Equal predictive ability tests based on panel data with applications to OECD and IMF forecasts

Oguzhan Akgun [a], Alain Pirotte [b,*], Giovanni Urga [c,d], Zhenlin Yang [e]

[a] *LEDi, University of Burgundy, France*
[b] *University of Paris Pantheon-Assas, CRED, France*
[c] *Bayes Business School (formerly Cass), London, United Kingdom*
[d] *Bergamo University, Italy*
[e] *School of Economics, Singapore Management University, Singapore*

ARTICLE INFO

Keywords:
Cross-sectional dependence
Forecast evaluation
Hypothesis testing

ABSTRACT

We propose two types of equal predictive ability (EPA) tests with panels to compare the predictions made by two forecasters. The first type, *S*-statistics, focuses on *the overall EPA hypothesis*, which states that the EPA holds, on average, over all panel units and over time. The second type, *C*-statistics, focuses on *the clustered EPA hypothesis* where the EPA holds jointly for a fixed number of clusters of panel units. The asymptotic properties of the proposed tests are evaluated under weak and strong cross-sectional dependence. An extensive Monte Carlo simulation shows that the proposed 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.

© 2023 Published by Elsevier B.V. on behalf of International Institute of Forecasters.

1. Introduction

Formal tests of the null hypothesis of no difference in forecast accuracy that use 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 a few exceptions: Keane and Runkle (1990), Davies and Lahiri (1995) and Qu, Timmermann, and Zhu (2022, QTZ, hereafter).

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 $\widehat{y}_{l,it}$ be the forecast of agent $l = 1, 2$ made for the target value y_{it} of time $t = 1, 2, \ldots, T$ for unit (e.g., country, firm) $i = 1, 2, \ldots, n$. We propose tests for comparing the predictive ability of two forecasting agents, using n time series of loss differentials, $\Delta L_{it} = L(y_{it}; \widehat{y}_{1,it}) - L(y_{it}; \widehat{y}_{1,it})$, of length T, where $L(\cdot)$ is a generic loss function (for different loss functions in use, see Gneiting, 2011). 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. In contrast, 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 with respect to these two papers.

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. The statistics of this type, which we call *S*-statistics, are useful when a researcher

* Correspondence to: 21 rue Valette, 75005, Paris, France.
E-mail address: Alain.Pirotte@u-paris2.fr (A. Pirotte).

is not interested in the differences of predictive ability for a specific unit or clusters of units but in the overall differences. The statistics of the second type, *C*-statistics, focus on *the clustered EPA hypothesis*, which states that the EPA holds jointly for a fixed number of clusters of units in the panel. QTZ considers the same set of null hypotheses. Our work differs from QTZ in how the test statistics are constructed, most notably the statistics for testing the clustered EPA hypothesis. Both works set up a joint null hypothesis as "all cluster means are zero" (see, (6) in our paper and (11) in QTZ). However, we formulate multivariate test statistics in line with the joint hypotheses (the *C*-statistics in Section 3.2) with a limiting null distribution of χ^2_G, where *G* is the number of clusters. On the other hand, QTZ formulates a univariate test statistic (the J^D_n given in (12) in QTZ) based on the "average of the cluster averages of the loss differentials". If cluster sizes are equal, the numerator of J^D_n is equivalent to these of $J^{DM}_{n,T}$ of QTZ and *S*-statistics in our paper, tests for the overall EPA. As a result, if the overall EPA hypothesis holds, J^D_n fails to detect departure from the clustered EPA hypothesis. See the detailed discussion at the end of Section 3.2.

The applied literature 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, Lewis, Dang, Jin, and Richardson (2014) show that the economic growth projections of the OECD for 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, following a similar pattern, as we highlight later in this paper. The results of Pain et al. (2014) also indicate that the effect of these common shocks is heterogeneous across economies, and some country clusters exist.

Following these insights, we build 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 error 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, Pesaran, and Tosetti (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 & Prucha, 2007; Kim & Sun, 2013). In addition, we propose a novel partial sample variance estimator for large panels that deals with 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 & Ng, 2002). Following the insights of Driscoll and Kraay (1998), we also propose tests for the case of an unknown number of common factors contrary to the ones based on the PCE. This last approach is flexible in that it is robust under different types of CD. Moreover, contrary to the PCE-based tests, they do not rely on a linear factor

model. They are also straightforward to calculate, as in the case of overall EPA testing; the proposed test is identical to applying the Diebold and Mariano (1995) test to cross-sectional averages of the loss differentials.

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 converge in distribution to standard normal and chi-square with *G* degrees of freedom under the null of interest, respectively. The finite sample properties of the tests are examined via Monte Carlo simulations. The results show, in general, that our tests have very good finite sample performance. Specifically, we find that Driscoll and Kraay (1998) based overall and clustered EPA tests are robust to arbitrary CD in the loss differentials. However, they have lower power than the tests developed specifically for the case of WCD if loss differentials are only weakly cross-sectionally correlated. Moreover, if the loss differentials satisfy the linear approximate factor model assumption, they have lower power than the tests based on PCE.

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 between 1998 and 2016. In the second application, the quality of the IMF consumer price inflation (CPI) forecasts is challenged by comparing them with random walk forecasts. This data set contains 127 countries over the period 1991–2019. We find strong evidence of SCD in loss differentials of forecast errors in both applications. The results of the first application show that neither the overall EPA hypothesis nor clustered EPA hypothesis can be rejected for the economic growth forecasts of the OECD and the IMF. In contrast, in the second application, we find significant evidence against the overall and clustered EPA hypotheses on comparing 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. Section 5 gives guidelines to practitioners willing to test the EPA hypotheses using panels. In Section 6, the economic growth forecast errors of the OECD and the IMF are compared using the proposed tests. Section 7 concludes. Three appendices contain additional justification of our testing framework, proofs of results, and the details of the second application comparing the IMF CPI forecasts with random walk forecasts.

Notation. Let $\mathbf{A} = [a_{ij}]$ be an $n \times n$ matrix. The column and row norms of $\mathbf{A}$ are $\|\mathbf{A}\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|$ and $\|\mathbf{A}\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|$, respectively. The Euclidean norm of an $n \times m$ matrix $\mathbf{B}$ is $\|\mathbf{B}\| = [\text{Tr}(\mathbf{B}'\mathbf{B})]^{1/2}$. *M* is a finite positive constant. $\overset{d}{\to}$ and $\overset{p}{\to}$ denote convergence in distribution and probability, respectively. Joint passage to infinity of *T* and *n* is denoted by $(T, n) \to \infty$.

2. Testing framework and the 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 differentials, $\Delta L_{it} = L(y_{it}; \widehat{y}_{1,it}) - L(y_{it}; \widehat{y}_{1,it})$, are potentially cross-sectionally correlated. The CD may be of strong or weak types or may even be generated by two distinct components of these two different types. To simplify the distinction, we assume the following structure on the loss differentials:

$$\Delta L_{it} = \mu_i + \boldsymbol{\lambda}_i' \mathbf{f}_t + \varepsilon_{it},$$

$$\varepsilon_{it} = \sum_{j=1}^{n} r_{ij} \epsilon_{jt}, \tag{1}$$

where $\mathbf{f}_t = (f_{1t}, f_{2t}, \ldots, f_{mt})'$ is an $m \times 1$ vector of unobservable common factors and $\boldsymbol{\lambda}_i = (\lambda_{1i}, \lambda_{2i}, \ldots, \lambda_{mi})'$ is an $m \times 1$ vector of fixed factor loadings. The coefficients r_{ij} are fixed, unknown elements of an $n \times n$ matrix $\mathbf{R}_n$. We impose $r_{ii} = 1$ for normalization of the variances. $\mathbf{f}_t$ and ϵ_{it} are assumed to be zero mean weakly stationary time series allowed to be autocorrelated through time. In addition, ϵ_{it} is cross-sectionally uncorrelated. The unit-specific means satisfy $|\mu_i| < \infty$.

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

$$\lim_{n \to \infty} \mathrm{Var}\left(n^{-1} \sum_{i=1}^{n} \varepsilon_{it} \right) = \lim_{n \to \infty} \left(n^{-2} \sum_{i,j=1}^{n} \mathbf{r}_{i.}' \boldsymbol{\gamma}_{n,0} \mathbf{r}_{j.} \right) = 0, \tag{2}$$

where $\mathbf{r}_{i.} = (r_{i1}, r_{i2}, \ldots, r_{in})'$, $\boldsymbol{\gamma}_{n,0} = \mathrm{diag}(\gamma_{11,0}, \ldots, \gamma_{nn,0})$ and $\gamma_{ii,0} = \mathrm{E}(\epsilon_{it}^2)$. For this to be true, it requires the row and column sum norms of $\mathbf{R}_n$ to be bounded; see Assumption 3 below. The common component in (1) is assumed to induce SCD such that the variance of its cross-sectional average is bounded away from zero:

$$\mathrm{Var}\left(n^{-1} \sum_{i=1}^{n} \boldsymbol{\lambda}_i' \mathbf{f}_t \right) = n^{-2} \sum_{i,j=1}^{n} \boldsymbol{\lambda}_i' \boldsymbol{\Gamma}_0 \boldsymbol{\lambda}_j > 0, \quad \text{for all } n, \tag{3}$$

where $\boldsymbol{\Gamma}_0 = \mathrm{E}(\mathbf{f}_t \mathbf{f}_t')$. The conditions under which this holds true are given below, in Assumption 6. The WCD can be modeled 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).

In this paper, we closely follow 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 directly concern the loss differentials. Forecast errors can be non-stationary and induce CD. The properties that are carried into the loss differentials are important, not the properties of the forecast errors specifically. As an example, in Appendix A, we show how the specification in (1) 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 the DM approach, any other (symmetric 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 & Ng, 2002).

The null hypotheses of EPA and the alternatives. The first hypothesis of interest is

$$H_0^a : \bar{\mu} = 0, \tag{4}$$

where $\bar{\mu} = \lim_{n \to \infty} \bar{\mu}_n$ with $\bar{\mu}_n = n^{-1} \sum_{i=1}^{n} \mu_i$. This hypothesis, also considered by QTZ, 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 in 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*. The alternative hypothesis, in this case, is that overall EPA does not hold:

$$H_1^a : \bar{\mu} \neq 0. \tag{5}$$

Of course, it is also possible to consider one-sided alternatives.

In a macro forecasting study, differences between clusters of units can have 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:

$$H_0^{ca} : \bar{\mu}_g = 0, \quad \text{for all } g = 1, 2, \ldots, G, \tag{6}$$

where $\bar{\mu}_g = \lim_{n_g \to \infty} \bar{\mu}_{g,n_g}$ with $\bar{\mu}_{g,n_g} = n_g^{-1} \sum_{i \in G_g} \mu_i$ and G_g is the set of indexes of n_g cross-sectional units which belong to cluster g. This paper refers to this hypothesis as *the clustered EPA hypothesis*. The alternative hypothesis associated is the following:

$$H_1^{ca} : \bar{\mu}_g \neq 0, \quad \text{for at least one } g = 1, 2, \ldots, G. \tag{7}$$

QTZ considers the null hypothesis H_0^{ca} without stating an alternative hypothesis. Its test against H_0^{ca} is inconsistent in particular cases; for instance, the clustered EPA test fails to reject H_0^{ca} if H_0^a and H_1^{ca} hold simultaneously. This issue is discussed theoretically below in Section 3.2. The difference between the two hypotheses is therefore important, and the choice depends on the specific interest.

Table 1
Test statistics proposed.

	Overall EPA tests S-statistics	Clustered EPA tests C-statistics
No cross-sectional dependence	$S_{nT}^{(1)}$	$C_{nT}^{(1)}$
Weak cross-sectional dependence		
Distance based tests	$S_{nT}^{(2)}$	$C_{nT}^{(2)}$
Tests with unknown distance	$\underline{S}_{nT}^{(2)}$	$\underline{C}_{nT}^{(2)}$
Strong cross-sectional dependence		
No restriction on CD	$S_{nT}^{(3)}, \tilde{S}_{nT}^{(3)}$	$C_{nT}^{(3)}$
Factor model-based tests	$\underline{S}_{nT}^{(3)}$	$\underline{C}_{nT}^{(3)}$

The overall EPA may hold even if the two forecasters have different predictive abilities for different clusters. This occurs if the average loss differentials of different country clusters are different from zero (that is H_1^{ca} holds), but they add up to zero when pooled (that is H_0^a holds). If these clusters have an economic meaning, testing the clustered EPA hypothesis is important.

We propose test statistics for these two null hypotheses under different assumptions on the CD structure in the loss differentials. These are No CD, WCD, or SCD. The main difference between the proposed test statistics is the methods of calculating the asymptotic variance depending on this structure. These test statistics are summarized in Table 1. Most of our test statistics do not require the approximate factor model given in (1) with the exceptions being $\underline{S}_{nT}^{(3)}$ and $\underline{C}_{nT}^{(3)}$. Nevertheless, we retain the factor model to simplify the discussion of different types of CDs.

3. The test statistics and their asymptotic properties

3.1. Tests for overall equal predictive ability: S-statistics

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 (4) 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 positive, non-random bandwidth parameters. In all cases, the limiting null distribution of the test statistics is obtained under Assumptions 1 and 2.

Assumption 1. ϵ_{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, $E|\psi_{it}|^4 < \infty$, $\max_{1 \le i \le n} \sum_{h=0}^{\infty} h|c_{ih}| < M < \infty$ and $\sum_{h=0}^{\infty} c_{ih} > 0$.

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

Assumption 1 is sufficient to obtain a CLT for the sample mean of ϵ_{it} for each i. Assumption 2(a) is standard in time series literature. Assumption 2(b) is a high-level

assumption used to estimate long-run variance for each i consistently. 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 $\lambda_i' \mathbf{f}_t = 0$ for all i, t and $r_{ij} = 0$ for all $i \ne j$. In order to test the null hypothesis H_0^a, we propose to use the following statistic:

$$S_{nT}^{(1)} = \frac{\sqrt{nT} \, \Delta \bar{L}_{nT}}{\hat{\sigma}_{1,nT}}, \tag{8}$$

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

Proposition 1. *Suppose ΔL_{it} follows the model in (1) with $\lambda_i' \mathbf{f}_t = 0$ for all i, t, $r_{ij} = 0$ for all $i \ne j$, and Assumptions 1–2 hold. Then, under H_0^a and as $(T, n) \to \infty$, $S_{nT}^{(1)} \xrightarrow{D} N(0, 1)$.*

To prove this result, it is sufficient to show that $\sqrt{nT}(\Delta \bar{L}_{nT} - \bar{\mu}_n)/\sigma_{1,nT} \xrightarrow{d} N(0, 1)$ where $\sigma_{1,nT}^2 = n^{-1} \sum_{i=1}^{n} \bar{\gamma}_{i,T}$ with $\bar{\gamma}_{i,T} = T^{-1} \sum_{t,s=1}^{T} \gamma_{i,d_{ts}}$, $\gamma_{i,d_{ts}} = E(\epsilon_{it} \epsilon_{is})$, and $\hat{\sigma}_{1,nT}^2 - \sigma_{1,nT}^2 \xrightarrow{p} 0$. It is easy to prove that the test statistic is divergent, hence consistent, under H_1^a. To see this, note that in this case $\Delta \bar{L}_{nT} \xrightarrow{p} \bar{\mu} \ne 0$, $\hat{\sigma}_{1,nT}^2 \xrightarrow{p} \sigma_1^2 > 0$ where $\sigma_1^2 = \lim_{n,T \to \infty} \sigma_{1,nT}^2$. It follows that $P[|S_{nT}^{(1)}| > c] \to 1$ for any constant $c \in \mathbb{R}$ as $(T, n) \to \infty$.

Overall EPA tests under WCD. Suppose $\lambda_i' \mathbf{f}_t = 0$ for all i, t but r_{ij} is not necessarily zero for all $i \ne j$. In this case, the loss differentials ΔL_{it} are no longer independent across i. Define $d_{ij} = d_{ji} \ge 0$ as the distance between units i and j. We make Assumptions 3 and 4 on the coefficients r_{ij}.

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

Assumption 4. $\sum_{j=1}^{n} |\mathbf{r}_{i.}' \mathbf{r}_{j.}| d_{ij}^\rho < \infty$ for some $\rho \ge 1$.

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, the correlation between two

panel units decreases as the distance increases. This, in turn, sets a basis for using the spatial kernel function $k_S(\cdot)$, which gives smaller weights to the covariance between units more distant from each other.

To deal with the WCD when $T = 1$, Kelejian and Prucha (2007) 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), and d_n a sequence of positive, non-random bandwidth parameters. 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 $n \to \infty$.

The condition in 5(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 5(b) are used by Kelejian and Prucha (2007), Moscone and Tosetti (2012) and Kim and Sun (2013). All these studies allow solely for kernels that truncate, i.e., those that equal 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 5(c) controls the expansion of d_n relative to n. This condition, as well as 5(b), is not necessary for our study's consistent estimation of the variance–covariance matrices. Nevertheless, these conditions are retained to compare our results with the existing literature.

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

$$S_{nT}^{(2)} = \frac{\sqrt{nT}\,\Delta\bar{L}_{nT}}{\hat{\sigma}_{2,nT}}, \tag{9}$$

where

$$\hat{\sigma}_{2,nT}^2 = \frac{1}{nT} \sum_{i,j=1}^n \sum_{t,s=1}^T k_S\left(\frac{d_{ij}}{d_n}\right) k_T\left(\frac{d_{ts}}{d_T}\right) \Delta\tilde{L}_{it}\,\Delta\tilde{L}_{js}. \tag{10}$$

The 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}, must 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

$$\underline{\hat{\sigma}}_{2,nT}^2 = \frac{1}{\underline{n}T} \sum_{i,j=1}^{\underline{n}} \sum_{t,s=1}^T k_T\left(\frac{d_{ts}}{d_T}\right) \Delta\tilde{L}_{it}\,\Delta\tilde{L}_{js}, \tag{11}$$

where $\underline{n}$, 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 deals with panel regression models with small T. Our variance estimator generalizes that of Moscone and Tosetti (2015) by allowing for large T with the help of a time kernel. The corresponding test statistic is given by

$$\underline{S}_{nT}^{(2)} = \frac{\sqrt{nT}\,\Delta\bar{L}_{nT}}{\underline{\hat{\sigma}}_{2,nT}}.$$

Proposition 2. *Suppose ΔL_{it} follows the model in (1) with $\lambda_i' \mathbf{f}_t = 0$ for all i, t and Assumptions 1–5 hold. Then, under H_0^a and as $(T, n) \to \infty$, (i) $S_{nT}^{(2)} \overset{d}{\to} N(0, 1)$ if $d_n \to \infty$, (ii) and $\underline{S}_{nT}^{(2)} \overset{d}{\to} N(0, 1)$ if $\underline{n} \to \infty$ such that $\underline{n}/n \to \kappa \in [0, 1]$.*

Under Assumptions 1–5, $\sqrt{nT}(\Delta\bar{L}_{nT} - \bar{\mu}_n)/\sigma_{2,nT} \overset{d}{\to} N(0, 1)$ as $(T, n) \to \infty$, where $\sigma_{2,nT}^2 = n^{-1} \sum_{i=1}^n \mathbf{r}_{i.}' \bar{\gamma}_{nT} \mathbf{r}_{i.}$ with $\bar{\gamma}_{nT} = T^{-1} \sum_{t,s=1}^T \gamma_{n,d_{ts}}$ and $\gamma_{n,d_{ts}} = \text{diag}(\gamma_{1,d_{ts}}, \gamma_{2,d_{ts}}, \ldots, \gamma_{n,d_{ts}})$. Hence, the stated results follow by the consistency of the variance estimators. Similar arguments made in the case of no CD lead to the consistency of the test statistics in the case in which the alternative H_1^a holds.

