3,t_2} \gamma_{l_2,l_4} \gamma_{l_4,l_1},
$$

where $\kappa(\cdot)$ is the cumulant of the fourth order between $\epsilon_{l_1,t_1}$, $\epsilon_{l_2,s_1}$, $\epsilon_{l_3,t_2}$ and $\epsilon_{l_4,s_2}$ (see Hannan, 1970, p. 23). In our case, by Assumption $\mathcal{H}$ $\epsilon_{it}$ are independent over $i$, hence, the cumulant is null. Furthermore, the covariances in the expression are null unless $l_i = l_j$ for $i, j = 1, 2, 3, 4$. Using these in (29), we find

$$
\text{Var}(A_2) = A_{21} + A_{22} + A_{23}. 
$$

All terms in the brackets are $O(T^{-1})$. Hence, $A_{21} = O(T^{-2})n^{-4} \sum_{i,j=1}^n r_{i1}^2 \sum_{l_1=1}^n r_{l_2}^2 = O[(nT)^{-2}]$. Similarly, $A_{22} = A_{23} = O[(nT)^{-2}]$. As a result, $A_2 = O_p((nT)^{-1}) = o_p(1)$.

For (ii), we write $\lambda_t f_t \hat{f}_s \lambda_j - \lambda_t f_{t'} \hat{f}_s \lambda_j = (\hat{X}_t f_t - \lambda_t f_t) \hat{f}_s \lambda_j + \lambda_t f_t (\hat{f}_s \lambda_j - f_{t'} \lambda_j)$. By Assumption $\mathcal{E}$.
\(X'_1f_t\), and by Lemma 3 \(\hat{X}'_1f'_s\) are \(O_p(1)\). This implies that \(\hat{X}'_1f'_s\hat{\lambda}_j - X'_1f'_s\lambda_j = O_p(\delta_{nT}^{-1})\) by Lemma 3. Then for the expression in the statement we obtain \(\frac{1}{n^2}\sum_{i,j=1}^n O_p(\delta_{nT}^{-1}) \left[ \frac{1}{T} \sum_{h=-T+1}^{T-1} k_T(h/d_T) \right] = O_p(\delta_{nT}^{-1})O(d_T/T) = o_p(1)\).

**Proof of Proposition 1.** This is a special case of Theorem 1 with \(R_n = I_n\) and \(G = 1\).

**Proof of Proposition 2.** This is a special case of Theorem 1 with \(G = 1\).

**Proof of Proposition 3.** This is a special case of Theorem 2 with \(G = 1\).

**Proof of Proposition 4.** This is a special case of Theorem 1 with \(R_n = I_n\).

**Proof of Theorem 1.** Define \(\Delta L_{n,t} = (\Delta L_{1,t}, \Delta L_{2,t}, \ldots, \Delta L_{n,t})'\). We have

\[
\Delta \tilde{L}_{nT} = D_{nT}^{-1} \frac{1}{T} \sum_{t=1}^T H'_n \Delta L_{n,t} = D_{nT}^{-1} \frac{1}{T} \sum_{t=1}^T H'_n (\mu + R_n \epsilon_{n,t}) = \bar{\mu}_n + D_{nT}^{-1} \frac{1}{T} \sum_{t=1}^T H'_n R_n \epsilon_{n,t},
\]

where \(\mu = (\mu_1, \mu_2, \ldots, \mu_n)'\). It follows that

\[
\sqrt{nT}(\Delta \tilde{L}_{nT} - \bar{\mu}_n) = \left( \frac{D_{nT}}{n} \right)^{-1} \frac{1}{\sqrt{nT}} \sum_{t=1}^T H'_n R_n \epsilon_{n,t}.
\]

