Comparing Forecasting Performance in Cross-sections

Ritong Qu* Allan Timmermann† Yinchu Zhu‡

October 13, 2021

Abstract

This paper develops new methods for pairwise comparisons of predictive accuracy with cross-sectional data. Using a common factor setup, we establish conditions on cross-sectional dependencies in forecast errors which allow us to test the null of equal predictive accuracy on a single cross-section of forecasts. We consider both unconditional tests of equal predictive accuracy as well as tests that condition on the realization of common factors and show how to decompose forecast errors into exposures to common factors and idiosyncratic components. An empirical application compares the predictive accuracy of financial analysts' short-term earnings forecasts across six brokerage firms.

Key words: Economic forecasting; Competing Models, Predictive Accuracy, Cross-sectional Data; Analysts' Earnings Forecasts

*Rady School of Management, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093, U.S.A.; ritong.qu@rady.ucsd.edu

†Rady School of Management, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093, U.S.A.; atimmermann@ucsd.edu

‡Department of Economics, Brandeis University, 1415 South Street, Waltham, MA 02453, U.S.A.; yinchuzhu@brandeis.edu

1 Introduction

What, if anything, can we learn about forecasting performance from a *single* cross-section of data? This question is becoming highly relevant as large cross-sections of forecasts are now routinely recorded for numerous economic and financial outcomes: financial analysts predict company earnings and revenues for hundreds of firms covering multiple industries; credit card companies conduct billions of forecasts for real-time transactions to guard against fraud; banks and international organizations forecast macroeconomic outcomes across many countries and sectors.

Comparisons of forecasting performance conducted on a single cross-section has the potential for yielding important economic insights that easily get masked by averaging performance over longer spans of time. First, forecasting performance may be state- and time-dependent. A test conducted on a single cross-section might find that model-based forecasts are inferior to survey forecasts during, say, the Covid-19 epidemic although the two forecasts are equally accurate when their performance gets averaged over a longer sample. Such a finding could indicate that survey participants possessed important forward-looking information about the impact of this event that was not reflected in past data. Second, when conducted on individual time periods, cross-sectional tests can be used to identify points in time during which one forecast performs relatively well or to identify shifts over time in forecasting performance. Third, performance evaluations conducted on individual cross-sections facilitate faster real-time comparisons of predictive accuracy than conventional methods that require calculating often lengthy time-series averages which tends to slow down discovery of deterioration or breakdown in forecasting performance. Fourth, inference conducted on a single cross-section dispenses with time-series stationarity assumptions that are unlikely to be valid in many situations.

From an inferential perspective, the key challenge for cross-sectional comparisons of forecasting performance is the likely presence of common components in forecast errors. Such common components can invalidate the use of a cross-sectional central limit theorem (CLT) to derive distributional results for test statistics based on cross-sectional averages. To address this challenge, we develop a common factor framework for capturing cross-sectional dependencies in forecast errors and separately consider the cases with homogeneous and heterogeneous factor loadings. The case with homogeneous factor loadings gives rise to tests of equal unconditionally expected squared error loss, while heterogeneous factor loadings lead to tests that condition on factor realizations. Although these tests are fundamentally different we show that, in practice, they lead to very similar inference. Forecast comparisons conducted on individual cross-sections are robust to changes in both the number of factors and in the factor loadings which can be an important concern in empirical work, see Cheng et al. (2016).

Common components in the forecast errors contain important economic information about the underlying models used by forecasters and the extent to which shocks are fundamentally unpredictable. Large shocks to outcomes that were unanticipated by *all* forecasters and, thus, are common, cancel out from pairwise comparisons of squared forecast error *differences* to the extent that they affect individual forecasters by the same amount. Conversely, idiosyncratic error components that are specific to individual forecasters do not cancel out from squared error loss differentials.