The test statistic $\underline{S}_{nT}^{(2)}$ is an asymptotic one, which, in practice, can be hard to implement. Despite its advantage of not relying on a distance metric, it needs special care as there are no clear indications of which cross-sectional units to use in the calculation. A solution in practice is to fix $\underline{n}$, calculate the test statistic for a large number of subsamples of cross-sectional units of size $\underline{n}$, and to take the smallest test statistic. This provides a conservative test statistic that would be correctly sized but potentially with low power. Nevertheless, below, we consider another test statistic, $S_{nT}^{(3)}$, which is robust to WCD and does not pose similar problems.

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 Assumption 6 on the common factors, their loadings, and the error terms.

Assumption 6. (a) $\mathbf{f}_t$ is independent of ε_{it} and follows the linear process $\mathbf{f}_t = \sum_{h=0}^\infty \mathbf{C}_h \Psi_{t-h}$, with $\Psi_t \sim iid(0, \mathbf{I}_m)$, $\mathbb{E}\|\Psi_t\|^4 < \infty$, $\sum_{h=0}^\infty h\|\mathbf{C}_h\| < M < \infty$ and $\sum_{h=0}^\infty \mathbf{C}_h$ is full rank. (b) Factor loadings λ_i are fixed parameters such that $\|\lambda_i\| < \infty$, $n^{-1} \sum_{i=1}^n \lambda_i \lambda_i' \to \Sigma_\lambda > 0$ for an $m \times m$ matrix Σ_λ. (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 $n \in \mathbb{Z}^+$.

It is easy to see that, $\lambda_i' \mathbf{f}_t$ 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 factor contributes to the cross-sectional averages' asymptotic variance. Although nonstandard, this assumption is not

restrictive as it does not affect the validity of our testing procedures.

In this case of SCD, the variance estimator given in (10) can be modified by setting $k_S(\cdot) = 1$. This variance estimator does not require knowledge of a distance measure between the units. Moreover, it assigns weights equal to one for all covariances from the same time period; hence it is robust to SCD and WCD. The test statistic takes the form:

$$S_{nT}^{(3)} = \frac{\sqrt{T}\Delta\bar{L}_{nT}}{\hat{\sigma}_{3,nT}}, \tag{12}$$

where

$$\hat{\sigma}_{3,nT}^2 = \frac{1}{n^2 T} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} k_T\left(\frac{d_{ts}}{d_T}\right) \Delta\tilde{L}_{it}\Delta\tilde{L}_{js}. \tag{13}$$

The variance estimator (13) is valid when T is large, regardless of n being finite or infinite (see Driscoll & Kraay, 1998). The special case of this test statistic where k_T is the Bartlett kernel corresponds to the test statistic $J_{n,T}^{DM}$ of QTZ. As the authors note, an important advantage of this test statistic is that it does not rely on the linear approximate factor model in (1). It is robust even under DGPs such as $\Delta L_{it} = \mu_i + g(\lambda_i, \mathbf{f}_t) + \varepsilon_{it}$ where $g(\cdot, \cdot)$ is a non-parametric function. Furthermore, the test statistic $S_{nT}^{(3)}$ is very easy to calculate as it is identical to the DM test statistic for the cross-sectional averages of loss differentials. To see this, it suffices to write $\Delta\bar{L}_{nT} = T^{-1}\sum_{t=1}^{T}\bar{L}_{n,t}$ and $\hat{\sigma}_{3,nT}^2 = T^{-1}\sum_{t,s=1}^{T} k_T(d_{ts}/d_T)\Delta\tilde{L}_{n,t}\Delta\tilde{L}_{n,s}$, where $\bar{L}_{n,t} = n^{-1}\sum_{i=1}^{n}\Delta L_{it}$ and $\Delta\tilde{L}_{n,t} = \bar{L}_{n,t} - \Delta\bar{L}_{nT}$.

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 robust to WCD and autocorrelation in error terms. It is obtained by minimizing the average squared residuals computed for m common factors:

$$V(m) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (\Delta\tilde{L}_{it} - \lambda_i'\mathbf{f}_t)^2 \tag{14}$$

subject to $Var(\mathbf{f}_t) = \mathbf{I}_m$ and $\Lambda_n'\Lambda_n$ being diagonal where $\Lambda_n = (\lambda_1, \lambda_2, \ldots, \lambda_n)'$. Then the solution for the estimates of the common factors, $\hat{\mathbf{f}}_t$, are given by $\sqrt{T}$ times the first m eigenvectors of the matrix $\sum_{i=1}^{n}\Delta\mathbf{L}_{i.}\Delta\mathbf{L}_{i.}'$ with $\Delta\mathbf{L}_{i.} = (\Delta\tilde{L}_{i1}, \Delta\tilde{L}_{i2}, \ldots, \Delta\tilde{L}_{iT})'$ and the factor loadings can be estimated as $\hat{\lambda}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\mathbf{f}}_t\Delta\tilde{L}_{it}$. We make the following high-level assumption on the asymptotic properties of the PC estimates.

Assumption 7. $\delta_{nT}(\widehat{\lambda_i'\mathbf{f}_t} - \lambda_i'\mathbf{f}_t) = O_p(1)$ where $\delta_{nT} = \min(\sqrt{T}, \sqrt{n})$.

This is a standard result in the literature on the approximate factor models. The conditions under which it holds are given in Bai (2003) (see their Theorem 3). Then the overall EPA hypothesis can be tested using

$$\underline{S}_{nT}^{(3)} = \frac{\sqrt{T}\Delta\bar{L}_{nT}}{\hat{\underline{\sigma}}_{3,nT}}, \tag{15}$$

where

$$\hat{\underline{\sigma}}_{3,nT}^2 = \frac{1}{n^2 T} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} k_T\left(\frac{d_{ts}}{d_T}\right) \widehat{\lambda_i}\widehat{\mathbf{f}_t}\widehat{\mathbf{f}_s}\widehat{\lambda_j}$$

$$+ \frac{1}{n^2 T} \sum_{i=1}^{n} \sum_{t,s=1}^{T} k_T\left(\frac{d_{ts}}{d_T}\right) \widehat{\varepsilon}_{it}\widehat{\varepsilon}_{is}, \tag{16}$$

with $\widehat{\varepsilon}_{it} = \Delta\tilde{L}_{it} - \widehat{\lambda_i}\widehat{\mathbf{f}_t}$.

Proposition 3. *Suppose ΔL_{it} follows the model in (1), Assumptions 1–3, 6 and 7 hold. Then, under H_0^a and as $(T, n) \to \infty$, (i) $S_{nT}^{(3)} \overset{d}{\to} N(0, 1)$, (ii) and $\underline{S}_{nT}^{(3)} \overset{d}{\to} N(0, 1)$.*

Under Assumptions 1–3 and 6, $\sqrt{T}(\Delta\bar{L}_{nT} - \bar{\mu}_n)/\sigma_{3,nT} \overset{d}{\to} N(0, 1)$, where

$$\sigma_{3,nT}^2 = \frac{1}{n^2} \sum_{i,j=1}^{n} \left(\lambda_i'\bar{\Gamma}_T\lambda_j + \mathbf{r}_{i.}'\bar{\gamma}_{nT}\mathbf{r}_{j.}\right), \tag{17}$$

with $\bar{\Gamma}_T = T^{-1}\sum_{t,s=1}^{T}\Gamma_{dts}$. 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). We can show the consistency of the test statistics using similar arguments to the ones made earlier, after Propositions 1 and 2. It can be easily shown that the variance estimator $\hat{\sigma}_{3,nT}^2$ is consistent under no CD as well as WCD. Hence the test statistic $S_{nT}^{(3)}$ is consistent in these cases as well. However, the divergence rate is smaller than those of Propositions 1 and 2. Thus, the test is expected to have a lower power against the null H_0^a.

In (17), the first term in parentheses dominates the second one. This is because, under Assumption 1, 3 and 6, $n^{-2}\sum_{i,j=1}^{n}\lambda_i'\bar{\Gamma}_T\lambda_j = O(1)$ but $n^{-2}\sum_{i,j=1}^{n}\mathbf{r}_{i.}'\bar{\gamma}_{nT}\mathbf{r}_{j.} = O(1/n)$. Hence, the latter is asymptotically negligible. This means that under SCD, one can use a simpler variance estimator without the second term in (16).

When T is fixed, but n is large, the test statistic $S_{nT}^{(3)}$ is not valid as it relies on the T-asymptotics. However, under additional assumptions on the DGP of the loss differentials, we can obtain a test statistic for fixed T. Specifically, assume that the loss differentials are based on optimal one-step ahead forecasts in the sense that they are serially uncorrelated. Consider the following test statistic:

$$\tilde{S}_{nT}^{(3)} = \frac{\sqrt{T}\Delta\bar{L}_{nT}}{\tilde{\sigma}_{3,nT}}, \tag{18}$$

where

$$\tilde{\sigma}_{3,nT}^2 = \frac{1}{T-1} \sum_{t=1}^{T} \Delta\tilde{L}_{n,t}^2. \tag{19}$$

We have the following result:

Corollary 1. *Suppose that the loss differentials ΔL_{it} follow the model in (1), they are serially uncorrelated, and one of the following holds: (a) $\lambda_i'\mathbf{f}_t = 0$ for all i, t and Assumptions 1–5 hold; (b) Assumptions 1–3, 6 and 7 hold and $\epsilon_{it} \sim N(0, v_i^2)$. Then, under H_0^a and as $n \to \infty$, $\tilde{S}_{nT}^{(3)} \overset{d}{\to} t(T-1)$.*

Here $t(T-1)$ represents the Student's t-distribution with $T-1$ degrees of freedom. The result shows that, under additional assumptions to those of Propositions 2 and 3, we can use the Student's t critical values after adjusting the degrees of freedom in the test statistic $S_{nT}^{(3)}$. If the loss differentials carry SCD, we need the normality of the error terms for the result to hold. If they are weakly cross-sectionally correlated, the normality assumption is not required as under the WCD assumption, cross-sectional averages of the loss differentials admit a CLT, i.e., they are normally distributed for each $t = 1, 2, \ldots, T$.

3.2. Tests for clustered equal predictive ability: C-statistics

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^{ca}, which can be written as

$$H_0^{ca} : \bar{\boldsymbol{\mu}} = \mathbf{0}, \tag{20}$$

where $\bar{\boldsymbol{\mu}} = (\bar{\mu}_1, \bar{\mu}_2, \ldots, \bar{\mu}_G)'$. 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 $\Delta \bar{L}_{g,n_g,T} = (n_g T)^{-1} \sum_{i \in G_g} \sum_{t=1}^{T} \Delta L_{it}$. We assume that the sets of indexes G_g, $g = 1, \ldots, G$ are known. Assumption 8 is placed to control the asymptotic number of units per cluster.

Assumption 8. For all $G > 1$, 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 $\boldsymbol{\lambda}_i' \mathbf{f}_t = 0$ for all i, t and $r_{ij} = 0$ for all $i \neq j$. We propose the following statistic to test the hypothesis in (20):

$$C_{nT}^{(1)} = nT \Delta \bar{\mathbf{L}}_{nT}' \widehat{\boldsymbol{\Omega}}_{1,nT}^{-1} \Delta \bar{\mathbf{L}}_{nT}, \tag{21}$$

where

$$\widehat{\boldsymbol{\Omega}}_{1,nT} = \frac{1}{T} \sum_{i=1}^{n} \frac{n}{n_{g_i}^2} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}_{g_i}' \Delta \tilde{L}_{it} \Delta \tilde{L}_{is},$$

with $g_i \in \{1, 2, \ldots, G\}$ being a variable which states the cluster which ith unit belongs to and $\boldsymbol{\iota}_{g_i}$ being the g_ith column of $\mathbf{I}_G$. Notice that $\widehat{\boldsymbol{\Omega}}_{1,nT}$ is a diagonal matrix that contains an estimate of the average long-run variances of each cluster as diagonal elements up to a factor which is asymptotically equal to τ_g^{-1}.

Proposition 4. *Suppose ΔL_{it} follows the model in (1) with $\boldsymbol{\lambda}_i' \mathbf{f}_t = 0$ for all i, t, $r_{ij} = 0$ for all $i \neq j$, and Assumption 1, 2 and 8 hold. Then, under H_0^{ca} and as $(T, n) \to \infty$, $C_{nT}^{(1)} \xrightarrow{d} \chi_G^2$.*

Under Assumption 1, 2 and 8, we have $\sqrt{nT} \boldsymbol{\Omega}_{1,nT}^{-1/2} (\Delta \bar{\mathbf{L}}_{nT} - \bar{\boldsymbol{\mu}}_n) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_G)$ as $(T, n) \to \infty$, where $\bar{\boldsymbol{\mu}}_n =$

$(\bar{\mu}_{1,n_1}, \bar{\mu}_{2,n_2}, \ldots, \bar{\mu}_{G,n_G})'$ and $\boldsymbol{\Omega}_{1,nT} = \sum_{i=1}^{n} \frac{n}{n_{g_i}^2} \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}_{g_i}' \bar{\gamma}_{i,T}$. This is a generalization of the CLT following Proposition 1, which is obtained when $G = 1$. It is easy to see that the test is consistent under the alternative hypothesis H_1^{ca}. This is because, under the assumptions of the proposition, the matrix $\widehat{\boldsymbol{\Omega}}_{1,nT}$ tends to a positive definite matrix as $(T, n) \to \infty$. However, at least one entry of the vector $\Delta \bar{\mathbf{L}}_{nT}$ tends to a non-zero constant provided that there exists a $g \in \{1, \ldots, G\}$ for which $\bar{\mu}_g \neq 0$. Furthermore, the test is consistent against H_0^a as this hypothesis fails if any cluster has a non-zero mean. However, we expect this test to have lower power against H_0^a compared to $S_{nT}^{(1)}$.

Clustered EPA tests under WCD. Suppose $\boldsymbol{\lambda}_i' \mathbf{f}_t = 0$ for all i, t but r_{ij} is not necessarily zero for all $i \neq j$. We can use the following statistic in order to test H_0^{ca}:

$$C_{nT}^{(2)} = nT \Delta \bar{\mathbf{L}}_{nT}' \widehat{\boldsymbol{\Omega}}_{2,nT}^{-1} \Delta \bar{\mathbf{L}}_{nT}, \tag{22}$$

where

$$\widehat{\boldsymbol{\Omega}}_{2,nT} = \frac{1}{T} \sum_{i,j=1}^{n} \frac{n}{n_{g_i} n_{g_j}} k_S \left(\frac{d_{ij}}{d_n} \right) \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}_{g_j}' \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}.$$

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 (11). Such an estimator is

$$\underline{\widehat{\boldsymbol{\Omega}}}_{2,nT} = \frac{1}{\underline{n}T} \sum_{i=1}^{\underline{n}_{g_i}} \sum_{j=1}^{\underline{n}_{g_j}} \frac{n^2}{n_{g_i} n_{g_j}} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}_{g_j}' \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}, \tag{23}$$

where $\underline{n}_g$ is the number of observations taken into account in the calculation of variance in cluster g, hence we have $\underline{n} = \sum_{g=1}^{G} \underline{n}_g$. The corresponding test statistic is

$$\underline{C}_{nT}^{(2)} = nT \Delta \bar{\mathbf{L}}_{nT}' \underline{\widehat{\boldsymbol{\Omega}}}_{2,nT}^{-1} \Delta \bar{\mathbf{L}}_{nT}. \tag{24}$$

Theorem 1. *Suppose ΔL_{it} follows the model in (1) with $\boldsymbol{\lambda}_i' \mathbf{f}_t = 0$ for all i, t and Assumptions 1–5 and 8 hold. Then, under H_0^{ca} and as $(T, n) \to \infty$, (i) $C_{nT}^{(2)} \xrightarrow{d} \chi_G^2$ if $d_n \to \infty$, (ii) and $\underline{C}_{nT}^{(2)} \xrightarrow{d} \chi_G^2$ if as $\underline{n}_g \to \infty$, $\forall g = 1, \ldots, G$ such that $\underline{n}_g/n_g \to \kappa \in [0, 1]$.*

Under the assumptions of the theorem, we have $\sqrt{nT} \boldsymbol{\Omega}_{2,nT}^{-1/2} (\Delta \bar{\mathbf{L}}_{nT} - \bar{\boldsymbol{\mu}}_n) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_G)$ as $(T, n) \to \infty$, where $\boldsymbol{\Omega}_{2,nT} = \sum_{i,j=1}^{n} \frac{n}{n_{g_i} n_{g_j}} \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}_{g_j}' \mathbf{r}_i' \bar{\gamma}_{nT} \mathbf{r}_j$. 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 $\mathbf{R}_n = \mathbf{I}_n$ and $G = 1$, the second is obtained with only $G = 1$, and the last is when only $\mathbf{R}_n = \mathbf{I}_n$. The consistency of the tests follows from similar arguments to the ones above, those under Proposition 4.

Clustered EPA tests under SCD. We now consider the case of no specific restrictions on the CD in the loss differentials. With SCD, the clustered EPA hypothesis H_0^{ca} can be tested using:

O. Akgun, A. Pirotte, G. Urga et al.

$$C_{nT}^{(3)} = T \Delta \bar{\mathbf{L}}_{nT}' \widehat{\boldsymbol{\Omega}}_{3,nT}^{-1} \Delta \bar{\mathbf{L}}_{nT}, \tag{25}$$

$$\widehat{\boldsymbol{\Omega}}_{3,nT} = \frac{1}{T} \sum_{i,j=1}^{n} \frac{1}{n_{g_i} n_{g_j}} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}_{g_j}' \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}.$$

As discussed previously, this type of variance estimator is robust to arbitrary CD and has the advantage of not requiring known distances between units. Furthermore, like $S_{nT}^{(3)}$, it is straightforward to calculate as it is identical to a Wald test applied to within-cluster cross-sectional averages where between cluster covariances as well as serial correlations are taken into account. 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

$$\underline{C}_{nT}^{(3)} = T \Delta \bar{\mathbf{L}}_{nT}' \underline{\widehat{\boldsymbol{\Omega}}}_{3,nT}^{-1} \Delta \bar{\mathbf{L}}_{nT}, \tag{26}$$

$$\underline{\widehat{\boldsymbol{\Omega}}}_{3,nT} = \frac{1}{T} \sum_{i,j=1}^{n} \frac{1}{n_{g_i} n_{g_j}} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}_{g_j}' \widehat{\boldsymbol{\lambda}}_i' \widehat{\mathbf{f}}_t \widehat{\mathbf{f}}_s' \widehat{\boldsymbol{\lambda}}_j$$

$$+ \frac{1}{T} \sum_{i=1}^{n} \frac{1}{n_{g_i}^2} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}_{g_i}' \widehat{\varepsilon}_{it} \widehat{\varepsilon}_{is}. \tag{27}$$

Theorem 2. *Suppose ΔL_{it} follows the model in* (1), *Assumptions* 1–3, 6–8 *hold. Then, under H_0^{ca} and as $(T, n) \rightarrow \infty$,* (i) $C_{nT}^{(3)} \xrightarrow{d} \chi_G^2$, (ii) *and* $\underline{C}_{nT}^{(3)} \xrightarrow{d} \chi_G^2$.

Under the assumptions of the theorem, we have $\sqrt{T} \boldsymbol{\Omega}_{3,nT}^{-1/2} (\Delta \bar{\mathbf{L}}_{nT} - \bar{\boldsymbol{\mu}}_n) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_G)$, where $\boldsymbol{\Omega}_{3,nT} = \sum_{i,j=1}^{n} \frac{1}{n_{g_i} n_{g_j}} \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}_{g_j}' (\boldsymbol{\lambda}_i' \bar{\boldsymbol{\Gamma}}_T \boldsymbol{\lambda}_j + \mathbf{r}_{i.}' \bar{\boldsymbol{\gamma}}_{nT} \mathbf{r}_{j.})$. Proposition 3 is a special case of this theorem with $G = 1$. It is easy to prove the consistency of the tests following arguments similar to the ones under Proposition 4 and Theorem 1. Similar to the discussion following Proposition 3, $C_{nT}^{(3)}$ is expected to be consistent under WCD but have lower power than $C_{nT}^{(2)}$ in this case.