Then \(\sqrt{nT}\Omega_{2,nT}^{-1/2}(\Delta \tilde{L}_{nT} - \bar{\mu}_n) \overset{D}{\rightarrow} N(0, I_G)\), where \(\Omega_{2,nT} = \sum_{i,j=1}^n \frac{\sigma_{ij}}{\sigma_{ij}^2} h_{ij} h'_{ij} \gamma_{nt} \gamma'_{nt}\) by Lemma 1(ii) and noting that \(n^{-1}D_n\) converges to a finite and nonsingular matrix under Assumption 3. The matrix \(\Omega_{2,nT}\) can be written as \(\Omega_{2,nT} = (D_{nT}/n)^{-1} V_{2,nT}(D_{nT}/n)^{-1}\). Since \(D_n\) is known, estimation of \(\Omega_{2,nT}\) requires only the estimation of \(V_{2,nT}\). Similarly, we write \(\hat{\Omega}_{2,nT} = (D_{nT}/n)^{-1} \hat{V}_{2,nT}(D_{nT}/n)^{-1}\)

where

\[
\hat{V}_{2,nT} = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T k_T(d_{ts}) k_s \left( \frac{d_{ij}}{d_n} \right) h_{ij} h'_{ij} \Delta \tilde{L}_{it} \Delta \tilde{L}_{js},
\]

The \(lw\)th element of the matrices \(V_{2,nT}\) and \(\hat{V}_{2,nT}\) are

\[
v_{2,nT}^{lw} = \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} r'_{ij} \delta_{nt} r_{ij}, \tag{30}
\]

and

\[
\hat{v}_{2,nT}^{lw} = \frac{1}{nT} \sum_{i \in G_l} \sum_{j \in G_w} \sum_{t,s=1}^T k_s \left( \frac{d_{ij}}{d_n} \right) k_T \left( \frac{d_{ts}}{d_T} \right) \Delta \tilde{L}_{it} \Delta \tilde{L}_{js},
\]

respectively. We will show that \(\hat{v}_{2,nT}^{lw} - v_{2,nT}^{lw} = o_p(1)\) which gives the first result in the proposition.
We have $\Delta \tilde{L}_{it} = \mathbf{r}'_i \tilde{e}_{n,t}$ which gives $\Delta \tilde{L}_{it} \Delta \tilde{L}_{js} = \mathbf{r}'_i \tilde{e}_{n,t} \tilde{e}'_{n,s} \mathbf{r}_j$. Then we have,

$$\tilde{v}^{lw}_{2,nT} - v^{lw}_{2,nT} = \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} k_S \left( \frac{d_{ij}}{d_n} \right) \mathbf{r}'_i \left[ \frac{1}{T} \sum_{t,s=1}^T k_T \left( \frac{d_{ts}}{d_T} \right) \tilde{e}_{n,t} \tilde{e}'_{n,s} \right] \mathbf{r}_j - \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} \mathbf{r}'_i \tilde{\gamma}_{nT} \mathbf{r}_j.$$

Since $\tilde{\gamma}_{nT} = O(1)$, and $\frac{1}{T} \sum_{t,s=1}^T k_T \left( \frac{d_{ts}}{d_T} \right) \tilde{e}_{n,t} \tilde{e}'_{n,s} = o_p(1)$ by Lemma 9(iv), it suffices to show that $\frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} k_S \left( \frac{d_{ij}}{d_n} \right) \mathbf{r}'_i \mathbf{r}_j = O(1)$ and $\frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} [1 - k_S \left( \frac{d_{ij}}{d_n} \right)] \mathbf{r}'_i \mathbf{r}_j = o(1)$ in order to prove the consistency of $v^{lw}_{2,nT}$. Starting with the latter, we have

$$\left| \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} \left[ 1 - k_S \left( \frac{d_{ij}}{d_n} \right) \right] \mathbf{r}'_i \mathbf{r}_j \right| \leq \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} \left| 1 - k_S \left( \frac{d_{ij}}{d_n} \right) \right| | \mathbf{r}'_i \mathbf{r}_j | \leq \frac{1}{n d_{ij}} \sum_{i \in G_l} \sum_{j \in G_w} | \mathbf{r}'_i \mathbf{r}_j | d_{ij} = O \left( \frac{1}{d_{ij}} \right) = o(1),$$