Furthermore, it is easy to see that the test is consistent against H_0^a, following similar arguments of the discussion of Proposition 4. However, it is important to note that none of the overall EPA test statistics is consistent under general alternatives of H_0^{ca}. To see this, suppose that $G = 2$ with $\bar{\mu}_{1,n_1} \rightarrow \mu^0 \neq 0$, $\bar{\mu}_{2,n_2} \rightarrow -\mu^0$ as $n \rightarrow \infty$. Then $\bar{\mu}_n \rightarrow 0$ as $n \rightarrow \infty$; hence, the overall EPA hypothesis holds asymptotically while the clustered EPA hypothesis fails. Therefore, the S-statistics follow their asymptotic null distributions under their respective assumptions, as stated in Propositions 1–3 and Corollary 1.

Comparison of C-statistics and J_n^D of QTZ. QTZ proposes a test statistic for the clustered EPA hypothesis of the form

$$J_n^D = \frac{\sqrt{G} \bar{D}}{\sqrt{(G-1)^{-1} \sum_{g=1}^{G} (D_g - \bar{D})^2}},$$

where $D_g = \frac{1}{\sqrt{n_g T}} \sum_{i \in G_g} \sum_{t=1}^{T} \Delta L_{it}$ and $\bar{D} = G^{-1} \sum_{g=1}^{G} D_g$ and they suggest to reject the null H_0^{ca} if $|J_n^D| > t_{G-1,1-\alpha/2}$ where $t_{G-1,1-\alpha}$ is the $1 - \alpha$ quantile of the $t(G-1)$

distribution. To see how J_n^D is related to the test statistics that we propose, let us first suppose that $n_g = n/G$ for all $g = 1, \ldots, G$, that is, the cluster sizes are identical. Then, the numerator of J_n^D is equal to $\sqrt{nT} \Delta \bar{L}_{nT}$, the numerator of our S-statistics, and that of $J_{n,T}^{DM}$ of QTZ:

$$\sqrt{G} \bar{D} = \sqrt{G} \left(G^{-1} \sum_{g=1}^{G} \frac{1}{\sqrt{n_g T}} \sum_{i \in G_g} \sum_{t=1}^{T} \Delta L_{it} \right)$$

$$= \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \sqrt{n_g T} \frac{1}{n_g T} \sum_{i \in G_g} \sum_{t=1}^{T} \Delta L_{it}$$

$$= \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \sqrt{\frac{n}{G}} T \Delta \bar{L}_{g,n_g,T}$$

$$= \sqrt{nT} \Delta \bar{L}_{nT}.$$

Hence, the test statistic is closer to the S-statistics rather than the C-statistics.

To have a clear idea of the asymptotic properties of J_n^D, let us suppose now that $G = 2$ and the cluster sizes are asymptotically identical: $\tau_1 = \tau_2 = 1/2$. Suppose also that $\bar{\mu}_{1,n_1} \rightarrow \mu^0 \neq 0$, $\bar{\mu}_{2,n_2} \rightarrow -\mu^0$ as $n \rightarrow \infty$, therefore $\bar{\mu}_n \rightarrow 0$. In this case H_0^a holds and the asymptotic distribution of $\sqrt{G} \bar{D}$ is identical to that of $\sqrt{nT} \Delta \bar{L}_{nT}$ under WCD. Hence, even if the clustered EPA hypothesis fails, the test cannot detect deviations from it.

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 specific economic importance. When this is the case, one may be interested in the following hypothesis:

$$H_0^j : E(\Delta L_{it}) = \mu_i = 0, \text{ for all } i = 1, 2, \ldots, n.$$

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^{ca} with the number of clusters equal to the number of units in the panel, that is $G = n$. However, in this case Assumption 8 is violated as for each $g = 1, \ldots, G$, $n_g = 1$ and therefore $n_g/n \rightarrow 0$. 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 that is robust to arbitrary CD is

$$J_{nT} = T \Delta \bar{\mathbf{L}}_{nT}' \widehat{\boldsymbol{\Omega}}_{nT}^{-1} \Delta \bar{\mathbf{L}}_{nT} \xrightarrow{D} \chi_n^2,$$

$$\widehat{\boldsymbol{\Omega}}_{nT} = \frac{1}{T} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \boldsymbol{\iota}_i \boldsymbol{\iota}_j' \Delta \tilde{L}_{it} \Delta \tilde{L}_{js},$$

with $\Delta \bar{\mathbf{L}}_{nT} = (\Delta \bar{L}_{1T}, \Delta \bar{L}_{2T}, \ldots, \Delta \bar{L}_{nT})'$, and $\boldsymbol{\iota}_i$ being the ith column of $\mathbf{I}_n$. Different cases of CD can be covered by modifications to the variance estimator. These cases are discussed at length in a previous version of this paper (Akgun, Pirotte, Urga, & Yang, 2020), 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 $\underline{S}_{nT}^{(2)}$, and clustered test statistics $C_{nT}^{(1)}$, $C_{nT}^{(2)}$ and $\underline{C}_{nT}^{(2)}$ are applicable to a single cross-section, i.e. when $T = 1$ after minor modifications. As an example, let us take $S_{nT}^{(1)}$. The modified statistic will take the following form:

$$S_n^{(1)} = \frac{\Delta \bar{L}_n}{\hat{\sigma}_{1,n}/\sqrt{n}},$$

where $\hat{\sigma}_{1,n}^2 = n^{-1} \sum_{i=1}^n \Delta \tilde{L}_{it} \Delta \tilde{L}_{it}$, with $\Delta \tilde{L}_{it} = \Delta L_{it} - \Delta \bar{L}_{n,t}$, where $\Delta \bar{L}_{n,t} = n^{-1} \sum_{i=1}^n \Delta L_{it}$. The modification is on calculating the variance, where we calculate the deviations from the overall mean instead of the individual means. A CLT can be easily obtained for a single cross-section as a basis for these statistics. For instance, 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

A set of Monte Carlo experiments are conducted to investigate the finite sample properties of the test statistics given above. Two thousand 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 that is controlled using a stationary spatial AR(1) process, which satisfies Assumption 3. 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. Define first

$$\zeta_{l,it} = \rho \sum_{j=1}^n w_{ij} \zeta_{l,jt} + u_{l,it}, \text{ with, } u_{l,it} \sim iid.N(0,1), \ l = 1, 2,$$
(28)

where w_{ij} is the element of the spatial matrix $\mathbf{W}_n$ in row i and column j with $w_{ii} = 0$ for all $i = 1, 2, \ldots, n$. Then the forecast error series $e_{l,it}$, $l = 1, 2$ are generated as the ith element of the n-vector

$$\mathbf{e}_{l,n,t} = \frac{1}{\sqrt{\bar{s}_2}} \mathbf{S}_n \mathbf{u}_{l,n,t},$$
(29)

where $\mathbf{u}_{l,n,t} = (\mathbf{u}_{l,1t}, \mathbf{u}_{l,2t}, \ldots, \mathbf{u}_{l,nt})'$, $\mathbf{S}_n = (\mathbf{I}_n - \rho \mathbf{W}_n)^{-1}$ and $\bar{s}_2 = n^{-1} \text{tr}(\mathbf{S}_n \mathbf{S}_n')$. To explore the size of various tests, we use these two forecast error series $e_{l,it}$, $l = 1, 2$. We set $\rho = 0.5$.[1] 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

$$\Delta L_{it} = \xi(\mu_i + \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + \varepsilon_{it}).$$
(30)

To investigate the size we set $\mu_i = 0$ for each $i = 1, 2, \ldots, n$ and generate factor loadings as $\lambda_{1i}, \lambda_{2i} \sim iid.N(1, 0.2)$. The common factors are formed by $f_{1t}, f_{2t} \sim iid.N(0, 1)$, The error series ε_{it} are generated in the same spirit as in (29), precisely we set $\varepsilon_{it} = e_{1,it}$. We finally set $\xi = \sqrt{1/3.4}$ to control for the variance of the loss differential series.

We also undertake a robustness check to study the small sample properties of the tests $S_n^{(3)}$ and $\tilde{S}_n^{(3)}$ by considering a heavy-tailed error distribution. We generate errors from a $t(6)$ distribution for half of the panel and a standard normal for the other half. Specifically, in (28) we set, for each $l = 1, 2$, $u_{l,it} \sim iid.t(6)$ if $i = 1, \ldots, N/2$ and $u_{l,it} \sim iid.N(0, 1)$ otherwise.

We explore the power of various tests under two different alternative hypotheses. For this purpose, in the case of DGP1, we generate a third series $e_{3i,t}$ as two different re-parametrizations of the series $e_{2,it}$. The first corresponds to the homogeneous alternative, and the second to the heterogeneous alternative. We generate $e_{3i,t} = \sqrt{1.2}e_{2,it}$ and report the results from testing the equality of forecast accuracy of $e_{1,it}$ and $e_{3i,t}$. In the heterogeneous scenario, we generate $e_{3i,t} = \sqrt{\theta_i}e_{2,it}$ 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 alternatives.

As the error series and the common factors are serially uncorrelated for each unit, it is implicitly assumed that we are dealing with one-step ahead forecasts. Hence, we set 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 row-normalized rook-type weight matrix.[2] The Euclidean distance gives the distance between two units. In the computation of the spatial kernel $k_S(\cdot)$, we use 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 Kelejian and Prucha (2007), we set the spatial kernel bandwidth to $d_n = \lceil n^{1/4} \rceil + 1$ where $\lceil \cdot \rceil$ stands for the smallest integer bigger than its argument. Similarly, the overall and clustered EPA test statistics using partial sample variance estimators are calculated by setting $\underline{n} = \lceil n^{1/2} \rceil + 1$ and $\underline{n}_1 = \underline{n}_2 = \lceil n^{1/2}/2 \rceil + 1$, respectively.

We consider three possibilities for selecting the number of factors for the tests using common factors. First, we calculate them assuming $m = 2$, meaning that for DGP2, the number of common factors is correctly specified. Second, we set $m = 1$ for the case of under-specification.

---

[1] We also tried $\rho = 0$ and $\rho = 0.9$. The findings are similar to those obtained with $\rho = 0.5$, and the results omitted to save space.

[2] The units are assumed to lie on a $p_1 \times p_2$ rectangular grid such that the first p_1 units are located in the first column of the grid, the second p_1 units are located in the second column, and so on. We, therefore, have $n = p_1 p_2$. For each $n \in \{10, 20, 30, 50, 100\}$, we choose p_1 as 2, 4, 6, 10, and 50, respectively.

Table 2
Small sample properties of the non-robust tests $S_{n,T}^{(1)}$ and $C_{n,T}^{(1)}$.

Overall EPA Tests: $S_{n,T}^{(1)}$						Clustered EPA Tests: $C_{n,T}^{(1)}$					
$n\|T$	10	20	30	50	100	$n\|T$	10	20	30	50	100
DGP1: Size											
10	13.6	13.1	12.1	12.7	13.4	10	15.4	16.0	14.5	14.8	14.8
20	13.1	12.5	12.2	10.5	11.5	20	14.8	15.0	14.1	12.5	13.2
30	11.4	11.7	11.9	11.4	10.5	30	14.1	12.5	12.6	12.0	11.1
50	14.3	11.6	10.4	11.1	10.1	50	16.0	12.9	11.4	12.1	12.2
100	13.5	11.1	12.8	12.4	11.4	100	12.9	10.6	11.6	10.9	10.2
DGP2: Size											
10	48.0	47.3	44.9	42.9	48.5	10	41.9	40.6	38.4	34.8	42.1
20	62.2	62.2	60.4	58.4	56.8	20	56.3	55.7	55.4	52.7	51.3
30	67.7	64.3	65.1	65.6	65.4	30	63.2	59.9	60.7	61.2	60.0
50	72.2	72.4	72.0	74.2	73.0	50	69.3	68.9	67.6	70.0	68.6
100	79.9	80.8	81.9	79.2	80.0	100	76.4	75.9	77.4	75.6	76.5
DGP1: Power											
10	24.2	33.1	38.2	54.8	76.3	10	23.6	29.8	34.2	49.3	71.7
20	32.0	49.2	61.4	78.3	95.9	20	28.7	44.8	57.3	73.2	93.5
30	41.0	61.1	74.7	89.5	99.4	30	38.2	54.0	67.7	85.8	98.7
50	55.5	78.3	91.1	98.5	100.0	50	51.7	73.1	87.3	97.6	100.0
100	79.9	96.4	99.3	100.0	100.0	100	73.0	94.3	98.7	100.0	100.0
DGP2: Power											
10	95.4	99.6	100.0	100.0	100.0	10	95.6	99.6	100.0	100.0	100.0
20	98.2	99.8	100.0	100.0	100.0	20	98.5	99.9	100.0	100.0	100.0
30	98.8	100.0	100.0	100.0	100.0	30	99.1	100.0	100.0	100.0	100.0
50	99.1	100.0	100.0	100.0	100.0	50	99.6	100.0	100.0	100.0	100.0
100	99.5	100.0	100.0	100.0	100.0	100	100.0	100.0	100.0	100.0	100.0

Note: Overall EPA Tests are introduced in Section 3.1 and Clustered EPA Tests are in Section 3.2. The nominal size is 5%. The power is calculated under the homogeneous alternative hypothesis.

Third, we select the number of common factors by minimizing the IC_{p1} information criterion[3] of Bai and Ng (2002) given by

$$IC_{p1} = V(m) + m \left(\frac{n+T}{nT} \right) \ln \left(\frac{nT}{n+T} \right) \qquad (31)$$

where $V(m)$ is defined by (14). The case of over-specification of the number of common factors is not separately considered because IC_{p1} almost always overestimates the number of common factors in small samples (see Table 9 below).

Before the discussion of the size and power 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 2. As is expected, all tests are incorrectly sized.

4.2. Size properties

The results on the size of CD-robust overall EPA tests with DGP1 are in Table 3. 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 $T = 10$

---

[3] The authors consider several information criteria. In their simulations IC_{p1} appears to be the best performing criterion under WCD.

with $n = 100$ it is 8.4%. In the largest sample, its size is 6.3% which is close to the nominal value of 5%. When the distance between the panel units is incorrectly specified, the test size improves with either dimension.

However, as expected, the size distortions are slightly larger in this case. In the 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%). In the case of large T, however, it performs better than the kernel robust test with incorrect distance. For instance, when $(T, n) = (100, 100)$ its size equals 7.7%.

The test $S_{nT}^{(3)}$ 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 $S_{nT}^{(2)}$. When T is greater than 50, it is correctly sized and even when $T = 30$ and $n = 100$ its size is 6.9% which makes it the preferred test over any version of $S_{nT}^{(2)}$.

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

Table 3
Size - DGP1: No common factors, spatial dependence.

Overall EPA Tests						Clustered EPA Tests					
$n\|T$	10	20	30	50	100	$n\|T$	10	20	30	50	100
$S_{n,T}^{(2)}$						$C_{n,T}^{(2)}$					
10	9.9	9.1	8.2	9.0	8.7	10	11.0	10.7	9.5	9.5	9.7
20	8.7	8.0	7.9	6.9	7.2	20	10.0	9.0	8.8	7.7	7.8
30	7.8	6.9	8.4	6.4	6.9	30	9.3	7.5	7.9	6.7	6.3
50	9.7	6.9	6.4	7.3	6.5	50	11.5	7.6	6.7	7.1	7.3
100	8.4	6.9	7.4	6.8	6.3	100	8.6	6.3	7.7	7.1	6.1
$S_{n,T}^{(2)}[ms]$						$C_{n,T}^{(2)}[ms]$					
10	12.5	11.1	10.3	10.8	11.0	10	13.3	13.0	11.7	11.7	11.6
20	11.1	10.3	10.2	8.8	9.3	20	12.3	11.1	11.2	10.2	10.2
30	9.7	9.2	10.5	8.6	8.4	30	11.1	9.3	10.1	9.2	8.3
50	11.8	9.2	7.9	9.6	8.4	50	13.0	9.8	8.8	8.8	9.3
100	9.8	8.4	9.7	9.0	8.3	100	8.3	6.1	8.0	7.2	6.1
$\underline{S}_{n,T}^{(2)}$						$\underline{C}_{n,T}^{(2)}$					
10	12.9	9.1	7.7	7.0	6.3	10	23.5	17.8	15.2	13.9	12.5
20	12.2	10.0	9.1	6.8	6.6	20	18.2	12.1	9.8	8.1	7.3
30	11.4	8.7	9.2	6.6	6.8	30	23.2	15.5	14.3	11.6	9.2
50	13.1	9.6	7.7	7.5	6.8	50	24.9	15.9	12.1	10.7	10.5
100	13.4	9.9	10.0	9.0	7.7	100	19.4	9.8	9.3	7.3	5.5
$S_{n,T}^{(3)}$						$C_{n,T}^{(3)}$					
10	8.3	7.2	5.9	5.8	5.9	10	13.8	9.9	7.4	6.3	6.9
20	9.7	7.4	6.9	6.0	5.7	20	14.2	9.2	7.9	6.6	5.8
30	8.7	6.9	7.5	5.3	5.2	30	13.9	8.4	7.7	6.3	5.0
50	10.9	7.4	5.3	5.9	5.7	50	16.9	9.3	6.7	6.9	5.8
100	10.0	7.2	7.0	6.5	5.4	100	15.7	8.6	8.6	7.9	5.7
$\underline{S}_{n,T}^{(3)}[m=2]$						$\underline{C}_{n,T}^{(3)}[m=2]$					
10	8.9	8.2	6.2	5.4	4.5	10	14.9	12.4	8.6	6.8	6.0
20	10.7	9.0	8.2	6.9	6.8	20	16.5	12.1	10.5	9.0	7.6
30	9.4	8.3	8.6	7.4	6.9	30	14.1	11.1	10.7	9.0	7.7
50	13.0	8.7	7.7	8.1	7.7	50	16.9	12.0	9.8	9.8	9.7
100	11.7	9.3	10.6	10.5	9.0	100	13.9	9.6	11.0	9.8	8.8
$\underline{S}_{n,T}^{(3)}[m=1]$						$\underline{C}_{n,T}^{(3)}[m=1]$					
10	10.3	9.4	8.1	7.2	7.2	10	16.6	14.3	11.0	10.2	10.6
20	11.9	10.1	9.6	8.6	8.1	20	16.9	13.5	12.1	10.6	9.7
30	10.3	9.5	9.5	8.9	8.4	30	13.9	12.3	11.6	9.9	8.8
50	13.1	9.7	9.1	9.6	9.0	50	17.1	12.7	10.6	10.8	10.6
100	11.9	10.0	11.2	11.6	10.1	100	13.6	10.1	10.9	10.4	9.5
$\underline{S}_{n,T}^{(3)}[ic]$						$\underline{C}_{n,T}^{(3)}[ic]$					
10	8.2	6.8	5.8	5.6	7.9	10	14.4	10.1	7.4	6.4	8.4
20	9.6	8.2	10.0	9.8	11.4	20	15.0	11.3	12.2	11.9	13.2
30	8.8	9.6	11.5	11.2	10.5	30	14.8	11.3	12.0	11.8	11.1
50	11.2	11.3	10.3	11.1	10.1	50	17.3	12.9	11.5	12.1	12.2
100	11.2	11.1	12.8	12.4	11.4	100	15.1	10.6	11.6	10.9	10.2

Note: See the note of Table 2. [*ms*] indicates that the test uses a misspecified distance metric. For $\underline{S}_{n,T}^{(3)}$, the number of common factors is shown in brackets. [*ic*] means that the number of common factors is chosen by IC.