where the last equality follows by the assumptions that $d_n \to \infty$ and $\rho_s \geq 1$. For the first term, write

$$\frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} k_S \left( \frac{d_{ij}}{d_n} \right) \mathbf{r}'_i \mathbf{r}_j = \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} \mathbf{r}'_i \mathbf{r}_j - \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} \left[ 1 - k_S \left( \frac{d_{ij}}{d_n} \right) \right] \mathbf{r}'_i \mathbf{r}_j = O(1), \quad (31)$$

which follows from the fact that the first term is $O(1)$ and the second term is $o(1)$ by the previous equation. This completes the proof of (i).

To show the consistency of $\tilde{\Omega}_{2,nT}$, we first write $\tilde{\Omega}_{2,nT} = (\mathbf{D}_n/n)^{-1} \tilde{\mathbf{V}}_{2,nT} (\mathbf{D}_n/n)^{-1}$, where

$$\tilde{\mathbf{V}}_{2,nT} = \frac{1}{n T} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t,s=1}^T k_T \left( \frac{d_{ts}}{d_T} \right) \mathbf{h}_{g_i} \mathbf{h}'_{g_j} \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}.$$  

The $l$th element of this matrix, corresponding to clusters $l$ and $w$ is

$$\tilde{v}^{lw}_{2,nT} = \frac{1}{n T} \sum_{i \in G_l} \sum_{j \in G_w} \mathbf{r}'_i \left[ \frac{1}{T} \sum_{t,s=1}^T k_T \left( \frac{d_{ts}}{d_T} \right) \tilde{e}_{n,t} \tilde{e}'_{n,t} \right] \mathbf{r}_j,$$

where $G_k, k = 1, \ldots, G$, is the set of indices in cluster $k$ which are used in the calculation of the partial variance estimate. This set has a cardinality of $n_k$. Using this expression together with (30), we can write

$$\tilde{v}^{lw}_{2,nT} - v^{lw}_{2,nT} = (\tilde{v}^{lw}_{2,nT} - v^{lw}_{2,nT}) + (v^{lw}_{2,nT} - v^{lw}_{2,nT}). \quad (32)$$
where \( \hat{v}_{2, nT}^{lw} = \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} r'_i \tilde{\gamma}_{nT} r_j. \) Our objective is to show that both terms in parentheses approach to zero. We have

\[
\hat{v}_{2, nT}^{lw} = \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} r'_i \left[ \frac{1}{T} \sum_{t, s=1}^T k_T \left( \frac{d_{ts}}{d_T} \right) \varepsilon_{n, t} \varepsilon'_{n, t} - \tilde{\gamma}_{nT} \right] r_j.
\]

The term in brackets is \( o_p(1) \) by Lemma 2.4. Then it is sufficient to show that \( \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} r'_i r_j \) is bounded. We have

\[
\left| \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} r'_i r_j \right| = \left| \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} \sum_{k=1}^n r_{ik} r_{jk} \right| \leq \left| \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} \sum_{k=1}^n r_{ik} \right| \left| \sum_{j \in G_w} r_{jk} \right| = \frac{n \mu}{n} O(1),
\]

which is \( O(1) \) if \( \kappa \in (0, 1) \) and it is \( o(1) \) if \( \kappa = 0 \). For the second term in (32) we have

\[
\hat{v}_{2, nT}^{lw} = \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} r'_i \tilde{\gamma}_{nT} r_j - \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} r'_i \tilde{\gamma}_{nT} r_j.
\]

Define \( G_k^c = G_k \setminus G_k \), the complement of \( G_k \) for cluster \( k \). Since \( \tilde{\gamma}_{nT} = O(1) \), we have

\[
\hat{v}_{2, nT}^{lw} - \hat{v}_{2, nT}^{lw} = O(1) \left( \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} \sum_{k=1}^n r_{ik} r_{jk} \right) \sum_{i \in G_j} \sum_{j \in G_w} \sum_{k=1}^n r_{ik} r_{jk} = O(1) \left[ \left( 1 - \frac{n}{n} \right) \left( \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} \sum_{k=1}^n r_{ik} r_{jk} \right) - \left( \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} \sum_{k=1}^n r_{ik} r_{jk} \right) \right].
\]

For the result to hold, it will suffice to show that the term in brackets goes to zero. By (33),

\[
\frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} \sum_{k=1}^n r_{ik} r_{jk} \]

is \( O(1) \) if \( \kappa \in (0, 1) \). Whereas, it is \( o(1) \) if \( \kappa = 0 \). If \( \kappa = 0 \), \( 1 - n/n = O(1) \), hence, in this case the first term in the brackets is \( o(1) \). If \( \kappa = 1 \), we have \( 1 - n/n = o(1) \) and

\[
\frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} \sum_{k=1}^n r_{ik} r_{jk} \]

is \( O(1) \), so the first term in the brackets is again \( o(1) \).

Focusing on the second term, we have

\[
\left| \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} \sum_{k=1}^n r_{ik} r_{jk} \right| \leq \left| \frac{1}{n} \sum_{i \in G_j} \sum_{j \in G_w} \sum_{k=1}^n r_{ik} \right| \left| \sum_{j \in G_w} r_{jk} \right| = \frac{n - n \mu}{n} O(1).
\]

We can write

\[
\frac{n - n \mu}{n} = \frac{n}{n} - \frac{n \mu}{n} \frac{n}{n}
\]

where all three ratios tend to \( \kappa \). Then the result in (ii) holds if (and only if) \( \kappa \in \{0, 1\} \) which is the
Proof of Theorem 2. We have
\[ \Delta \tilde{L}_{nT} = \mu_n + D_n^{-1} \frac{1}{T} \sum_{t=1}^{T} H_n' R_n \epsilon_{n,t} + D_n^{-1} \frac{1}{T} \sum_{t=1}^{T} H_n' \Lambda_n f_t, \]
from which, we obtain
\[ \sqrt{T}(\Delta \tilde{L}_{nT} - \bar{\mu}_n) = \left( \frac{D_n}{n} \right)^{-1} \frac{1}{n \sqrt{T}} \sum_{t=1}^{T} H_n' R_n \epsilon_{n,t} + \left( \frac{D_n}{n} \right)^{-1} \frac{1}{n \sqrt{T}} \sum_{t=1}^{T} H_n' \Lambda_n f_t. \]
From Lemma [iii] it follows that \( \frac{1}{n \sqrt{T}} \sum_{t=1}^{T} H_n' R_n \epsilon_{n,t} = O_p(n^{-1/2}) \). Then, \( \sqrt{T} \Omega_{3,nT}^{-1/2}(\Delta \tilde{L}_{nT} - \bar{\mu}_n) \overset{D}{\rightarrow} N(0, I_G) \), where \( \Omega_{3,nT} = \sum_{i,j=1}^{n} \frac{1}{n_{g_i} n_{g_j}} h_{g_i} h_{g_j}' (\lambda_i' \Gamma_T \lambda_j + r_i' \gamma_{n,t} r_j) \) by noting that \( n^{-1} D_n \) converges to a finite and nonsingular matrix under Assumption [ii] and Lemma [iii]. The matrix \( \Omega_{3,nT} \) can be written as \( \Omega_{3,nT} = (D_n/n)^{-1} V_{3,nT}(D_n/n)^{-1} \). As in the proof of Theorem [i] estimation of \( \Omega_{2,nT} \), requires only the estimation of \( V_{2,nT} \) because \( D_n \) is known. Write \( \tilde{\Omega}_{3,nT} = (D_n/n)^{-1} \tilde{V}_{3,nT}(D_n/n)^{-1} \) where
\[ \tilde{V}_{3,nT} = \frac{1}{n^2 T} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} k_T \left( \frac{d t_s}{d T} \right) h_{g_i} h_{g_j}' \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}. \]
The \( lw \)th element of the matrices \( V_{3,nT} \) and \( \tilde{V}_{3,nT} \), corresponding to clusters \( l \) and \( w \) are
\[ v_{3,nT}^{lw} = \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{j \in G_w} \sum_{t,s=1}^{T} (\lambda_i' \Gamma_{dt_i} \lambda_j + r_i' \gamma_{n,t} r_j), \quad (35) \]
and
\[ \tilde{v}_{3,nT}^{lw} = \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{j \in G_w} \sum_{t,s=1}^{T} k_T \left( \frac{d t_s}{d T} \right) \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}, \]
respectively. We will show that \( \tilde{v}_{3,nT}^{lw} - v_{3,nT}^{lw} = o_p(1) \) which gives the first result in the proposition.
We have \( \Delta \tilde{L}_{it} = \lambda_i' \tilde{f}_t + r_i' \tilde{e} \) which gives
\[ \Delta \tilde{L}_{it} \Delta \tilde{L}_{js} = \lambda_i' \tilde{f}_t \lambda_j + \lambda_j' \tilde{f}_t \lambda_j + r_i' \tilde{e} \lambda_j + r_j' \tilde{e} \lambda_j. \]
Using the last three equations, we obtain
\[
\hat{v}_{3,nT}^{lw} - v_{3,nT}^{lw} = \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \lambda_i' \left[ \frac{1}{T} \sum_{t,s=1}^{T} k_T \left( \frac{d_t s}{d_T} \right) \tilde{f}_t \tilde{f}_s' - \tilde{f}_T \right] \lambda_j \\
+ \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \lambda_i' \left[ \frac{1}{T} \sum_{t,s=1}^{T} k_T \left( \frac{d_t s}{d_T} \right) \tilde{f}_t \tilde{z}_s' \right] \textbf{r}_j \\
+ \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \textbf{r}_i' \left[ \frac{1}{T} \sum_{t,s=1}^{T} k_T \left( \frac{d_t s}{d_T} \right) \tilde{e}_t \tilde{f}_s \right] \lambda_j \\
+ \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \textbf{r}_i' \left[ \frac{1}{T} \sum_{t,s=1}^{T} k_T \left( \frac{d_t s}{d_T} \right) \tilde{e}_t \tilde{z}_s' - \tilde{\gamma}_n T \right] \textbf{r}_j.
\]