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 9. It is seen that when either of the dimension of the panel grows, the performance of the IC increases. However, this improvement is prolonged when one of the dimensions is small. In DGP1, only when one of the dimensions is greater than 50 is the performance at 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 that the test converges to the non-robust test in these cases. Hence, the empirical size and power of the test equals the size and power of the non-robust test given in Table 2. 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 based on IC. However, it should not be understood that the performance, in this case, best describes our testing procedure. In practice, if a suitable test rejects the no CD hypothesis, 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, $S_{nT}^{(2)}$ or $S_{nT}^{(3)}$. As in the case of zero common factors $\underline{S}_{nT}^{(3)}$ is identical to $S_{nT}^{(1)}$, 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 clustered EPA tests are reported in the right block of Table 3. 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 dimension 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 in the number of observations. The performance of the test based on the partial sample estimator, namely $\underline{C}_{nT}^{(2)}$, is unsatisfactory in small samples, especially for small n. For $T = 10$, even for the largest n we have, the empirical size of the test equals 19.4%. However, its performance improves with T, reaching 5.5% in the largest sample.

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 9.9%,

Table 4
Size - DGP2: Common factors, spatial dependence.

Overall EPA Tests							Clustered EPA Tests					
	$n\|T$	10	20	30	50	100		$n\|T$	10	20	30	50	100
$S_{n,T}^{(2)}$	10	30.7	30.5	26.6	24.7	29.6	$C_{n,T}^{(2)}$	10	24.1	22.2	19.6	17.0	20.7
	20	30.0	29.4	28.1	25.9	25.0		20	21.5	20.6	19.5	17.3	17.1
	30	37.3	34.7	35.6	34.8	32.6		30	29.3	25.7	27.4	25.6	24.1
	50	47.4	43.6	42.9	44.4	42.7		50	39.0	35.1	33.9	35.2	32.7
	100	59.3	59.1	58.0	56.3	58.0		100	52.7	51.9	52.2	49.6	49.9
$S_{n,T}^{(2)}[ms]$	10	37.7	36.1	33.5	31.3	38.1	$C_{n,T}^{(2)}[ms]$	10	30.5	29.9	25.9	24.7	29.1
	20	45.5	44.9	43.5	41.4	40.3		20	36.9	35.5	35.2	32.4	31.6
	30	52.8	49.5	51.2	50.8	50.0		30	46.0	42.2	43.8	42.0	40.3
	50	60.9	59.3	58.9	60.2	59.8		50	55.7	51.9	51.9	53.3	52.0
	100	68.2	67.3	67.6	66.4	67.4		100	61.2	61.0	60.0	58.5	59.8
$\underline{S}_{n,T}^{(2)}$	10	24.3	23.0	20.3	18.9	22.4	$\underline{C}_{n,T}^{(2)}$	10	35.3	29.2	26.4	23.1	25.3
	20	34.8	33.7	32.9	30.5	29.0		20	33.7	28.8	26.5	24.2	23.0
	30	40.4	37.6	38.4	37.4	35.9		30	46.9	39.9	40.9	39.0	35.5
	50	45.7	41.8	41.3	42.0	40.5		50	50.8	44.3	42.0	43.8	39.3
	100	54.0	54.3	54.2	51.6	52.3		100	56.8	53.1	53.6	50.2	49.3
$S_{n,T}^{(3)}$	10	10.2	8.5	6.0	6.3	5.7	$C_{n,T}^{(3)}$	10	16.9	9.8	7.5	6.7	7.1
	20	9.4	7.4	6.7	5.2	5.9		20	15.4	10.1	8.0	6.5	6.2
	30	10.3	6.5	7.7	5.8	5.4		30	15.7	8.8	8.3	6.7	6.0
	50	8.7	6.9	6.5	5.5	5.0		50	15.2	9.3	7.3	6.8	4.8
	100	8.5	7.1	5.3	5.7	4.7		100	14.9	9.0	7.5	7.2	5.2
$\underline{S}_{n,T}^{(3)}[m=2]$	10	9.9	8.0	5.6	6.0	5.3	$\underline{C}_{n,T}^{(3)}[m=2]$	10	20.0	12.9	11.0	9.1	9.8
	20	9.3	7.3	6.3	5.0	5.4		20	21.1	15.8	15.4	13.8	12.8
	30	10.1	6.5	7.7	5.5	5.4		30	22.0	16.8	16.6	16.0	14.6
	50	8.6	6.8	6.4	5.4	4.9		50	23.8	18.6	15.8	15.9	14.3
	100	8.5	7.1	5.2	5.7	4.6		100	13.3	8.7	7.3	7.4	5.4
$\underline{S}_{n,T}^{(3)}[m=1]$	10	9.8	7.7	5.6	5.8	5.2	$\underline{C}_{n,T}^{(3)}[m=1]$	10	25.3	20.3	18.7	16.8	15.6
	20	9.3	7.2	6.2	4.9	5.3		20	24.8	19.2	18.5	16.2	15.8
	30	10.1	6.4	7.6	5.5	5.3		30	24.2	18.6	18.2	17.2	15.1
	50	8.6	6.7	6.4	5.4	4.9		50	25.5	19.4	16.6	16.3	14.5
	100	8.6	7.1	5.2	5.7	4.6		100	12.4	8.8	7.2	7.7	5.3
$\underline{S}_{n,T}^{(3)}[ic]$	10	10.1	8.3	6.0	6.2	5.5	$\underline{C}_{n,T}^{(3)}[ic]$	10	16.7	9.5	7.3	6.1	6.3
	20	9.3	7.3	6.4	5.1	5.5		20	15.9	10.3	9.1	8.0	7.7
	30	10.2	6.5	7.7	5.5	5.4		30	16.4	11.4	13.8	13.7	13.5
	50	8.6	6.8	6.4	5.4	4.9		50	16.7	17.5	15.2	15.9	14.3
	100	8.5	7.1	5.2	5.7	4.6		100	14.4	8.8	7.3	7.4	5.4

Note: See the note of Table 3.

which is better than that of the kernel robust test (10.7%). When $n = 100$ and $T = 20$ the performance of the latter improves dramatically and reaches 6.3% whereas that of $C_{nT}^{(3)}$ is 8.6%.

Finally, we focus on the size of the test using estimated common factors, $\underline{C}_{nT}^{(3)}$. 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 $m = 2$ performs similarly to the others. In fact for the smallest n and largest T that we consider it has an empirical size of 6% which is better than that of $C_{nT}^{(2)}$ (9.7%) and $C_{nT}^{(3)}$ (6.9%). 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 non-robust test, as expected.

The size results for DGP2 are reported in Table 4. As expected, for this DGP, the overall tests $S_{nT}^{(2)}$ and $\underline{S}_{nT}^{(2)}$ 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 $\underline{S}_{nT}^{(3)}$. This test performs better than $S_{nT}^{(3)}$ for all sample sizes considered. An important finding is that the test performs very well even

when the number of common factors is underspecified. We also see that the three test versions are equivalent in large samples. These three tests have equal empirical sizes for $T > 10$ and $n > 30$.

The findings concerning the clustered EPA tests robust to SCD are similar to those in the case of DGP1, with a few points worth mentioning. The test $C_{nT}^{(3)}$ 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 $S_{nT}^{(3)}$. The tests based on estimated common factors are found to be oversized in small samples when the number of common factors is chosen by the researcher. The IC-based version shows fewer size distortions. For small T it outperforms $C_{nT}^{(3)}$ overall. Finally, the IC-based test and the test with $m = 2$ have identical sizes in the largest samples. 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 9.

We conclude this section by reporting the following two simulation exercises. In the first, we evaluate the performance of the test statistics $S_{nT}^{(3)}$ and $\tilde{S}_{nT}^{(3)}$. In the second exercise, we study the properties of the J_n^D test statistic of QTZ. We note that the modified test $\tilde{S}_{nT}^{(3)}$ is

Table 5
Small sample properties of $S_{n,T}^{(3)}$ and $\tilde{S}_{n,T}^{(3)}$ under deviations from normality.

DGP1						DGP2					
$n|T$	10	20	30	50	100	$n|T$	10	20	30	50	100
Size											
$S_{n,T}^{(3)}$						$S_{n,T}^{(3)}$					
10	8.6	6.3	6.4	5.1	5.1	10	9.3	7.6	6.8	5.8	5.7
20	8.4	7.4	5.0	6.0	5.3	20	10.2	7.3	6.9	5.9	5.3
30	10.5	6.6	6.3	6.2	5.7	30	9.6	7.7	6.7	5.2	5.6
50	8.3	7.6	6.1	6.0	5.5	50	10.1	7.2	7.4	5.9	6.6
100	8.9	7.0	5.8	5.3	5.0	100	8.1	6.2	5.6	5.1	5.6
$\tilde{S}_{n,T}^{(3)}$						$\tilde{S}_{n,T}^{(3)}$					
10	4.1	3.9	5.0	4.4	4.8	10	4.6	5.2	5.3	5.3	5.5
20	3.6	4.6	3.7	5.2	4.8	20	4.9	5.4	5.1	4.8	4.9
30	4.8	4.7	5.1	5.4	5.1	30	5.2	5.5	5.6	4.4	5.3
50	4.4	5.2	4.6	5.2	5.4	50	5.3	5.0	6.1	5.4	5.9
100	4.4	4.5	4.5	4.6	4.7	100	3.9	4.0	4.6	4.3	5.1
Power											
$S_{n,T}^{(3)}$						$S_{n,T}^{(3)}$					
10	13.4	14.9	20.8	27.8	45.3	10	72.8	93.5	98.2	100.0	100.0
20	21.3	25.9	33.6	49.3	73.5	20	74.4	95.6	99.2	100.0	100.0
30	25.2	34.6	46.7	63.8	87.2	30	77.6	95.3	98.9	100.0	100.0
50	32.0	49.4	63.1	83.6	98.0	50	77.9	95.4	99.1	100.0	100.0
100	52.5	75.4	88.0	98.0	100.0	100	78.9	96.8	99.3	100.0	100.0
$\tilde{S}_{n,T}^{(3)}$						$\tilde{S}_{n,T}^{(3)}$					
10	7.0	11.2	17.0	25.4	43.9	10	60.5	90.6	97.4	100.0	100.0
20	12.1	20.8	29.5	46.2	72.2	20	62.2	93.3	98.7	100.0	100.0
30	14.9	28.2	42.2	62.0	86.6	30	65.8	93.2	98.8	100.0	100.0
50	20.7	41.8	59.5	82.1	97.6	50	65.9	93.8	98.9	100.0	100.0
100	38.3	68.9	85.4	97.6	100.0	100	66.4	95.3	99.2	100.0	100.0

Note: The tests are introduced in Section 3.1. The nominal size is 5%. The power is calculated under the homogeneous alternative hypothesis.

valid only for one-step-ahead forecasts. Furthermore, it requires normality under SCD.

Table 5 reports the finite sample properties of $S_{nT}^{(3)}$ and $\tilde{S}_{nT}^{(3)}$ under the failure of normality assumption. As can be seen, the test statistic $\tilde{S}_{nT}^{(3)}$ is perfectly sized irrespective of the sample size, even when normality fails, under both DGPs. Turning to the performance of J_n^D, we note that the test is valid only under WCD (DGP1). This is confirmed by the results reported in Table 6, where the test's size is very close to the nominal size irrespective of the sample size for DGP1, while it is grossly oversized under DGP2.

To summarize, in the case of both DGPs, the 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 test $S_{nT}^{(3)}$ has very good properties in both DGPs. In addition, $\tilde{S}_{nT}^{(3)}$ can be used for one step ahead forecasts with size practically equal to the nominal size, irrespective of the type of CD in the loss differentials. 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 $C_{nT}^{(3)}$ is preferred over others; however, $\underline{C}_{nT}^{(3)}$ also has 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 7. In the previous subsection, we have seen that the size of the overall EPA tests $S_{nT}^{(2)}$, $\underline{S}_{nT}^{(2)}$ and $S_{nT}^{(3)}$ approach to the nominal level for this DGP. Here, it is seen that the power of the tests $\underline{S}_{nT}^{(2)}$ and $S_{nT}^{(2)}$ converge to 100%, for the

Table 6
Size properties of the clustered test J_n^D of QTZ.

$n|T$	DGP1					DGP2				
	10	20	30	50	100	10	20	30	50	100
10	5.5	5.0	5.4	5.1	5.6	15.4	15.8	15.7	15.1	15.7
20	5.6	5.1	5.8	5.0	5.2	19.5	20.3	20.7	20.3	19.9
30	5.5	5.6	5.3	5.5	5.5	23.0	22.7	23.3	23.9	22.9
50	5.9	5.3	6.0	5.6	4.4	28.6	28.6	29.3	30.5	27.8
100	5.5	5.2	6.7	6.1	6.3	47.0	48.5	46.8	46.2	47.9

latter, the distance metric is unimportant. Hence the test is consistent in all cases. For moderate to large T, the test $S_{nT}^{(3)}$ is correctly sized. We observe that its power is slightly lower compared to that of $S_{nT}^{(2)}$ in these sample sizes. For instance, when $T = 100$ and $n = 10$ the power of $S_{nT}^{(2)}$ equals 70.8% whereas those of $S_{nT}^{(3)}$ and $\underline{S}_{nT}^{(3)}$ are 63.5% and 57.2%, respectively. Hence, even though they wrongly assume that there are common factors in the DGP, the power of the factor-robust tests $\underline{S}_{nT}^{(3)}$ are close to that of $S_{nT}^{(3)}$. To conclude, we see that the factor robust tests have lower power under WCD, as expected.

The previous results of the clustered 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 19.1%, and 17.3%, respectively, and their power reaches 100% in the largest sample.

Table 8 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 oversized, even asymptotically. Hence, we focus on the tests $S_{nT}^{(3)}$ and $\underline{S}_{nT}^{(3)}$.

Table 7

Power under homogeneous alternative – DGP 1: No common factors, spatial dependence.

Overall EPA Tests							Clustered EPA Tests					
	$n\|T$	10	20	30	50	100		$n\|T$	10	20	30	50	100
$S_{n,T}^{(2)}$	10	19.1	26.6	30.8	47.6	70.8	$C_{n,T}^{(2)}$	10	17.3	22.3	26.8	40.3	64.1
	20	24.6	40.1	53.1	71.2	93.2		20	21.7	34.3	45.7	63.7	88.4
	30	33.6	51.9	67.2	85.2	98.9		30	29.5	44.6	57.8	78.4	97.3
	50	48.6	71.6	87.8	97.5	100.0		50	42.8	63.5	81.1	94.8	100.0
	100	71.2	94.1	98.4	100.0	100.0		100	63.4	89.8	96.8	100.0	100.0
$S_{n,T}^{(2)}[ms]$	10	21.8	29.9	34.9	51.6	73.8	$C_{n,T}^{(2)}[ms]$	10	20.0	26.1	30.6	45.1	68.3
	20	28.7	45.3	57.3	75.1	95.0		20	24.9	39.6	51.9	68.5	91.1
	30	37.6	56.3	71.2	87.9	99.3		30	33.9	49.6	62.8	82.4	98.3
	50	52.4	75.2	89.8	98.1	100.0		50	47.0	68.0	84.6	96.2	100.0
	100	74.3	95.2	98.7	100.0	100.0		100	67.4	91.8	97.8	100.0	100.0
$\underline{S}_{n,T}^{(2)}$	10	20.3	23.3	27.0	41.0	64.2	$\underline{C}_{n,T}^{(2)}$	10	31.5	30.9	33.0	45.5	67.7
	20	27.0	38.7	52.0	68.8	92.0		20	30.9	36.8	45.5	61.2	87.0
	30	36.2	52.9	65.6	83.4	98.6		30	43.7	52.3	65.1	82.6	97.8
	50	50.6	72.5	85.8	97.1	100.0		50	56.0	70.9	83.9	95.4	100.0
	100	73.3	94.5	98.5	100.0	100.0		100	66.6	86.8	94.6	99.6	100.0
$S_{n,T}^{(3)}$	10	16.5	21.7	25.4	39.1	63.5	$C_{n,T}^{(3)}$	10	18.5	19.9	22.3	32.0	53.5
	20	23.8	37.3	48.9	66.2	90.4		20	26.0	32.5	42.3	57.8	84.9
	30	31.8	48.8	63.0	82.2	98.3		30	33.1	42.9	54.3	73.2	96.3
	50	47.0	67.7	84.3	96.3	100.0		50	45.5	60.7	78.4	92.8	100.0
	100	68.5	92.2	97.8	99.9	100.0		100	66.2	88.4	95.8	99.8	100.0
$\underline{S}_{n,T}^{(3)}[m=2]$	10	17.4	20.8	25.1	36.0	57.2	$\underline{C}_{n,T}^{(3)}[m=2]$	10	21.0	21.8	23.5	31.0	47.4
	20	25.5	38.6	50.4	67.1	90.1		20	28.0	37.8	46.3	61.4	84.9
	30	34.3	52.3	66.9	84.1	98.1		30	34.4	47.8	58.8	79.0	96.9
	50	50.3	72.1	87.3	97.3	100.0		50	48.7	67.6	81.9	95.5	100.0
	100	73.0	94.4	98.7	100.0	100.0		100	68.6	92.0	97.5	99.9	100.0
$\underline{S}_{n,T}^{(3)}[m=1]$	10	19.4	24.0	28.3	41.5	61.9	$\underline{C}_{n,T}^{(3)}[m=1]$	10	23.4	25.6	27.3	38.7	57.2
	20	28.3	42.8	53.5	71.8	92.1		20	28.3	40.2	50.8	66.8	88.3
	30	36.3	54.7	70.1	86.3	98.8		30	36.1	50.6	62.7	81.3	97.5
	50	52.4	74.4	89.1	97.7	100.0		50	50.4	69.9	84.0	96.4	100.0
	100	75.5	95.3	98.9	100.0	100.0		100	70.4	92.8	97.9	100.0	100.0
$\underline{S}_{n,T}^{(3)}[ic]$	10	16.3	21.0	23.8	36.8	64.9	$\underline{C}_{n,T}^{(3)}[ic]$	10	19.1	19.7	20.9	30.6	56.4
	20	23.8	38.7	56.0	76.3	95.8		20	26.6	35.3	51.6	71.1	93.3
	30	31.6	56.0	73.1	89.0	99.4		30	33.5	49.9	65.7	85.4	98.7
	50	48.2	77.0	90.9	98.5	100.0		50	47.0	71.8	87.1	97.6	100.0
	100	71.8	96.4	99.3	100.0	100.0		100	68.0	94.2	98.6	100.0	100.0

Note: See the note of Table 3.

It is seen that the power of both tests is 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 $\underline{C}_{nT}^{(3)}$ when the number of common factors is chosen by the IC.

The power of the tests $S_{nT}^{(3)}$ and $\tilde{S}_{nT}^{(3)}$ under non-normality are reported in the second panel of Table 5. As is stated in the previous subsection, $\tilde{S}_{nT}^{(3)}$ improves greatly over $S_{nT}^{(3)}$ in terms of size. We now see $\tilde{S}_{nT}^{(3)}$ has less power for any sample size compared to $S_{nT}^{(3)}$. It has to be noted, however, that $\tilde{S}_{nT}^{(3)}$ still performs very well, and its power equals 100% as the sample size increases.

We do not report the power results under the heterogeneous alternative hypothesis to save space. The main findings are summarized in Fig. 1 where the power of $S_{nT}^{(3)}$, $C_{nT}^{(3)}$ and J_n^D are shown under DGP1 so that the assumptions of QTZ are satisfied. It can be seen that under the homogeneous alternative, the power of our tests $S_{nT}^{(3)}$ and $C_{nT}^{(3)}$ rapidly approach 100%. Whereas the power of J_n^D is very low compared to our proposed tests, although it increases with sample size. Under the heterogeneous alternative on the other hand, the power of the overall test $S_{nT}^{(3)}$ approaches the nominal size. This is because the overall EPA

hypothesis holds under this alternative hypothesis. On the other hand, the expected value of the loss differentials is different from zero for clusters of panel units; hence the clustered EPA test $C_{nT}^{(3)}$ has power under this scenario. Under this alternative hypothesis, we see that the power of J_n^D approaches zero. Hence, it is inconsistent contrary to our proposed test $C_{nT}^{(3)}$.