We will show that each of these four terms are $o_p(1)$. By Lemma 2 all expressions in square brackets are $o_p(1)$. The first term can be written as $D_1 = \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \lambda_i \lambda_j o_p(1)$. By Hölder’s inequality we have $|\lambda_i \lambda_j| \leq ||\lambda_i|| ||\lambda_j||$ where the right hand side is bounded by Assumption 6(b).

This shows that $D_1 = o_p(1)$. Other terms can be shown to be $o_p(1)$ similarly which in turn gives $\hat{V}_{3,nT} - V_{3,nT} = o_p(1)$ and hence $\hat{\Omega}_{3,nT} - \Omega_{3,nT} = o_p(1)$. The consistency of $\hat{\Omega}_{3,nT}$ in turn implies the asymptotic null distribution which completes the proof of (i).

For the second result, we write $\hat{\Omega}_{3,nT} = (D_n/n)^{-1} \hat{V}_{3,nT}(D_n/n)^{-1}$, where
\[
\hat{V}_{3,nT} = \frac{1}{n^2 T} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} k_T \left( \frac{d_t s}{d_T} \right) h_{i,j} h_{i,j}' \hat{f}_t \hat{f}_s + \frac{1}{n^2 T} \sum_{i=1}^{n} \sum_{t,s=1}^{T} k_T \left( \frac{d_t s}{d_T} \right) h_{i,j} h_{i,j}' \hat{\epsilon}_t \hat{\epsilon}_s.
\]
The $lw$th element of this matrix is
\[
\hat{v}_{l,nT}^{lw} = \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{j \in G_w} \sum_{t,s=1}^{T} k_T \left( \frac{d_t s}{d_T} \right) \hat{f}_i \hat{f}_j' \hat{f}_s + \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{j \in G_w} \sum_{t,s=1}^{T} k_T \left( \frac{d_t s}{d_T} \right) \hat{\epsilon}_t \hat{\epsilon}_s.
\]
Using this expression together with (35), we find

\[ \sum_{i \in G_1, j \in G_w} \sum_{t, s=1}^{T} k_T \left( \frac{d t_s}{d T} \right) \lambda_i' \lambda_j' f_t f_s - \lambda_i' \Gamma_T \lambda_j \]

\[ + \sum_{i \in G_1, j \in G_w} \sum_{t, s=1}^{T} k_T \left( \frac{d t_s}{d T} \right) \tilde{\varepsilon}_i \tilde{\varepsilon}_j - \sum_{i \in G_1, j \in G_w} r_i' \tilde{\gamma}_n T r_j. \]

\[ = \sum_{i \in G_1, j \in G_w} \sum_{t, s=1}^{T} k_T \left( \frac{d t_s}{d T} \right) \lambda_i' \lambda_j' f_t f_s - \lambda_i' f_t f_j \]

\[ + \sum_{i \in G_1, j \in G_w} \sum_{t, s=1}^{T} k_T \left( \frac{d t_s}{d T} \right) \tilde{\varepsilon}_i \tilde{\varepsilon}_j - \sum_{i \in G_1, j \in G_w} r_i' \tilde{\gamma}_n T r_j. \]

The desired result now follows from Lemma 4(i), Lemma 4(ii), Lemma 2(i) and by noting that the last term is \( O(n^{-1}) \). This completes the proof.

**Appendix C  An Evaluation of the IMF Consumer Price Inflation Forecasts**

In this appendix, we evaluate the consumer price inflation forecasts (CPI) made by the IMF by comparing them with no change, i.e. random walk (RW) forecasts. As in the previous application, the data for the IMF forecasts come from the Fund’s Historical WEO Forecasts Database. Once again we focus on their summer forecasts made for the following year, hence we are dealing with one year ahead forecasts. Our data set contains 127 countries for which the forecasts are available from 1991 to 2019, i.e. the panel is balanced. We exclude 5 countries from the data set as their loss differentials are very different from the rest of the sample. These countries are BRA, COD, PER, VEN and NIC. Notice that all of these countries experienced hyperinflation in early to mid-90’s or in late 2010’s, in the case of VEN. For the first four countries there are very big drops or jumps in the CPI inflation, hence their RW forecasts are very poor for at least one year. Whereas for the last country the situation is the contrary, that is IMF forecasts are much worse than the RW forecasts. Our conclusions should be understood to apply to the 127 countries in our sample which includes the
G7 countries, 35 OECD countries (no data available for LVA and LTU) and 92 non-OECD countries.

As in the application reported in the main text, we start our analysis by applying the DM test to each country in the sample using two loss functions, i.e. absolute and quadratic losses. The loss differentials are computed by taking IMF as the first predictor and RW the second. Since we have a large number of countries, we report the statistics in histograms with 10 bins. Here, we set the time series kernel bandwidth to \( l_T = 1 \) as our autocorrelation test below indicates that the loss differentials are serially uncorrelated. The results for the individual DM tests are reported in Figure 3. As can be seen, for each loss function, the average DM statistics over the countries is very close to zero and a considerable amount of heterogeneity exists in the sample. We identified only 22 countries in the sample for which at least one loss function provides a significant DM statistic at 5% level.

**Cross-sectional and temporal dependence in loss differentials: CPI forecasts.** Before looking into the EPA test results we apply the methodology that we propose and check if we can find evidence for CD in our sample and identify its type. To save space, we do not report the diagnostic results in tables in this subsection. The two CD tests, namely BP and modified BP provide \( p \)-values which are practically zero for both loss functions. Hence, we conclude that the loss differentials contain CD. Following the CD tests, we use the \( IC_{p1} \) which indicate that there are 6 common factors in the loss differential series. We therefore find out that both loss differential series display SCD and apply our tests robust to SCD.