5. Guidelines for empirical applications

The application of the panel EPA tests requires 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 must have some information on the type of the CD in loss differentials, as WCD and SCD may require 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.
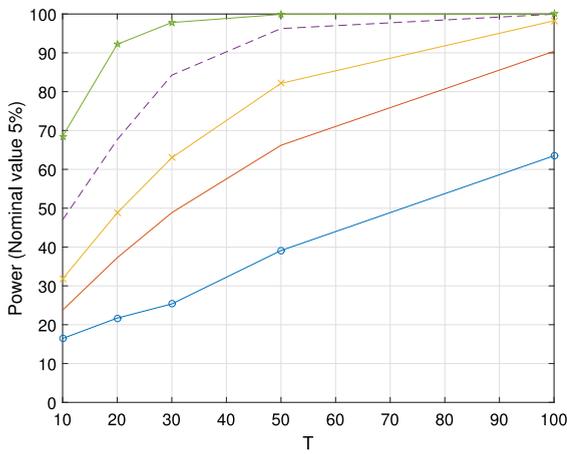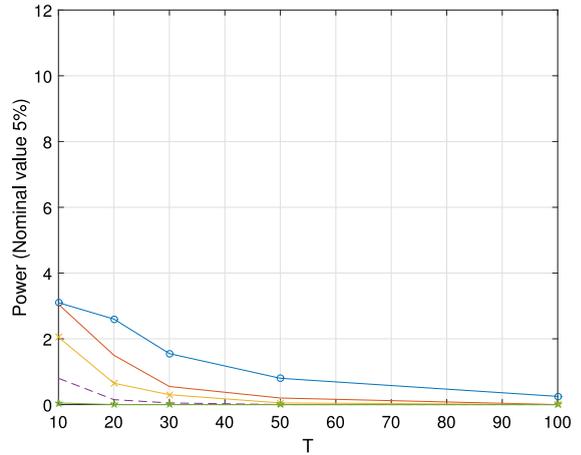
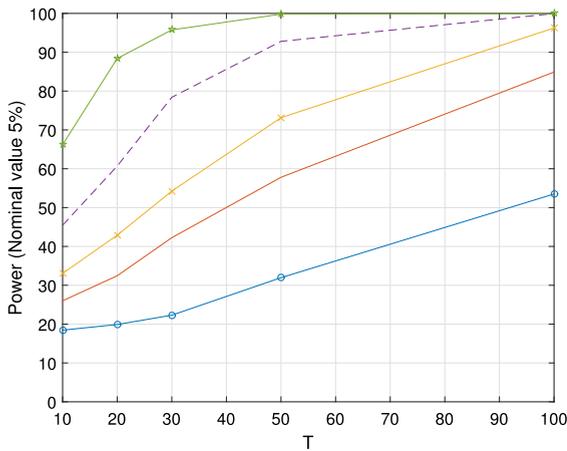(a) $S_{n,T}^{(3)}$, Homogeneous Alternative
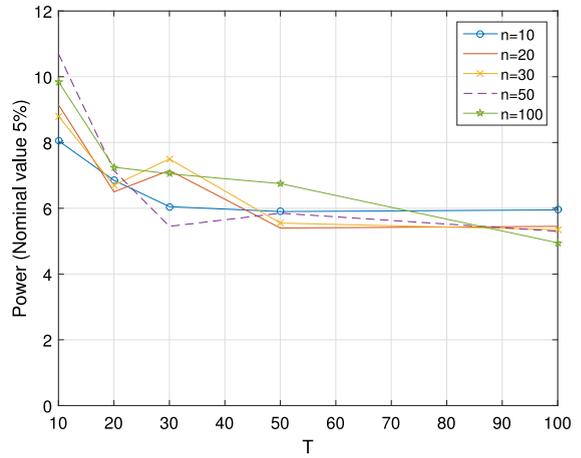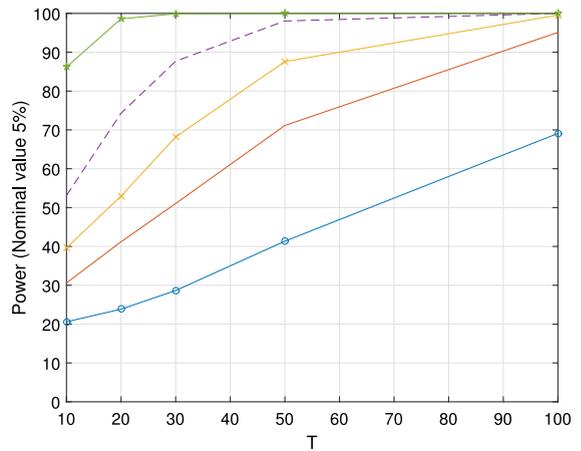
(b) $S_{n,T}^{(3)}$, Heterogeneous Alternative
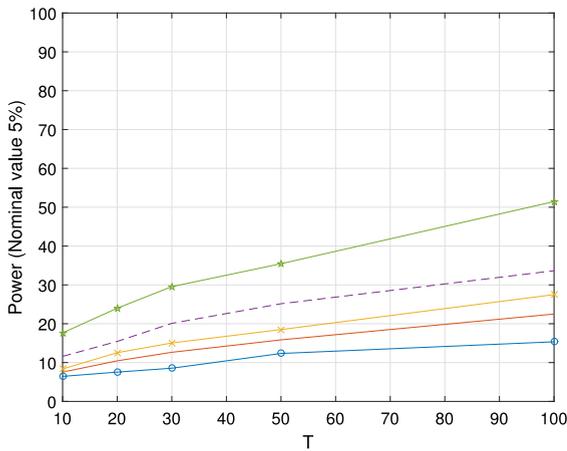
(c) J_n^D, Homogeneous Alternative

(d) J_n^D, Heterogeneous Alternative

(e) $C_{n,T}^{(3)}$, Homogeneous Alternative

(f) $C_{n,T}^{(3)}$, Heterogeneous Alternative

Fig. 1. Power of Selected Tests Under Different Alternative Hypotheses for DGP1 (5% Nominal Size)

Table 8
Power under homogeneous alternative – DGP 2: Common factors, spatial dependence.

Overall EPA Tests							Clustered EPA Tests					
	$n\|T$	10	20	30	50	100		$n\|T$	10	20	30	50	100
$S_{n,T}^{(2)}$	10	91.8	98.9	99.9	100.0	100.0	$C_{n,T}^{(2)}$	10	90.0	98.6	99.9	100.0	100.0
	20	92.8	99.2	100.0	100.0	100.0		20	91.1	99.2	100.0	100.0	100.0
	30	95.6	99.5	100.0	100.0	100.0		30	94.9	99.6	100.0	100.0	100.0
	50	97.1	99.9	100.0	100.0	100.0		50	97.1	100.0	100.0	100.0	100.0
	100	98.5	100.0	100.0	100.0	100.0		100	99.9	100.0	100.0	100.0	100.0
$S_{n,T}^{(2)}[ms]$	10	93.8	99.3	100.0	100.0	100.0	$C_{n,T}^{(2)}[ms]$	10	93.0	99.3	99.9	100.0	100.0
	20	96.3	99.6	100.0	100.0	100.0		20	95.9	99.7	100.0	100.0	100.0
	30	97.6	99.8	100.0	100.0	100.0		30	97.4	99.8	100.0	100.0	100.0
	50	98.3	100.0	100.0	100.0	100.0		50	98.9	100.0	100.0	100.0	100.0
	100	99.2	100.0	100.0	100.0	100.0		100	99.7	100.0	100.0	100.0	100.0
$\underline{S}_{n,T}^{(2)}$	10	88.8	98.3	99.9	100.0	100.0	$\underline{C}_{n,T}^{(2)}$	10	93.1	98.8	100.0	100.0	100.0
	20	93.6	99.3	100.0	100.0	100.0		20	95.6	99.6	100.0	100.0	100.0
	30	95.5	99.6	100.0	100.0	100.0		30	98.1	100.0	100.0	100.0	100.0
	50	96.3	99.9	100.0	100.0	100.0		50	99.3	100.0	100.0	100.0	100.0
	100	98.4	99.9	100.0	100.0	100.0		100	100.0	100.0	100.0	100.0	100.0
$S_{n,T}^{(3)}$	10	74.5	93.1	99.0	100.0	100.0	$C_{n,T}^{(3)}$	10	81.3	94.9	99.2	100.0	100.0
	20	74.9	94.8	99.4	100.0	100.0		20	85.7	98.2	99.6	100.0	100.0
	30	77.7	95.6	99.3	100.0	100.0		30	89.4	99.2	99.8	100.0	100.0
	50	78.1	95.4	99.5	100.0	100.0		50	94.3	99.8	100.0	100.0	100.0
	100	79.0	96.6	99.7	100.0	100.0		100	99.7	100.0	100.0	100.0	100.0
$\underline{S}_{n,T}^{(3)}[m=2]$	10	73.8	92.8	98.9	100.0	100.0	$\underline{C}_{n,T}^{(3)}[m=2]$	10	82.4	95.1	99.3	100.0	100.0
	20	74.6	94.6	99.3	100.0	100.0		20	88.5	98.4	99.7	100.0	100.0
	30	77.4	95.6	99.2	100.0	100.0		30	91.8	99.4	99.8	100.0	100.0
	50	77.9	95.4	99.5	100.0	100.0		50	96.3	99.9	100.0	100.0	100.0
	100	78.9	96.6	99.7	100.0	100.0		100	99.7	100.0	100.0	100.0	100.0
$\underline{S}_{n,T}^{(3)}[m=1]$	10	73.6	92.6	98.9	100.0	100.0	$\underline{C}_{n,T}^{(3)}[m=1]$	10	86.1	96.2	99.3	100.0	100.0
	20	74.2	94.5	99.3	100.0	100.0		20	90.1	98.9	99.7	100.0	100.0
	30	77.4	95.6	99.2	100.0	100.0		30	92.8	99.5	99.9	100.0	100.0
	50	77.8	95.4	99.5	100.0	100.0		50	97.0	99.9	100.0	100.0	100.0
	100	78.9	96.6	99.7	100.0	100.0		100	99.8	100.0	100.0	100.0	100.0
$\underline{S}_{n,T}^{(3)}[ic]$	10	74.3	93.0	99.0	100.0	100.0	$\underline{C}_{n,T}^{(3)}[ic]$	10	81.3	94.7	99.2	100.0	100.0
	20	74.9	94.7	99.4	100.0	100.0		20	86.0	98.0	99.7	100.0	100.0
	30	77.6	95.6	99.2	100.0	100.0		30	89.7	99.3	99.8	100.0	100.0
	50	78.1	95.4	99.5	100.0	100.0		50	94.5	99.9	100.0	100.0	100.0
	100	79.0	96.6	99.7	100.0	100.0		100	99.7	100.0	100.0	100.0	100.0

Note: See the note of Table 3.

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, Ullah, and Yamagata (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:

$$\Delta L_{it} = \pi_1 \Delta L_{i,t-1} + \pi_2 \Delta L_{i,t-2} + \cdots + \pi_p \Delta L_{i,t-p} + a_i + \zeta_{it}, \quad (32)$$

and the hypothesis $H_0^{ac} : \pi_1 = \pi_2 = \cdots = \pi_p$.

1. Run a fixed effects regression on (32) and test H_0^{ac} using a Wald test with a variance robust to panel-level heteroskedasticity.
2. Run a fixed effects regression on (32) 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^a use $S_{nT}^{(1)}$, and for the clustered EPA hypothesis H_0^{ca} use $C_{nT}^{(1)}$.
2. If IC_{p1} indicates $m = 0$, use $S_{nT}^{(2)}$, $\underline{S}_{nT}^{(2)}$ or $S_{nT}^{(3)}$ to test H_0^a, and $C_{nT}^{(2)}$, $\underline{C}_{nT}^{(2)}$ or $C_{nT}^{(3)}$ to test H_0^{ca}. If IC_{p1} indicates $m > 0$, use $S_{nT}^{(3)}$ or $\underline{S}_{nT}^{(3)}$ to test H_0^a, and $C_{nT}^{(3)}$ or $\underline{C}_{nT}^{(3)}$ to test H_0^{ca}.

In Step A, we suggest two tests of CD. The null hypothesis of the BP test is the joint absence of CD between all pairs in the panel. The statistic is distributed as χ_q^2 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 $n \rightarrow \infty$ 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. We suggest the information criterion IC_{p1} in (31) to see if the loss differentials contain common factors. As seen in our simulations (Table 9), this IC performs quite well to choose between WCD and SCD.

Table 9
Small sample properties of the information criterion IC_{p1}.

$n\|T$	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	2.25	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

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 (32) 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. A critical point in Step B is the determination of the lag length p. Han, Phillips, and Sul (2017) found that the BIC is inconsistent in panel autoregressions even in the absence of fixed effects. We suggest using the general-to-specific methodology that they propose. The method starts with p_{max} chosen by the researcher and 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 non-zero. 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^a or H_0^{ca} based on the outcome of the two previous steps. As documented in this methodology, pre-testing may affect our proposed tests' properties. This is an interesting issue worth exploring but goes beyond this paper's scope. In the next section, we follow this methodology.

6. Empirical applications

6.1. Comparing OECD and IMF economic growth forecasts

We compare the OECD and IMF GDP growth forecasts using the EPA methodology described above. Our application can be seen as complementary to the application of QTZ comparing the IMF forecasts to Consensus Economics forecasts. The data for the IMF forecasts come from their Historical WEO Forecasts Database. The database includes historical h-years ahead forecast values, $h = 1, 2, \ldots, 5$, for the GDP growth rate and covers up to 192 countries and starts from the early 1990s. We collected similar data from the past vintages of the Economic Outlook of the OECD. The Economic Outlook contains only 1-year

ahead forecasts. Both organizations publish their forecasts twice a year. Our application focuses on their summer forecasts made for the following year. The OECD publishes these forecasts yearly in June, whereas IMF forecasts are published in April. Publishing dates are close, hence the forecast errors are comparable. To compute the forecast errors, we use the GDP growth outturns by the IMF. We also tried using the GDP growth published by the OECD, and the results are unchanged (the Pearson correlation coefficient between the two outturn series equals 0.995; hence the difference is negligible). Eventually, we constructed a balanced panel data set of annual GDP growth forecast errors of 29 OECD countries from the two organizations between 1998 and 2016.

We use quadratic loss, which is defined as

$$\Delta L_{it}^{(q)} = e_{1,it}^2 - e_{2,it}^2,$$

where, as is throughout this application, the first organization is the OECD. This loss function is arguably the most frequently used one and is robust to measurement errors contrary to other loss functions such as absolute loss (Hoga & Dimitriadis, 2022).

We begin the analysis with 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.[4] The DM test results are given in Table 10 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. We see that roughly half of the statistics are negative. Second, most of these statistics are statistically insignificant, except for Belgium, Spain, Hungary, and Luxembourg. For Belgium, a country where the predictive ability of the IMF is superior, the EPA hypothesis can be rejected at 10%. For Spain's OECD predictions, Hungary's IMF predictions outperform them. For Luxembourg, we can reject the EPA hypothesis at the 10% level, and the predictive ability of the IMF is found to be superior for this country.

Finally, at the bottom right of Table 10, we report average loss differentials over clusters of countries and the period 1998–2016. The average quadratic loss differential over all 29 OECD countries in the sample is 0.009. This shows that, when we average over all countries, the difference between the predictive ability of the two institutions is positive but very small.

An interesting question is whether there are clusters of countries for which the forecast performance of the two institutions differs dramatically. Dreher, Marchesi, and Vreeland (2008) test the hypothesis that the forecast performance of the IMF differs with respect to the

---

[4] It is worth noticing that we use the summer forecasts of the two organizations, and they are expected to be autocorrelated by construction. Nevertheless, as the autocorrelation test indicates the absence of autocorrelation, its effect must be negligible. Below, for the panel tests, we also use a different bandwidth for robustness, and this expectation is confirmed.

Table 10

Average loss differentials for the economic growth forecasts, DM test statistics, and their *p*-values, 1998–2016 (OECD *vs.* IMF).

Country	Statistics	Country	Statistics
Australia	−0.052	Iceland	−0.484
	(−0.451)		(−0.459)
	[0.652]		[0.646]
Austria	−0.035	Italy†	0.492
	(−0.293)		(1.047)
	[0.770]		[0.295]
Belgium	**0.555**	Japan†	0.191
	(1.701)		(1.040)
	[0.089]		[0.298]
Canada†	0.273	South Korea	0.571
	(1.504)		(0.564)
	[0.133]		[0.573]
Switzerland	0.272	Luxembourg	**1.951**
	(1.131)		**(1.757)**
	[0.258]		**[0.079]**
Czech Republic	−1.295	Mexico	0.009
	(−1.014)		(0.007)
	[0.311]		[0.994]
Germany†	−0.627	Netherlands	0.405
	(−1.125)		(1.549)
	[0.260]		[0.121]
Denmark	−0.150	Norway	−0.165
	(−0.771)		(−0.614)
	[0.440]		[0.539]
Spain	**−0.657**	New Zealand	−0.463
	(−1.726)		(−1.443)
	[0.084]		[0.149]
Finland	0.062	Poland	−0.477
	(0.079)		(−0.903)
	[0.937]		[0.367]
France†	0.216	Portugal	0.062
	(1.505)		(0.128)
	[0.132]		[0.898]
United Kingdom†	−0.143	Sweden	0.269
	(−0.740)		(0.501)
	[0.459]		[0.616]
Greece	−2.111	Türkiye	−0.746
	(−1.465)		(−0.306)
	[0.143]		[0.760]
Hungary	**1.303**	United States†	0.150
	(1.873)		(0.460)
	[0.061]		[0.645]
Ireland	0.875	*Average*	0.009
	(0.761)	*Average (G7)*	0.079
	[0.447]	*Average (Non-G7)*	−0.014

Note: [†] G7 countries. DM statistics in parentheses are calculated as $S_{i,T}^{(0)} = \sqrt{T}(\Delta\bar{L}_{i,T}/\hat{\sigma}_{i,T}) \xrightarrow{d} N(0,1)$ 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.

direct influence of a country on the institution. They use a country's GDP as a proxy of political influence and find evidence that the forecast bias of the institution declines with GDP. To see if we can find similar evidence regarding 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 accounted 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 the average losses of the G7 and non-G7 countries.

The table shows that the average quadratic loss for G7 countries is 0.079. For non-G7 countries, this average is −0.014. This shows that for G7 countries, the IMF has 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 CD's degree and nature. Furthermore, applying the correct EPA tests requires 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 our dataset's CD and autocorrelation properties.

The results of BP and Modified BP tests show that the null hypothesis of no CD is rejected using any test in conventional significance levels. The results are omitted to save space but are available from the authors. This means that the tests which allow for CD are more reliable for our data set. Next, IC_{p1} indicates the existence of six 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 (32) with $p = 1$, which was chosen by the general-to-specific 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, first we set $d_T = 1$. To see the impact of the change in the time series bandwidth, we also try $d_T = T^{1/3}$.