**Panel tests for the EPA hypotheses: CPI forecasts.** The results are given in Table 13. As before, we report the tests robust to SCD as well as the results for the non-robust tests as a benchmark. We start the analysis with overall EPA tests and continue with clustered EPA tests. We consider 4 different country clusters in this application. These are G7, non-G7 OECD, OECD and non-OECD countries. We further divide the sample into pre- and post-global financial crises periods and compute the average loss differentials. The results for the average loss differentials of are given in Table 12. In the full sample with the absolute error loss, we see that the IMF does better than the RW for all clusters except the non-OECD countries. The global average is found to be positive which shows that the IMF does worse than the RW overall. The results are similar for the quadratic loss function except that the average loss is practically zero for G7 countries. In the pre-crisis period the overall differences are more pronounced between the IMF and the RW model.
In the post-crisis period however, the differences are very close to zero, especially for the quadratic loss. In what follows we use our test to check the significance of these averages.

First, in Panel (a) we see that with absolute loss function, all three overall EPA test statistics are statistically significant in 10% level. We also see here the effect of taking CD into account: with $S_{n,T}^{(1)}$, we can reject the overall EPA hypothesis in 5% level but this is not the case for the tests $S_{n,T}^{(3)}$ and $S_{n,T}^{(3)}$. We remind that the difference between the biases of IMF and RW forecasts is 0.45% which is reported in the last row of Table 12. We conclude that, overall, there is a small but statistically significant difference between the biases of IMF and RW forecasts in favor of the latter.

With quadratic loss we cannot reject the overall EPA hypothesis in conventional levels using any test. The question therefore is, if we can reject the clustered EPA hypothesis using the clusters we consider above. In the second block of the table, we have the results for the clustered EPA tests using two clusters: OECD and non-OECD countries. For these clusters, we can strongly reject the joint EPA hypothesis with both loss functions using the $C_{n,T}^{(1)}$. However, when we consider the CD in the loss differentials, the test statistics drop dramatically and they are insignificant. In the last block of the table, we have the results for three clusters: G7, non-G7 OECD, and non-OECD countries. Here, a similar picture arises such that we can reject the clustered EPA hypothesis with $C_{n,T}^{(1)}$ but this is not the case for $C_{n,T}^{(3)}$ and $C_{n,T}^{(3)}$.

In Panel (b), the results for the pre-crisis period are reported. The results obtained using the overall tests are similar to those of the full sample, that is we can reject the overall EPA hypothesis with the absolute loss function but this is not the case with the quadratic loss function. Similarly, with two country clusters (OECD and non-OECD countries), we can reject the clustered EPA hypothesis using only $C_{n,T}^{(1)}$ for both loss functions. When we consider the three country clusters however, we can reject the clustered EPA hypothesis with the absolute loss function using any test, at least in %10 significance level. To conclude, a significant difference between the biases of IMF and RW forecasts exists in the pre-crisis period.

As can be seen in Panel (c), the results are slightly different for the post-crisis period. First, the overall test statistics are negative for the absolute loss as IMF has less bias in this period. However, these overall differences are not statistically significant. When we look at the cluster based tests with two clusters, the differences are again statistically insignificant for any loss function. If we consider
the three country clusters, similar to the pre-crisis period, we can reject the clustered EPA hypothesis with the absolute loss function using any test. Quadratic loss differentials are insignificant as before. This is not surprising as we have found that the average loss differentials are very close to zero using quadratic loss, as is reported in Table 12.