To see the time series profile of the common factors in the loss differential series, we report the plot of the first 6 PCs of the loss differential series in panels (a) and (b) of Fig. 2. The PCs are numbered in decreasing order with respect to their eigenvalues. For a better interpretation of the common factor estimates, we report a focus on the first three PCs of the quadratic loss differential. Associated factor loadings estimates are reported in Table 11. Estimated loadings for other PCs are dropped to save space, but they are available from the authors upon request. The estimates of the common factors in loss differentials show the effect of the financial crisis. The first three common factors in quadratic loss fell in 2009. As can be seen in Table 11, for 20 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 for 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
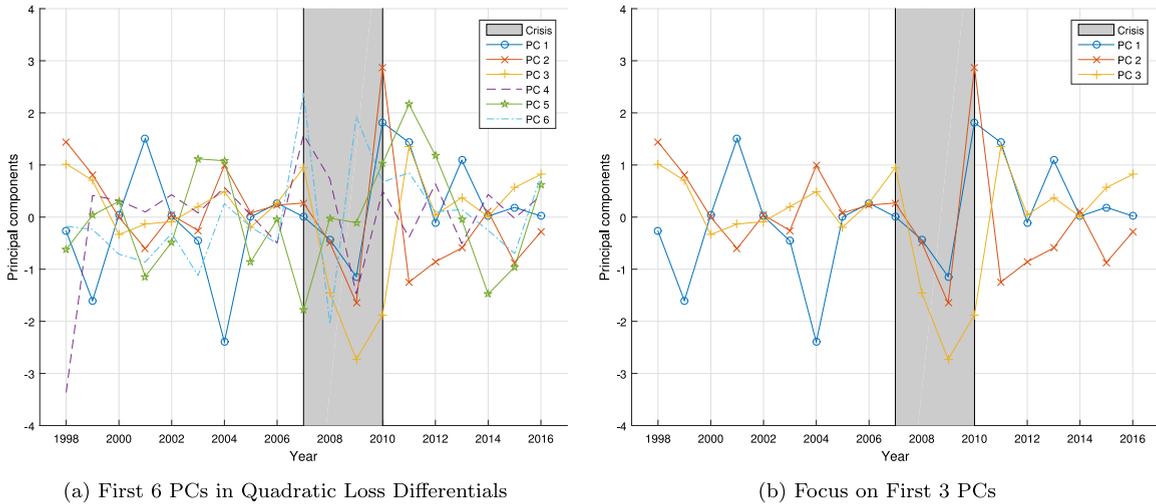
ARTICLE IN PRESS

O. Akgun, A. Pirotte, G. Urga et al.   International Journal of Forecasting xxx (xxxx) xxx

(a) First 6 PCs in Quadratic Loss Differentials

(b) Focus on First 3 PCs

Fig. 2. PC Estimates of the Common Factors in the Loss Differentials of the Economic Growth Forecasts Errors (OECD *vs.* IMF)

Table 11
PC estimates of the factor loadings in the loss differentials of the economic growth forecasts (OECD *vs.* IMF).

Countries	PC1	PC2	PC3
Australia	−0.05	−0.20	0.26
Austria	−0.14	0.27	−0.21
Belgium	0.37	0.26	−1.10
Canada	0.26	0.14	−0.20
Switzerland	−0.08	−0.48	−0.42
Czech Republic	0.07	−2.97	−1.71
Germany	−1.01	−1.66	0.65
Denmark	0.24	0.08	0.09
Spain	0.17	1.01	0.95
Finland	−1.06	−2.69	0.61
France	−0.08	−0.09	−0.28
United Kingdom	−0.08	0.09	0.58
Greece	−1.12	3.76	−3.58
Hungary	−0.03	−0.17	−2.38
Ireland	−1.55	−3.53	0.18
Iceland	0.79	−0.86	1.42
Italy	−0.65	−1.29	−0.85
Japan	0.04	−0.34	0.06
South Korea	−2.56	−2.31	1.10
Luxembourg	−1.04	−1.67	−3.53
Mexico	−2.13	−3.24	−1.47
Netherlands	−0.08	−0.70	−0.50
Norway	0.43	0.28	0.80
New Zealand	0.18	0.08	0.66
Poland	−0.78	0.66	−1.04
Portugal	−0.36	−0.10	−0.42
Sweden	−0.31	−0.56	−1.33
Türkiye	−10.32	2.03	0.93
United States	−0.50	−0.92	0.09

tests $S_{nT}^{(3)}$, $C_{nT}^{(3)}$, $\underline{S}_{nT}^{(3)}$ and $\underline{C}_{nT}^{(3)}$. 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 12. The first part of the table reports the tests of the overall EPA hypothesis. Hence, they are the significance tests of the overall difference reported in Table 10, which is 0.009. The test statistics for the non-robust test $S_{nT}^{(1)}$ are equal to 0.06 for both bandwidth parameters considered. The conclusion

does not change using the robust tests for which the test statistics are slightly weaker. Our results are in line with those obtained by QTZ on comparing IMF forecasts with Consensus Economic forecasts because they do not find a significant difference between the predictive ability of the IMF and Consensus Economics using EPA tests based on the summer forecasts of the two institutions.

The results for the clustered EPA tests are reported in the second part of the table. The clusters we consider are G7 countries and the non-G7 countries for which the average loss differentials are reported in Table 10. It is seen that for these clusters, the clustered EPA hypothesis cannot be rejected using any test and any bandwidth parameter. Hence, the predictive ability differences between the two institutions are insignificant.

6.2. An evaluation of the IMF consumer price inflation forecasts

In a second application, we challenge the quality of the IMF CPI forecasts by comparing them with random walk (RW) forecasts. In a similar exercise, QTZ compares IMF CPI inflation forecasts with those obtained from the AR(1) model. In this subsection, we summarize the application and its results. The details can be found in Appendix C. The data on the IMF CPI forecasts used in this application come from the Fund's Historical WEO Forecasts Database. The sample contains 127 countries over the period 1991–2019. We restrict our attention to the period also considered by QTZ, but we use the data from all countries available, contrary to their sample, which covers 86 countries. In this application, in addition to the quadratic loss function, we use the absolute loss function, defined as

$$\Delta L_{it}^{(a)} = |e_{1,it}| - |e_{2,it}|,$$

where $e_{1,it}$ is the forecast error made by the IMF for country i and year t, and $e_{2,it}$ is the RW forecast errors calculated using the CPI data published by the IMF. We consider two different country clustering schemes. The

Table 12
Panel tests of EPA for the economic growth forecasts (OECD *vs.* IMF).

	Overall EPA tests			Clustered EPA tests	
Test	$d_T = 0$	$d_T = T^{1/4}$	Test	$d_T = 0$	$d_T = T^{1/4}$
$S_{n,T}^{(1)}$	0.06	0.06	$C_{n,T}^{(1)}$	0.41	0.50
	(0.95)	(0.95)		(0.82)	(0.78)
$S_{n,T}^{(3)}$	0.04	0.04	$C_{n,T}^{(3)}$	0.42	0.59
	(0.97)	(0.97)		(0.81)	(0.74)
$\underline{S}_{n,T}^{(3)}$	0.04	0.04	$\underline{C}_{n,T}^{(3)}$	0.42	0.41
	(0.97)	(0.97)		(0.81)	(0.81)

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

first divides the sample of countries into OECD and non-OECD countries, and the second divides them into G7 countries, non-G7 OECD countries, and non-OECD countries. In addition to the full sample analysis, we split the sample into pre- and post-global financial crisis periods.

The results show a small but statistically significant difference between the average absolute losses of the IMF and the RW forecasts in favor of the latter in the full sample. When we divide the sample into pre- and post-crisis periods, we see that these differences are driven by the differences in the pre-crisis period. The clustered EPA hypothesis can be rejected for the absolute loss in both pre- and post-crisis periods.

Using quadratic loss, QTZ finds that the IMF CPI forecasts are significantly more accurate than the AR(1) forecasts. Our results show that IMF CPI forecasts are less accurate than RW forecasts. However, for this loss function, neither the overall EPA hypothesis nor clustered EPA hypothesis can be rejected at conventional levels using CD-robust tests.

7. Conclusions

In this paper, we proposed a 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. Consistent clustered EPA tests were also built, which can 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 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. We compare the prediction performance of two major organizations, the OECD and the IMF, on their historical economic growth forecasts. In a second application, we evaluate the CPI forecasts of the IMF by comparing them with random walk forecasts. We found evidence of strong cross-sectional dependence in loss differentials of forecast errors in both applications. The results showed that there are only minor differences between the predictive ability of the OECD and the IMF in terms of their economic growth forecasts, and

they are statistically insignificant. We found some evidence that the RW forecasts of consumer price inflation are significantly more accurate than those made by the IMF.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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 the 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

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

where $\mathbf{g}_{lt} = (g_{l1,t}, \ldots, g_{lm_l,t})'$ is an $m_l \times 1$ vector of common factors, $\boldsymbol{\theta}_{li} = (\theta_{li,1}, \ldots, \theta_{li,m_l})'$ 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 differentials, we will not focus on the weak dependence in the errors. We assume that $E(u_{l,it}|\mathbf{g}_t) = 0$ and $E(u_{l,it}^2|\mathbf{g}_t) = \sigma_l^2$ for $l = 1, 2$ where $\mathbf{g}_t = (\mathbf{g}_{1t}', \mathbf{g}_{2t}')'$. QTZ has used a similar factor model for forecast errors, but they impose $c_{li} = 0$ for all $i = 1, \ldots, n$, and $l = 1, 2$.

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

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

Hence, the model for the loss differentials is obtained with $m = m_1 + m_2$ in the case of the 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

$$
\begin{aligned}
e_{l,it}^2 &= \boldsymbol{\theta}_{li}'\mathbf{g}_{lt}\mathbf{g}_{lt}'\boldsymbol{\theta}_{li} + 2\boldsymbol{\theta}_{li}'\mathbf{g}_{lt}u_{l,it} + u_{l,it}^2 \\
&= \sum_{k_1=1}^{m_l}\sum_{k_2=1}^{m_l} \theta_{li,k_1}\theta_{li,k_2}g_{lk_1,t}g_{lk_2,t} + 2\boldsymbol{\theta}_{li}'\mathbf{g}_{lt}u_{l,it} + u_{l,it}^2 \\
&= \sum_{k=1}^{m_l^2} \gamma_{li,k}h_{lk,t} + 2\boldsymbol{\theta}_{li}'\mathbf{g}_{lt}u_{l,it} + u_{l,it}^2 \\
&= \boldsymbol{\gamma}_{li}'\boldsymbol{\iota}_{lt} + 2\boldsymbol{\theta}_{li}'\mathbf{g}_{lt}u_{l,it} + u_{l,it}^2,
\end{aligned}
$$

with straightforward definitions of $\gamma_{li,k}$ and $h_{lk,t}$ which are the kth elements of $m_l^2 \times 1$ vectors $\boldsymbol{\gamma}_{li}$ and $\boldsymbol{\iota}_{lt}$, respectively. Then, we obtain the model for the loss differentials as $E(d_{2,it}|\mathbf{g}_t) = \mu_i + \boldsymbol{\lambda}_i'\mathbf{f}_t$ with $\mu_i = \sigma_1^2 - \sigma_2^2$, $\boldsymbol{\lambda}_i = (\boldsymbol{\gamma}_{1i}', -\boldsymbol{\gamma}_{2i}')'$, $\mathbf{f}_t = (\boldsymbol{\iota}_{1t}', \boldsymbol{\iota}_{2t}')'$ 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 $\mathbf{H}_n$ be an $n \times G$ matrix which has $\boldsymbol{\iota}_{g_i}'$, the g_ith column of $\mathbf{I}_G$, in its ith row with $g_i \in \{1, 2, \ldots, G\}$ being a variable which states the cluster which ith unit belongs to. Define also $\boldsymbol{\epsilon}_{n,t} = (\epsilon_{1t}, \epsilon_{2t}, \ldots, \epsilon_{nt})'$, $\mathbf{c}_n = \text{diag}(c_1, c_2, \ldots, c_n)'$, $c_i = \sum_{h=0}^{\infty} c_{ih}$, $\boldsymbol{\Lambda}_n = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_n)'$, $\mathbf{C} = \sum_{h=0}^{\infty}\mathbf{C}_h$, $\tilde{\mathbf{f}}_t = \mathbf{f}_t - \frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t$, $\tilde{\boldsymbol{\epsilon}}_{n,t} = \boldsymbol{\epsilon}_{n,t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{n,t}$, $\mathbf{D}_n = \text{diag}(n_1, n_2, \ldots, n_G)$ and $\gamma_{ij,d_{ts}} = E(\epsilon_{it}\epsilon_{js})$.

Define

$$
\mathbf{V}_{1,nT} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\iota}_{g_i}\boldsymbol{\iota}_{g_i}'\bar{\gamma}_{i,T},
$$

$$
\mathbf{V}_{2,nT} = \frac{1}{n}\sum_{i,j=1}^{n} \boldsymbol{\iota}_{g_i}\boldsymbol{\iota}_{g_j}'\mathbf{r}_{i.}'\bar{\boldsymbol{\gamma}}_{nT}\mathbf{r}_{j.},
$$

$$
\mathbf{V}_{3,nT} = \frac{1}{n^2}\sum_{i,j=1}^{n} \boldsymbol{\iota}_{g_i}\boldsymbol{\iota}_{g_j}'\boldsymbol{\lambda}_i'\bar{\boldsymbol{\Gamma}}_T\boldsymbol{\lambda}_j + \frac{1}{n^2}\sum_{i,j=1}^{n} \boldsymbol{\iota}_{g_i}\boldsymbol{\iota}_{g_j}'\mathbf{r}_{i.}'\bar{\boldsymbol{\gamma}}_{nT}\mathbf{r}_{j.}.
$$

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

Lemma 1. *Suppose Assumption 1, 3 and 6 hold. Then, as $T, n \to \infty$,*

(i) $\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\mathbf{H}_n'\boldsymbol{\epsilon}_{n,t} \overset{d}{\to} N(\mathbf{0}, \mathbf{V}_1)$, *where* $\mathbf{V}_1 = \lim_{(n,T)\to\infty} \mathbf{V}_{1,nT}$,

(ii) $\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\mathbf{H}_n'\mathbf{R}_n\boldsymbol{\epsilon}_{n,t} \overset{d}{\to} N(\mathbf{0}, \mathbf{V}_2)$, *where* $\mathbf{V}_2 = \lim_{(n,T)\to\infty} \mathbf{V}_{2,nT}$,

(iii) $\frac{1}{n\sqrt{T}}\sum_{t=1}^{T}\mathbf{H}_n'\boldsymbol{\Lambda}_n\mathbf{f}_t \overset{d}{\to} N(\mathbf{0}, \mathbf{V}_3)$, *where* $\mathbf{V}_3 = \lim_{(n,T)\to\infty} \mathbf{V}_{3,nT}$.

Lemma 2. *Suppose Assumption 1, 2 and 6 hold. Then, for any fixed n, as $T \to \infty$*

(i) $T^{-1}\sum_{t,s=1}^{T} k_T (d_{ts}/d_T) \mathbf{f}_t\mathbf{f}_s' - \bar{\boldsymbol{\Gamma}}_T = o_p(1)$,

(ii) $T^{-1}\sum_{t,s=1}^{T} k_T (d_{ts}/d_T) \tilde{\mathbf{f}}_t\tilde{\mathbf{f}}_s' - \bar{\boldsymbol{\Gamma}}_T = o_p(1)$,

(iii) $T^{-1}\sum_{t,s=1}^{T} k_T (d_{ts}/d_T) \tilde{\mathbf{f}}_t\tilde{\boldsymbol{\epsilon}}_{n,s}' = o_p(1)$,

(iv) $T^{-1}\sum_{t,s=1}^{T} k_T (d_{ts}/d_T) \tilde{\boldsymbol{\epsilon}}_{n,t}\tilde{\boldsymbol{\epsilon}}_{n,t}' - \bar{\boldsymbol{\gamma}}_{nT} = o_p(1)$.

Lemma 3. *Suppose Assumption 1, 3, 6 and 7 hold. Then,*

(i) $\frac{1}{n^2T}\sum_{i=1}^{n}\sum_{t,s=1}^{T} k_T (d_{ts}/d_T) \widehat{\varepsilon}_{it}\widehat{\varepsilon}_{is} = o_p(1)$,

(ii) $\frac{1}{n^2}\sum_{i,j=1}^{n}\left[\frac{1}{T}\sum_{t,s=1}^{T} k_T (d_{ts}/d_T)\left(\widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{f}}_t\widehat{\boldsymbol{\lambda}}_j'\widehat{\mathbf{f}}_s - \boldsymbol{\lambda}_i'\mathbf{f}_t\mathbf{f}_s'\boldsymbol{\lambda}_j\right)\right]$
$= o_p(1)$.

Proof of Lemma 1. (i) is a special case of (ii) with $\mathbf{R}_n = \mathbf{I}_n$; hence we focus on the two last results. We start by applying the Beveridge–Nelson (BN) decomposition (see, for instance, Phillips & Solo, 1992) to each component of $\boldsymbol{\epsilon}_{n,t}$. 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

$$
\begin{aligned}
\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\mathbf{H}_n'\mathbf{R}_n\boldsymbol{\epsilon}_{n,t} &= \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\mathbf{H}_n'\mathbf{R}_n\mathbf{c}_n\boldsymbol{\psi}_{n,t} \\
&\quad + \frac{1}{\sqrt{nT}}\mathbf{H}_n'\mathbf{R}_n\tilde{\boldsymbol{\psi}}_{n,1} - \frac{1}{\sqrt{nT}}\mathbf{H}_n'\mathbf{R}_n\tilde{\boldsymbol{\psi}}_{n,T} \\
&= A_1 + A_2 + A_3,
\end{aligned}
$$

where $\boldsymbol{\psi}_{n,t} = (\psi_{1t}, \psi_{2t}, \ldots, \psi_{nt})'$ and $\tilde{\boldsymbol{\psi}}_{n,t} = (\tilde{\psi}_{1t}, \tilde{\psi}_{2t}, \ldots, \tilde{\psi}_{nt})'$. The variance of the first term is $\text{Var}(A_1) = n^{-1}\sum_{i,j=1}^{n}\boldsymbol{\iota}_{g_i}\mathbf{r}_{i.}'\mathbf{c}_n\mathbf{c}_n'\mathbf{r}_{j.}\boldsymbol{\iota}_{g_j}'$. A typical element of this matrix, say lwth, satisfies

$$
\begin{aligned}
&\left| \frac{1}{n}\sum_{i \in G_l}\sum_{j \in G_w}\mathbf{r}_{i.}'\mathbf{c}_i\mathbf{c}_j'\mathbf{r}_{j.} \right| \\
&= \left| \frac{1}{n}\sum_{i \in G_l}\sum_{j \in G_w}\sum_{k=1}^{n}r_{ik}r_{jk}\sum_{h=0}^{\infty}\sum_{h'=0}^{\infty}c_{kh}c_{kh'} \right| \\
&\leq \frac{1}{n}\sum_{i \in G_l}\sum_{j \in G_w}\sum_{k=1}^{n}|r_{ik}||r_{jk}|\sum_{h=0}^{\infty}\sum_{h'=0}^{\infty}|c_{kh}||c_{kh'}| \\
&\leq \frac{1}{n}\sum_{k=1}^{n}\sum_{i \in G_l}|r_{ik}|\sum_{j \in G_w}|r_{jk}|\sum_{h=0}^{\infty}|c_{kh}|\sum_{h'=0}^{\infty}|c_{kh'}| = O(1),
\end{aligned}
$$

where we used the fact that absolute summability is implied by the summability condition we impose on co-efficients c_{ih} in Assumption 1. Hence we showed that A_1 is $O_p(1)$. The variances of the second and the third terms are $\mathrm{Var}\,(A_2) = (nT)^{-1}\sum_{i,j=1}^{n}\iota_{g_i}\mathbf{r}'_{i.}\mathrm{E}(\tilde{\boldsymbol{\psi}}_{n,1}\tilde{\boldsymbol{\psi}}'_{n,1})\mathbf{r}_{j.}\iota'_{g_j}$ and $\mathrm{Var}\,(A_3) = (nT)^{-1}\sum_{i,j=1}^{n}\iota_{g_i}\mathbf{r}'_{i.}\mathrm{E}(\tilde{\boldsymbol{\psi}}_{n,T}\tilde{\boldsymbol{\psi}}'_{n,T})\mathbf{r}_{j.}\iota'_{g_j}$. respectively. For (ii) to hold, we need to show that these are asymptotically negligible. This holds if elements of $\mathrm{E}(\tilde{\boldsymbol{\psi}}_{n,1}\tilde{\boldsymbol{\psi}}'_{n,1})$ and $\mathrm{E}(\tilde{\boldsymbol{\psi}}_{n,T}\tilde{\boldsymbol{\psi}}'_{n,T})$ are finite which holds under Assumption 1 as shown by the BN Lemma in Phillips and Solo (1992). Then the lwth elements of $\mathrm{Var}\,(A_2)$ and $\mathrm{Var}\,(A_3)$ satisfy $(nT)^{-1}\left|\sum_{i\in G_l}\sum_{j\in G_w}\mathbf{r}'_{i.}\mathbf{r}_{j.}O(1)\right|$ which is $O(T^{-1})$ under Assumption 3. 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 & Prucha, 1998) and (ii) follows.

To prove (iii), we apply the multivariate generalization of the BN decomposition (see, for instance, Phillips & Solo, 1992, p. 985) to $\mathbf{f}_t$. We have

$$\frac{1}{n\sqrt{T}}\sum_{t=1}^{T}\mathbf{H}'_n\boldsymbol{\Lambda}_n\mathbf{f}_t = \frac{1}{n\sqrt{T}}\sum_{t=1}^{T}\mathbf{H}'_n\boldsymbol{\Lambda}_n\mathbf{C}\,\boldsymbol{\Psi}_t + \frac{1}{n\sqrt{T}}\mathbf{H}'_n\boldsymbol{\Lambda}_n\,\tilde{\boldsymbol{\Psi}}_1$$

$$-\frac{1}{n\sqrt{T}}\mathbf{H}'_n\boldsymbol{\Lambda}_n\,\tilde{\boldsymbol{\Psi}}_T$$

$$=B_1 + B_2 + B_3,$$

where $\tilde{\boldsymbol{\Psi}}_t = \sum_{h=0}^{\infty}\tilde{\mathbf{C}}_h\boldsymbol{\Psi}_{t-h}$, $\tilde{\mathbf{C}}_h = \sum_{j=h+1}^{\infty}\mathbf{C}_j$. The variance of the first term is given by $\mathrm{Var}\,(B_1) = \frac{1}{n^2}\sum_{i,j=1}^{n}\iota_{g_i}\boldsymbol{\lambda}'_i\mathbf{C}\mathbf{C}'\boldsymbol{\lambda}_j\iota'_{g_j}$. The lwth element of this matrix satisfies

$$\left|\frac{1}{n^2}\sum_{i\in G_l}\sum_{j\in G_w}\boldsymbol{\lambda}'_i\mathbf{C}\mathbf{C}'\boldsymbol{\lambda}_j\right| \le \frac{1}{n^2}\sum_{i\in G_l}\sum_{j\in G_w}\left|\boldsymbol{\lambda}'_i\mathbf{C}\mathbf{C}'\boldsymbol{\lambda}_j\right|$$

$$\le \frac{1}{n^2}\sum_{i\in G_l}\sum_{j\in G_w}\|\mathbf{C}\|^2\|\boldsymbol{\lambda}_i\|\|\boldsymbol{\lambda}_j\| = O(1),$$

where we used the fact that $\|\mathbf{C}\|^2 \le \sum_{h=0}^{\infty}\|\mathbf{C}_h\|^2 < \infty$. It follows that B_1 is $O_p(1)$. The variances of the second and the third terms are $\mathrm{Var}\,(B_2) = \frac{1}{n^2 T}\mathbf{H}'_n\boldsymbol{\Lambda}_n\mathrm{E}(\tilde{\boldsymbol{\Psi}}_1\tilde{\boldsymbol{\Psi}}'_1)\boldsymbol{\Lambda}'_n\mathbf{H}_n$ and $\mathrm{Var}\,(B_3) = \frac{1}{n^2 T}\mathbf{H}'_n\boldsymbol{\Lambda}_n\mathrm{E}(\tilde{\boldsymbol{\Psi}}_T\tilde{\boldsymbol{\Psi}}'_T)\boldsymbol{\Lambda}'_n\mathbf{H}_n$, respectively. Similar to the reasoning above, we have $\mathrm{E}(\tilde{\boldsymbol{\Psi}}_1\tilde{\boldsymbol{\Psi}}'_1) < \infty$ and $\mathrm{E}(\tilde{\boldsymbol{\Psi}}_T\tilde{\boldsymbol{\Psi}}'_T) < \infty$, under the summability condition we impose on the coefficients of the process $\mathbf{f}_T$ (see, for instance Phillips & Magdalinos, 2009). Furthermore, the lwth element of the matrix $\frac{1}{n^2}\mathbf{H}'_n\boldsymbol{\Lambda}_n\boldsymbol{\Lambda}'_n\mathbf{H}_n$ satisfies $\left|n^{-2}\sum_{i\in G_l}\sum_{j\in G_w}\sum_{k,k'=1}^{m}\lambda_{ik}\lambda_{jk'}\right| = O(1)$. Hence, B_2 and B_3 are $o_p(1)$. Since the first term is $O_p(1)$, it dominates the second and third terms, and a central limit theorem for triangular arrays (Kelejian & 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) follows from the proof of Theorem 2 of Jansson (2002) under Assumptions 1 and 2. (ii) follows immediately from their proof noting that

$T^{-1}\sum_{t=1}^{T}\mathbf{f}_t = o_p(1)$ under Assumption 66. (iii) holds again by their proof, as by Assumption 66 $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_{n,s}$ are independent for every t, s, and noting that $T^{-1}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{n,t} = o_p(1)$. Similarly, (iv) is a result of their proof under Assumption 1.

Proof of Lemma 3. We have $\widehat{\varepsilon}_{it} = \Delta L_{it} - \Delta\bar{L}_{i,T} - \widehat{\boldsymbol{\lambda}}'_i\widehat{\mathbf{f}}_t = (\mu_i - \Delta\bar{L}_{i,T}) + (\boldsymbol{\lambda}'_i\mathbf{f}_t - \widehat{\boldsymbol{\lambda}}'_i\widehat{\mathbf{f}}_t) + \varepsilon_{it}$. By Assumptions 1 and 6, $\mu_i - \Delta\bar{L}_{i,T} = O_p(T^{-1/2})$, $\varepsilon_{it} = O_p(1)$ and by Assumption 7 $\boldsymbol{\lambda}'_i\mathbf{f}_t - \widehat{\boldsymbol{\lambda}}'_i\widehat{\mathbf{f}}_t = O_p(\delta_{nT}^{-1})$. After trivial algebra, this gives $\widehat{\varepsilon}_{it}\widehat{\varepsilon}_{is} = \varepsilon_{it}\varepsilon_{is} + O_p(\delta_{nT}^{-1})$. Then $(n^2 T)^{-1}\sum_{i=1}^{n}\sum_{t,s=1}^{T}k_T\,(d_{ts}/d_T)\,\widehat{\varepsilon}_{it}\widehat{\varepsilon}_{is} = (n^2 T)^{-1}\sum_{i=1}^{n}\sum_{t,s=1}^{T}k_T\,(d_{ts}/d_T)\,O_p(\delta_{nT}^{-1}) + (n^2 T)^{-1}\sum_{i=1}^{n}\sum_{t,s=1}^{T}k_T\,(d_{ts}/d_T)\,\varepsilon_{it}\varepsilon_{is} = A_1 + A_2$, say. We have $A_1 = n^{-2}\sum_{i=1}^{n}O_p(\delta_{nT}^{-1})\left[T^{-1}\sum_{h=-T+1}^{T-1}k_T\,(h/d_T)\right]$ where the term in brackets is $O(d_T/T)$ by Assumption 2, which in turn gives $A_1 = O(n^{-1})O_p(\delta_{nT}^{-1})O(d_T/T) = o_p(1)$. For the second term, we find $\mathrm{E}(A_2) = \frac{1}{n^2 T}\sum_{i=1}^{n}\sum_{t,s=1}^{T}k_T\,(d_{ts}/d_T)\,\mathbf{r}'_{i.}\boldsymbol{\gamma}_{n,d_{ts}}\mathbf{r}_{i.} = n^{-2}\sum_{i,k=1}^{n}r_{ik}^2\left[T^{-1}\sum_{h=-T+1}^{T-1}k_T\,(h/d_T)\,\gamma_{k,h}\right]$. By Assumption 1 the spectral density function of ϵ_{it}, $f_i(\cdot)$, exists and is bounded. By Theorem 2 of Jansson (2002), under Assumption 2, $\lim_{T\to\infty}T^{-1}\sum_{h=-T+1}^{T-1}k_T\,(h/d_T)\,\gamma_{i,h} = T^{-1}2\pi f_i(0) = O(T^{-1})$ for each i. Then we have $\mathrm{E}(A_2) = O(T^{-1})n^{-2}\sum_{i,k=1}^{n}r_{ik}^2 = O[(nT)^{-1}]$. For the variance, we find

$$\mathrm{Var}(A_2) = \mathrm{E}\left[\frac{1}{n^4 T^2}\sum_{i,j=1}^{n}\sum_{t_1,s_1,t_2,s_2=1}^{T}k_T\left(\frac{d_{t_1 s_1}}{d_T}\right)k_T\right.$$

$$\times\left.\left(\frac{d_{t_2 s_2}}{d_T}\right)\varepsilon_{it_1}\varepsilon_{is_1}\varepsilon_{jt_2}\varepsilon_{js_2}\right]$$

$$= \frac{1}{n^4 T^2}\sum_{i,j=1}^{n}\sum_{t_1,s_1,t_2,s_2=1}^{T}k_T\left(\frac{d_{t_1 s_1}}{d_T}\right)k_T$$

$$\times\left(\frac{d_{t_2 s_2}}{d_T}\right)\mathrm{E}(\varepsilon_{it_1}\varepsilon_{is_1}\varepsilon_{jt_2}\varepsilon_{js_2})$$

$$= \frac{1}{n^4 T^2}\sum_{i,j=1}^{n}\sum_{t_1,s_1,t_2,s_2=1}^{T}k_T\left(\frac{d_{t_1 s_1}}{d_T}\right)k_T$$

$$\times\left(\frac{d_{t_2 s_2}}{d_T}\right)\mathrm{E}(\mathbf{r}'_{i.}\boldsymbol{\epsilon}_{n,t_1}\boldsymbol{\epsilon}'_{n,s_1}\mathbf{r}_{i.}\mathbf{r}'_{j.}\boldsymbol{\epsilon}_{n,t_2}\boldsymbol{\epsilon}'_{n,s_2}\mathbf{r}_{j.})$$

$$= \frac{1}{n^4 T^2}\sum_{i,j=1}^{n}\sum_{t_1,s_1,t_2,s_2=1}^{T}k_T\left(\frac{d_{t_1 s_1}}{d_T}\right)k_T\left(\frac{d_{t_2 s_2}}{d_T}\right)$$

$$\times\mathrm{E}\left(\sum_{l_1,l_2=1}^{n}r_{il_1}r_{il_2}\epsilon_{l_1,t_1}\epsilon_{l_2,s_1}\sum_{l_3,l_4=1}^{n}r_{jl_3}r_{jl_4}\epsilon_{l_3,t_2}\epsilon_{l_4,s_2}\right)$$

$$= \frac{1}{n^4}\sum_{i,j,l_1,l_2,l_3,l_4=1}^{n}r_{il_1}r_{il_2}r_{jl_3}r_{jl_4}\left[\frac{1}{T^2}\sum_{t_1,s_1,t_2,s_2=1}^{T}k_T\right.$$

$$\times\left.\left(\frac{d_{t_1 s_1}}{d_T}\right)k_T\left(\frac{d_{t_2 s_2}}{d_T}\right)\mathrm{E}(\epsilon_{l_1,t_1}\epsilon_{l_2,s_1}\epsilon_{l_3,t_2}\epsilon_{l_4,s_2})\right].$$

$$(33)$$

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

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

where $\kappa(\cdot)$ is the cumulant of the fourth order between ϵ_{l_1,t_1}, ϵ_{l_2,s_1}, ϵ_{l_3,t_2} and ϵ_{l_4,s_2} (see Hannan, 1970, p. 23). In our case, by Assumption 1 ϵ_{it} are

$$\text{Var}(A_2) = \frac{1}{n^4}\sum_{i,j=1}^{n}\sum_{l_1=1}^{n}r_{il_1}^2\left[\frac{1}{T}\sum_{h_1=-T+1}^{T-1}k_T\left(\frac{h_1}{d_T}\right)\gamma_{l_1,h_1}\right]$$
$$\times \sum_{l_2=1}^{n}r_{jl_2}^2\left[\frac{1}{T}\sum_{h_2=-T+1}^{T-1}k_T\left(\frac{h_2}{d_T}\right)\gamma_{l_2,h_2}\right]$$
$$+ \frac{1}{n^4}\sum_{i,j=1}^{n}\sum_{l_1=1}^{n}r_{il_1}^2\left[\frac{1}{T}\sum_{h_1=-T+1}^{T-1}k_T\left(\frac{h_1}{d_T}\right)\gamma_{l_1,h_1}\right]$$
$$\times \sum_{l_3=1}^{n}r_{jl_3}^2\left[\frac{1}{T}\sum_{h_2=-T+1}^{T-1}k_T\left(\frac{h_2}{d_T}\right)\gamma_{l_3,h_2}\right]$$
$$+ \frac{1}{n^4}\sum_{i,j=1}^{n}\sum_{l_1=1}^{n}r_{il_1}^2\left[\frac{1}{T}\sum_{h_1=-T+1}^{T-1}k_T\left(\frac{h_1}{d_T}\right)\gamma_{l_1,h_1}\right]$$
$$\times \sum_{l_4=1}^{n}r_{jl_4}^2\left[\frac{1}{T}\sum_{h_2=-T+1}^{T-1}k_T\left(\frac{h_2}{d_T}\right)\gamma_{l_4,h_2}\right]$$
$$= A_{21} + A_{22} + A_{23}.$$

All terms in the brackets are $O(T^{-1})$. Hence, $A_{21} = O(T^{-2})n^{-4}\sum_{i,j=1}^{n}\sum_{l_1=1}^{n}r_{il_1}^2\sum_{l_2=1}^{n}r_{jl_2}^2 = O[(nT)^{-2}]$. Similarly, $A_{22} = A_{23} = O[(nT)^{-2}]$. As a result, $A_2 = O_p[(nT)^{-1}] = o_p(1)$.

For (ii), we write $\widehat{\lambda_i'\widehat{\mathbf{f}_t}\widehat{\mathbf{f}_s}'\widehat{\lambda}_j} - \lambda_i'\mathbf{f}_t\mathbf{f}_s'\lambda_j = (\widehat{\lambda_i'\widehat{\mathbf{f}_t}} - \lambda_i'\mathbf{f}_t)\widehat{\mathbf{f}_s}'\widehat{\lambda}_j + \lambda_i'\mathbf{f}_t(\widehat{\mathbf{f}_s'\widehat{\lambda}_j} - \mathbf{f}_s'\lambda_j)$. By Assumption 6 $\lambda_i'\mathbf{f}_t$, and by Assumption 7 $\widehat{\lambda_i'\widehat{\mathbf{f}_t}}$ are $O_p(1)$. This implies that $\widehat{\lambda_i'\widehat{\mathbf{f}_t}\widehat{\mathbf{f}_s}'\widehat{\lambda}_j} - \lambda_i'\mathbf{f}_t\mathbf{f}_s'\lambda_j = O_p(\delta_{nT}^{-1})$ by Assumption 7. 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 $\mathbf{R}_n = \mathbf{I}_n$ 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 $\mathbf{R}_n = \mathbf{I}_n$.

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

$$\Delta\bar{\mathbf{L}}_{nT} = \mathbf{D}_n^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{H}_n'\Delta\mathbf{L}_{n,t} = \mathbf{D}_n^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{H}_n'(\boldsymbol{\mu} + \mathbf{R}_n\boldsymbol{\epsilon}_{n,t})$$
$$= \bar{\boldsymbol{\mu}}_n + \mathbf{D}_n^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{H}_n'\mathbf{R}_n\boldsymbol{\epsilon}_{n,t},$$

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

$$\sqrt{nT}(\Delta\bar{\mathbf{L}}_{nT} - \bar{\boldsymbol{\mu}}_n) = \left(\frac{\mathbf{D}_n}{n}\right)^{-1}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\mathbf{H}_n'\mathbf{R}_n\boldsymbol{\epsilon}_{n,t}.$$

Then $\sqrt{nT}\boldsymbol{\Omega}_{2,nT}^{-1/2}(\Delta\bar{\mathbf{L}}_{nT} - \bar{\boldsymbol{\mu}}_n) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}_G)$, where $\boldsymbol{\Omega}_{2,nT} = \sum_{i,j=1}^{n}\frac{n}{n_{g_i}n_{g_j}}\boldsymbol{\iota}_{g_i}\boldsymbol{\iota}_{g_j}'\mathbf{r}_{i.}'\bar{\boldsymbol{\gamma}}_{nT}\mathbf{r}_{j.}$ by Lemma 1(ii) and noting that $n^{-1}\mathbf{D}_n$ converges to a finite and nonsingular matrix under Assumption 8. The matrix $\boldsymbol{\Omega}_{2,nT}$ can be written as $\boldsymbol{\Omega}_{2,nT} = (\mathbf{D}_n/n)^{-1}\mathbf{V}_{2,nT}(\mathbf{D}_n/n)^{-1}$. Since $\mathbf{D}_n$ is known, estimation of $\boldsymbol{\Omega}_{2,nT}$, requires only the estimation of $\mathbf{V}_{2,nT}$. Similarly, we write $\widehat{\boldsymbol{\Omega}}_{2,nT} = (\mathbf{D}_n/n)^{-1}\widehat{\mathbf{V}}_{2,nT}(\mathbf{D}_n/n)^{-1}$ where

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

The lwth element of the matrices $\mathbf{V}_{2,nT}$ and $\widehat{\mathbf{V}}_{2,nT}$ are

$$v_{2,nT}^{lw} = \frac{1}{n}\sum_{i\in G_l}\sum_{j\in G_w}\mathbf{r}_{i.}'\bar{\boldsymbol{\gamma}}_{nT}\mathbf{r}_{j.}, \tag{34}$$

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 leads to the first result in the proposition. We have $\Delta\tilde{L}_{it} = \mathbf{r}_{i.}'\tilde{\boldsymbol{\epsilon}}_{n,t}$ which gives $\Delta\tilde{L}_{it}\Delta\tilde{L}_{js} = \mathbf{r}_{i.}'\tilde{\boldsymbol{\epsilon}}_{n,t}\tilde{\boldsymbol{\epsilon}}_{n,s}'\mathbf{r}_{j.}$. Then,

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

Since $\bar{\boldsymbol{\gamma}}_{nT} = O(1)$, and $T^{-1}\sum_{t,s=1}^{T}k_T(d_{ts}/d_T)\tilde{\boldsymbol{\epsilon}}_{n,t}\tilde{\boldsymbol{\epsilon}}_{n,s}' - \bar{\boldsymbol{\gamma}}_{nT} = o_p(1)$ by Lemma 2(iv), it suffices to show that $\frac{1}{n}\sum_{i\in G_l}\sum_{j\in G_w}k_S(d_{ij}/d_n)\mathbf{r}_{i.}'\mathbf{r}_{j.} = O(1)$ and $\frac{1}{n}\sum_{i\in G_l}\sum_{j\in G_w}$

$[1 - k_S(d_{ij}/d_n)]\mathbf{r}'_{i.}\mathbf{r}_{j.} = o(1)$ in order to prove the consistency of $\hat{v}^{lw}_{2,nT}$. Starting with the latter, we have

$$\left| \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} \left[1 - k_S \left(\frac{d_{ij}}{d_n} \right) \right] \mathbf{r}'_{i.}\mathbf{r}_{j.} \right|$$

$$\leq \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} \left| 1 - k_S \left(\frac{d_{ij}}{d_n} \right) \right| |\mathbf{r}'_{i.}\mathbf{r}_{j.}|$$

$$\leq \frac{1}{n d_n^{\rho_s}} \sum_{i \in G_l} \sum_{j \in G_w} |\mathbf{r}'_{i.}\mathbf{r}_{j.}| d_{ij}^{\rho_s} = O\left(\frac{1}{d_n^{\rho_s}} \right) = o(1),$$

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

$$\frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} k_S \left(\frac{d_{ij}}{d_n} \right) \mathbf{r}'_{i.}\mathbf{r}_{j.} = \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w} \mathbf{r}'_{i.}\mathbf{r}_{j.} - \frac{1}{n} \sum_{i \in G_l} \sum_{j \in G_w}$$

$$\times \left[1 - k_S \left(\frac{d_{ij}}{d_n} \right) \right] \mathbf{r}'_{i.}\mathbf{r}_{j.} = O(1),$$

which follows from the fact that the first term is $O(1)$ and the second term is $o(1)$ by the previous equation. Hence, $\hat{v}^{lw}_{2,nT} - v^{lw}_{2,nT} = o_p(1)$ and the result follows. This completes the proof of (i).