Figure 3: Histogram of DM Test Statistics for the CPI Inflation Forecasts Country, 1991-2016 (IMF vs. Random Walk)
### Table 12: Average Loss Differentials for the CPI Inflation Forecasts of Different Country Clusters (IMF vs. Random Walk)

| Cluster          | Full Sample | 1991-2006 (Pre-crisis) | 2009-2019 (Post-crisis) |
|------------------|-------------|------------------------|------------------------|
|                  | Absolute Loss | Quadratic Loss | Absolute Loss | Quadratic Loss | Absolute Loss | Quadratic Loss |
| G7               | -0.0004      | 0.0000       | 0.0006      | 0.0000       | -0.0019     | -0.0001       |
| Non-G7 OECD      | -0.0033      | -0.0397      | -0.0063     | -0.0719      | 0.0002      | 0.0000        |
| OECD             | -0.0026      | -0.0301      | -0.0047     | -0.0546      | -0.0003     | 0.0000        |
| Non-OECD         | 0.0065       | 0.0134       | 0.0121      | 0.0243       | -0.0014     | -0.0002       |
| All              | 0.0045       | 0.0035       | 0.0083      | 0.0063       | -0.0012     | -0.0002       |
Table 13: Panel Tests of EPA for the CPI Inflation Forecasts (IMF vs. Random Walk)

| Test | Panel (a): Full Sample | Panel (b): 1991-2006 (Pre-crisis) | Panel (c): 2009-2019 (Post-crisis) |
|------|------------------------|----------------------------------|----------------------------------|
|      | Overall Tests          | Cluster 1: OECD                  | Cluster 1: G7                     |
|      | Cluster 1: Non-OECD    | Cluster 2: Non-G7 OECD            | Cluster 3: Non-OECD               |
|      |                        |                                  |                                  |
| $S_{n,T}^{(1)}$ | 2.04, 0.42             | 2.15, 0.43                       | -1.59, -0.82                     |
|      | (0.04)                 | (0.03)                           | (0.11)                           |
| $S_{n,T}^{(3)}$ | 1.70, 0.50             | 2.02, 0.51                       | -0.45, -0.40                     |
|      | (0.09)                 | (0.04)                           | (0.66)                           |
| $\Sigma_{n,T}^{(3)}$ | 1.78, 0.50             | 1.99, 0.51                       | -0.45, -0.40                     |
|      | (0.08)                 | (0.05)                           | (0.65)                           |
| $C_{n,T}^{(1)}$ | 8.16, 7.02             | 9.08, 7.29                       | 2.66, 2.25                       |
|      | (0.02), (0.03)         | (0.01), (0.03)                   | (0.26), (0.33)                   |
| $C_{n,T}^{(3)}$ | 3.23, 3.65             | 4.40, 4.09                       | 0.30, 0.23                       |
|      | (0.20), (0.16)         | (0.11), (0.13)                   | (0.86), (0.89)                   |
| $\Sigma_{n,T}^{(3)}$ | 3.44, 3.65             | 4.28, 4.08                       | 0.41, 0.24                       |
|      | (0.18), (0.16)         | (0.12), (0.13)                   | (0.82), (0.89)                   |

Note: The values shown in parentheses are $p$-values.
This figure "Fig_11.png" is available in "png" format from:

http://arxiv.org/ps/2003.02803v2
This figure "Fig_12.png" is available in "png" format from:

http://arxiv.org/ps/2003.02803v2
This figure "Fig_13.png" is available in "png" format from:

http://arxiv.org/ps/2003.02803v2
This figure "Fig_14.png" is available in "png" format from:

http://arxiv.org/ps/2003.02803v2
This figure "Fig_21.png" is available in "png" format from:

http://arxiv.org/ps/2003.02803v2
This figure "Fig_22.png" is available in "png" format from:

http://arxiv.org/ps/2003.02803v2
This figure "Fig_23.png" is available in "png" format from:

http://arxiv.org/ps/2003.02803v2
This figure "Fig_24.png" is available in "png" format from:

http://arxiv.org/ps/2003.02803v2
This figure "Fig_31.png" is available in "png" format from:

http://arxiv.org/ps/2003.02803v2
This figure "Fig_32.png" is available in "png" format from:

http://arxiv.org/ps/2003.02803v2