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

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

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

$$\hat{\underline{v}}^{lw}_{2,nT} = \frac{1}{\underline{n}T} \sum_{i \in \underline{G}_l} \sum_{j \in \underline{G}_w} \mathbf{r}'_{i.} \left[\frac{1}{T} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \tilde{\boldsymbol{\epsilon}}_{n,t} \tilde{\boldsymbol{\epsilon}}'_{n,t} \right] \mathbf{r}_{j.},$$

where G_g, $g = 1, \ldots, G$, 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 $\underline{n}_g$. Using this expression together with (34), we can write

$$\hat{\underline{v}}^{lw}_{2,nT} - v^{lw}_{2,nT} = (\hat{\underline{v}}^{lw}_{2,nT} - \underline{v}^{lw}_{2,nT}) + (\underline{v}^{lw}_{2,nT} - v^{lw}_{2,nT}). \qquad (35)$$

where $\underline{v}^{lw}_{2,nT} = \frac{1}{\underline{n}} \sum_{i \in \underline{G}_l} \sum_{j \in \underline{G}_w} \mathbf{r}'_{i.} \bar{\boldsymbol{\gamma}}_{nT} \mathbf{r}_{j.}$. Our objective is to show that both terms in parentheses approach zero. We have

$$\hat{\underline{v}}^{lw}_{2,nT} - \underline{v}^{lw}_{2,nT} = \frac{1}{\underline{n}} \sum_{i \in \underline{G}_l} \sum_{j \in \underline{G}_w} \mathbf{r}'_{i.} \left[\frac{1}{T} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \tilde{\boldsymbol{\epsilon}}_{n,t} \tilde{\boldsymbol{\epsilon}}'_{n,t} - \bar{\boldsymbol{\gamma}}_{nT} \right] \mathbf{r}_{j.}.$$

The term in brackets is $o_p(1)$ by Lemma 2(iv). Then it is sufficient to show that $\frac{1}{\underline{n}} \sum_{i \in \underline{G}_l} \sum_{j \in \underline{G}_w} \mathbf{r}'_{i.}\mathbf{r}_{j.}$ is bounded. We have

$$\left| \frac{1}{\underline{n}} \sum_{i \in \underline{G}_l} \sum_{j \in \underline{G}_w} \mathbf{r}'_{i.}\mathbf{r}_{j.} \right| = \left| \frac{1}{\underline{n}} \sum_{i \in \underline{G}_l} \sum_{j \in \underline{G}_w} \sum_{k=1}^{n} r_{ik} r_{jk} \right|$$

$$\leq \frac{1}{\underline{n}} \sum_{i \in \underline{G}_l} \sum_{k=1}^{n} |r_{ik}| \sum_{j \in \underline{G}_w} |r_{jk}| = \frac{\underline{n}_l}{\underline{n}} O(1),$$

which is $O(1)$ because $\underline{n}_l/\underline{n} \to \tau_l \in (0, 1)$. For the second term in (35) we note that, by definition, $\lim_{(n,T) \to \infty} \underline{v}^{lw}_{2,nT} =$

$\lim_{(n,T) \to \infty} v^{lw}_{2,nT} = v^{lw}_2$ where v^{lw}_2 is the lwth element of the matrix $\mathbf{V}_2$ defined in Lemma 1(ii).

Proof of Theorem 2. We have

$$\Delta \bar{\mathbf{L}}_{nT} = \bar{\boldsymbol{\mu}}_n + \mathbf{D}_n^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbf{H}'_n \mathbf{R}_n \boldsymbol{\epsilon}_{n,t} + \mathbf{D}_n^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbf{H}_n^{-1} \boldsymbol{\Lambda}_n \mathbf{f}_t,$$

from which we obtain

$$\sqrt{T}(\Delta \bar{\mathbf{L}}_{nT} - \bar{\boldsymbol{\mu}}_n) = \left(\frac{\mathbf{D}_n}{n} \right)^{-1} \frac{1}{n\sqrt{T}} \sum_{t=1}^{T} \mathbf{H}'_n \mathbf{R}_n \boldsymbol{\epsilon}_{n,t}$$

$$+ \left(\frac{\mathbf{D}_n}{n} \right)^{-1} \frac{1}{n\sqrt{T}} \sum_{t=1}^{T} \mathbf{H}'_n \boldsymbol{\Lambda}_n \mathbf{f}_t.$$

From Lemma 1(ii) it follows that $\frac{1}{n\sqrt{T}} \sum_{t=1}^{T} \mathbf{H}'_n \mathbf{R}_n \boldsymbol{\epsilon}_{n,t} = O_p(n^{-1/2})$. Then, $\sqrt{T} \boldsymbol{\Omega}_{3,nT}^{-1/2}(\Delta \bar{\mathbf{L}}_{nT} - \bar{\boldsymbol{\mu}}_n) \overset{D}{\to} N(\mathbf{0}, \mathbf{I}_G)$, where $\boldsymbol{\Omega}_{3,nT} = \sum_{i,j=1}^{n} \frac{1}{n_{g_i} n_{g_j}} \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}'_{g_j} (\boldsymbol{\lambda}'_i \bar{\boldsymbol{\Gamma}}_T \boldsymbol{\lambda}_j + \mathbf{r}'_{i.} \bar{\boldsymbol{\gamma}}_{nT} \mathbf{r}_{j.})$ by noting that $n^{-1}\mathbf{D}_n$ converges to a finite and nonsingular matrix under Assumption 8 and Lemma 1(iii). The matrix $\boldsymbol{\Omega}_{3,nT}$ can be written as $\boldsymbol{\Omega}_{3,nT} = (\mathbf{D}_n/n)^{-1}\mathbf{V}_{3,nT}(\mathbf{D}_n/n)^{-1}$. As in the proof of Theorem 1, estimation of $\boldsymbol{\Omega}_{2,nT}$, requires only the estimation of $\mathbf{V}_{2,nT}$ because $\mathbf{D}_n$ is known. Write $\widehat{\boldsymbol{\Omega}}_{3,nT} = (\mathbf{D}_n/n)^{-1}\widehat{\mathbf{V}}_{3,nT}(\mathbf{D}_n/n)^{-1}$ where

$$\widehat{\mathbf{V}}_{3,nT} = \frac{1}{n^2 T} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \boldsymbol{\iota}_{g_i} \boldsymbol{\iota}'_{g_j} \Delta \tilde{L}_{it} \Delta \tilde{L}_{js}.$$

The lwth element of the matrices $\mathbf{V}_{3,nT}$ and $\widehat{\mathbf{V}}_{3,nT}$, corresponding to clusters l and w are

$$v^{lw}_{3,nT} = \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{j \in G_w} \sum_{t,s=1}^{T} (\boldsymbol{\lambda}'_i \boldsymbol{\Gamma}_{d_{ts}} \boldsymbol{\lambda}_j + \mathbf{r}'_{i.} \boldsymbol{\gamma}_{n,d_{ts}} \mathbf{r}_{j.}), \qquad (36)$$

and

$$\hat{v}^{lw}_{3,nT} = \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{j \in G_w} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \Delta \tilde{L}_{it} \Delta \tilde{L}_{js},$$

respectively. We will show that $\hat{v}^{lw}_{3,nT} - v^{lw}_{3,nT} = o_p(1)$ which gives the first result in the proposition. We have $\Delta \tilde{L}_{it} = \boldsymbol{\lambda}'_i \tilde{\mathbf{f}}_t + \mathbf{r}'_{i.} \tilde{\boldsymbol{\epsilon}}_{n,t}$ which gives

$$\Delta \tilde{L}_{it} \Delta \tilde{L}_{js} = \boldsymbol{\lambda}'_i \tilde{\mathbf{f}}_t \tilde{\mathbf{f}}'_s \boldsymbol{\lambda}_j + \boldsymbol{\lambda}'_i \tilde{\mathbf{f}}_t \tilde{\boldsymbol{\epsilon}}'_{n,s} \mathbf{r}_{j.} + \mathbf{r}'_{i.} \tilde{\boldsymbol{\epsilon}}_{n,t} \tilde{\mathbf{f}}'_s \boldsymbol{\lambda}_j + \mathbf{r}'_{i.} \tilde{\boldsymbol{\epsilon}}_{n,t} \tilde{\boldsymbol{\epsilon}}'_{n,s} \mathbf{r}_{j.}.$$

Using the last three equations, we obtain

$$\hat{v}^{lw}_{3,nT} - v^{lw}_{3,nT} = \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \boldsymbol{\lambda}'_i \left[\frac{1}{T} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \tilde{\mathbf{f}}_t \tilde{\mathbf{f}}'_s - \bar{\boldsymbol{\Gamma}}_T \right] \boldsymbol{\lambda}_j$$

$$+ \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \boldsymbol{\lambda}'_i \left[\frac{1}{T} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \tilde{\mathbf{f}}_t \tilde{\boldsymbol{\epsilon}}'_{n,s} \right] \mathbf{r}_{j.}$$

$$+ \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \mathbf{r}'_{i.} \left[\frac{1}{T} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \tilde{\boldsymbol{\epsilon}}_{n,t} \tilde{\mathbf{f}}'_s \right] \boldsymbol{\lambda}_j$$

$$+ \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \mathbf{r}'_{i.} \left[\frac{1}{T} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \right.$$

$$\times \tilde{\boldsymbol{\epsilon}}_{n,t} \tilde{\boldsymbol{\epsilon}}'_{n,t} - \bar{\boldsymbol{\gamma}}_{nT} \Big] \mathbf{r}_{j.}$$

$$= D_1 + D_2 + D_3 + D_4.$$

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

For the second result, we write $\underline{\widehat{\Omega}}_{3,nT} = (\mathbf{D}_n/n)^{-1} \underline{\widehat{\mathbf{V}}}_{3,nT} (\mathbf{D}_n/n)^{-1}$, where

$$\underline{\widehat{\mathbf{V}}}_{3,nT} = \frac{1}{n^2 T} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \iota_{g_i} \iota_{g_j}' \widehat{\lambda_i \mathbf{f}_t} \widehat{\lambda_j \mathbf{f}_s}$$

$$\times + \frac{1}{n^2 T} \sum_{i=1}^{n} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \iota_{g_i} \iota_{g_i}' \widehat{\varepsilon}_{it} \widehat{\varepsilon}_{is}.$$

The lwth element of this matrix is

$$\underline{\hat{v}}_{3,nT}^{lw} = \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{j \in G_w} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \widehat{\lambda_i \mathbf{f}_t} \widehat{\lambda_j \mathbf{f}_s}$$

$$+ \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{j \in G_w} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \widehat{\varepsilon}_{it} \widehat{\varepsilon}_{is}.$$

Using this expression together with (36), we find

$$\hat{v}_{3,nT}^{lw} - v_{3,nT}^{lw}$$

$$= \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \left[\frac{1}{T} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \widehat{\lambda_i \mathbf{f}_t} \widehat{\lambda_j \mathbf{f}_s} - \lambda_i' \bar{\Gamma}_T \lambda_j \right]$$

$$+ \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \widehat{\varepsilon}_{it} \widehat{\varepsilon}_{is} - \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \mathbf{r}_{i.}' \bar{\gamma}_{nT} \mathbf{r}_{j.}$$

$$= \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \left[\frac{1}{T} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \left(\widehat{\lambda_i \mathbf{f}_t} \widehat{\lambda_j \mathbf{f}_s} - \lambda_i' \mathbf{f}_t \mathbf{f}_s' \lambda_j \right) \right]$$

$$+ \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{j \in G_w} \lambda_i' \left[\frac{1}{T} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \mathbf{f}_t \mathbf{f}_s' - \bar{\Gamma}_T \right] \lambda_j$$

$$+ \frac{1}{n^2 T} \sum_{i \in G_l} \sum_{t,s=1}^{T} k_T \left(\frac{d_{ts}}{d_T} \right) \widehat{\varepsilon}_{it} \widehat{\varepsilon}_{is} - \frac{1}{n^2} \sum_{i \in G_l} \sum_{j \in G_w} \mathbf{r}_{i.}' \bar{\gamma}_{nT} \mathbf{r}_{j.}$$

The desired result now follows from Lemma 3(i), Lemma 3(ii), Lemma 2(i) and by noting that the last term is $O(n^{-1})$. This completes the proof.

Appendix C. Details on the evaluation of the IMF consumer price inflation forecasts

In this appendix, we present the evaluation of the IMF CPI forecasts summarized in Section 6.2. As in the application of the comparison of the economic growth forecasts of the OECD and the IMF, 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 five countries from the original IMF data set as their loss differentials are very different from the rest of the sample. These countries are Brazil, the Democratic Republic of the Congo, Peru, Venezuela, and Nicaragua. Notice that all these countries experienced hyperinflation in the late 2010s in the case of Venezuela and in the early to mid-90s for the rest. For the first four countries, there are huge drops or jumps in CPI inflation; hence their RW forecasts are very poor for at least one year. Whereas for the last country, the situation is the contrary; 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, 22 OECD countries that are not part of G7, and 98 non-OECD countries. As stated in the paper, in addition to the quadratic loss function, we use the absolute loss function in this application.

Cross-sectional and temporal dependence in loss differentials: CPI forecasts. Before looking into the EPA test results, we apply the methodology described in Section 5 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 that 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}, indicating six 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. 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 the overall EPA tests and continue with the clustered EPA tests. As stated in the main paper, we consider two different country clustering schemes: the first divides the sample of countries into OECD and non-OECD countries, whereas the second consists of G7 countries, non-G7 OECD countries, and non-OECD countries. We further split the sample into pre- and post-global financial crisis periods and computed the average loss differentials. The results for the average loss differentials are given in Table C.13. 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 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 between the IMF and the RW model were more pronounced. 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) of Table C.14, we see that all three overall EPA test statistics are statistically significant at the 10% level for the absolute loss function. We also see here the effect of taking CD into account: with $S_{n,T}^{(1)}$, we can reject the overall EPA hypothesis at 5% level, but this is

Table C.13

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 C.14

Panel tests of EPA for the CPI inflation forecasts (IMF *vs.* random walk).

	Overall EPA tests			Clustered EPA tests			
				Cluster 1: OECD Cluster 2: Non-OECD		Cluster 1: G7 Cluster 2: Non-G7 OECD Cluster 3: Non-OECD	
Panel (a): Full sample							
Test	Absolute loss	Quadratic loss	Test	Absolute loss	Quadratic loss	Absolute loss	Quadratic Loss
$S_{n,T}^{(1)}$	2.04 (0.04)	0.42 (0.67)	$C_{n,T}^{(1)}$	8.16 (0.02)	7.02 (0.03)	8.64 (0.03)	10.14 (0.02)
$S_{n,T}^{(3)}$	1.70 (0.09)	0.50 (0.62)	$C_{n,T}^{(3)}$	3.23 (0.20)	3.65 (0.16)	3.62 (0.31)	4.49 (0.21)
$\underline{S}_{n,T}^{(3)}$	1.78 (0.08)	0.50 (0.62)	$\underline{C}_{n,T}^{(3)}$	3.44 (0.18)	3.65 (0.16)	3.62 (0.31)	5.05 (0.17)
Panel (b): 1991–2006 (Pre-crisis)							
$S_{n,T}^{(1)}$	2.15 (0.03)	0.43 (0.67)	$C_{n,T}^{(1)}$	9.08 (0.01)	7.29 (0.03)	9.78 (0.02)	7.53 (0.06)
$S_{n,T}^{(3)}$	2.02 (0.04)	0.51 (0.61)	$C_{n,T}^{(3)}$	4.40 (0.11)	4.09 (0.13)	7.22 (0.07)	4.26 (0.23)
$\underline{S}_{n,T}^{(3)}$	1.99 (0.05)	0.51 (0.61)	$\underline{C}_{n,T}^{(3)}$	4.28 (0.12)	4.08 (0.13)	8.20 (0.04)	4.23 (0.24)
Panel (c): 2009–2019 (Post-crisis)							
$S_{n,T}^{(1)}$	−1.59 (0.11)	−0.82 (0.41)	$C_{n,T}^{(1)}$	2.66 (0.26)	2.25 (0.33)	8.29 (0.04)	6.31 (0.10)
$S_{n,T}^{(3)}$	−0.45 (0.66)	−0.40 (0.69)	$C_{n,T}^{(3)}$	0.30 (0.86)	0.23 (0.89)	9.87 (0.02)	4.97 (0.17)
$\underline{S}_{n,T}^{(3)}$	−0.45 (0.65)	−0.40 (0.69)	$\underline{C}_{n,T}^{(3)}$	0.41 (0.82)	0.24 (0.89)	8.27 (0.04)	6.34 (0.10)

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

not the case for the tests $S_{n,T}^{(3)}$ and $\underline{S}_{n,T}^{(3)}$. We remind that the average absolute loss differential of IMF and RW forecasts is 0.45, which is reported in the last row of Table C.13. We conclude that, overall, there is a small but statistically significant difference between the average absolute loss differential of the IMF and RW forecasts in favor of the latter.

With the quadratic loss function, we cannot reject the overall EPA hypothesis at conventional levels using any test. Therefore, the question is whether we can reject the clustered EPA hypothesis using the clusters under consideration. 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 clustered EPA hypothesis with both loss functions using the $C_{n,T}^{(1)}$. However, when we consider the CD in the loss differentials, the magnitudes of the test statistics decline dramatically and are insignificant. In the last block of the table, we have the results for the case of the 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 $\underline{C}_{n,T}^{(3)}$.

In Panel (b), the results for the pre-crisis period are reported. The results obtained from the overall tests are similar to those of the full sample; we can reject the overall EPA hypothesis with the absolute loss function, but this is not the case with the quadratic loss. 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. However, when we consider the three country clusters, we can reject the clustered EPA hypothesis with the absolute loss function using any test, at least at the 10% level. To conclude, using the absolute loss function, a significant difference between the IMF and RW forecast accuracy exists in the pre-crisis period.

As 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 clustered

EPA tests with two clusters, the differences are again statistically insignificant for both loss functions. Considering the case of three country clusters, similar to the pre-crisis period, we can reject the clustered EPA hypothesis with the absolute loss function using any test. The statistics for the 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 the quadratic loss, as is reported in Table C.13.

References

Akgun, O., Pirotte, A., Urga, G., & Yang, Z. (2020). Equal predictive ability tests for panel data with an application to OECD and IMF forecasts. arXiv preprint arXiv:2003.02803.

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

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

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

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

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

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

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

Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Dreher, A., Marchesi, S., & Vreeland, J. R. (2008). The political economy of IMF forecasts. *Public Choice*, *137*(1–2), 145–171.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, *80*(4), 549–560.

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

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762.

Han, C., Phillips, P. C., & 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.

Hoga, Y., & Dimitriadis, T. (2022). On testing equal conditional predictive ability under measurement error. *Journal of Business & Economic Statistics*, 1–13.

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

Keane, M. P., & Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. *The American Economic Review*, *80*, 714–735.

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

Kim, M. S., & 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., & Preve, D. (2012). Statistical tests for multiple forecast comparison. *Journal of Econometrics*, *169*(1), 123–130.

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

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

Pain, N., Lewis, C., Dang, T.-T., Jin, Y., & Richardson, P. (2014). OECD forecasts during and after the financial crisis. In *OECD Economics Department Working Papers*, *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., & Yamagata, T. (2008). A bias-adjusted LM test of error cross-section independence. *The Econometrics Journal*, *11*(1), 105–127.

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

Phillips, P. C., & Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, *20*, 971–1001.

Qu, R., Timmermann, A., & Zhu, Y. (2022). Comparing forecasting performance with panel data. Unpublished manuscript.

Stock, J. H., & 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.