To get a better sense of the commonality and predictability of economic shocks, we propose a new decomposition of the squared forecast error differential into a squared bias component, which tracks differences in forecast exposures to common factors, and an idiosyncratic error variance component. Only the total squared forecast error differential is observed, so we develop three approaches to estimate the common factors in forecast errors, namely (i) a cluster method that imposes homogeneity restrictions on factor loadings within clusters of variables and can be computed on a single cross-section; (ii) a common correlated effects estimator based on Pesaran (2006); and (iii) a principal components approach. Unlike the cluster approach, the second and third approach require the availability of time-series data to estimate factor loadings. Moreover, these approaches work under different assumptions about the number of factors and patterns in factor loadings and cover many of the situations encountered by applied researchers.

We illustrate our new tests in an empirical application to financial analysts' short-term forecasts of individual firms' quarterly earnings. We compare the predictive accuracy across six brokerages covering a total of between 1,400 and 1,800 different firms during a sample that spans twenty years. We find evidence of highly significant correlation across brokerage firms' earnings forecast errors, most of which can be captured through their loadings on a single common factor. Empirically, we find that our cross-sectional tests of equal predictive accuracy across brokerage firms are highly robust regardless of whether factor loadings are assumed to be homogeneous or heterogeneous and so yield similar results for the conditional and unconditional cases. For the vast majority of quarters, brokerage firms produce similarly accurate earnings forecasts, but we also identify some quarters with rejections of the null of equal predictive accuracy.

Using our decompositions we find that, in general, differences in idiosyncratic error variances account for more of the variation in squared error loss differences in brokerage firms' earnings forecasts than the squared bias. Differences in the accuracy of earnings forecasts in individual quarters thus appear to be mostly driven by differences in brokerage firms' ability to reduce uncertainty about the idiosyncratic earnings component and is less a reflection of differences in exposures to common factor shocks.

Our paper expands to a cross-sectional setting a large literature that compares the predictive accuracy of time-series forecasts. Chong and Hendry (1986) propose tests of forecast encompassing. More recently, Diebold and Mariano (1995) and West (1996) develop tests for comparing the null of equal predictive accuracy. Clark and McCracken (2001) and McCracken (2007) focus on comparisons of predictive accuracy for forecasts that are generated by nested models, while accounting for the effect of recursive updating in the parameter estimates used to generate forecasts. Giacomini and White (2006) propose a test of equal predictive accuracy that accounts for the presence of non-vanishing parameter estimation error and develop methods for conditional forecast comparisons. We build on these earlier contributions, but show how the presence of a cross-sectional dimension can enrich the set of economic hypotheses that can be tested and dispenses with the need for restrictive assumptions on time-series stationarity for the underlying data generating process.

A related literature evaluates the efficiency of forecasts with panel data; see, e.g., Keane and Runkle (1990), Davies and Lahiri (1995), and Patton and Timmermann (2012). However, this literature does not provide methods for comparing the relative accuracy of different forecasts or for conducting tests of the null of equal predictive accuracy across different forecasts. An advantage of our new tests is that they can be computed using only a single cross-section-provided that cross-sectional dependencies are properly accounted for. This makes the tests particularly useful in microeconomic forecast applications which often have short time-series dimensions since such surveys are conducted infrequently or due to the attrition of individual households that enter and exit.[1]

The outline of the paper is as follows. Section 2 presents our new tests for comparing predictive accuracy with individual cross-sections, while Section 3 develops our decomposition of the mean squared forecast errors into a squared bias and an idiosyncratic error variance component and derives statistics for testing the null that these two components are of the same magnitude across different forecasts. Section 4 conducts an empirical analysis that compares the predictive accuracy of firm-level short-term earnings forecasts across six brokerage firms. Section 5 uses Monte Carlo simulations to explore the finite-sample size and power properties of our tests in a variety of settings and Section 6 concludes. Technical proofs are in an Appendix.

[1] Giacomini et al. (2019) discuss micro forecasting approaches for annual PSID panels while Liu et al. (2018) and Liu et al. (2019) develop ways to forecast in panels with very short time-series dimensions.

2 Tests for Cross-sectional Comparisons of Predictive Accuracy

Formal tests used in comparisons of forecasting performance such as the well-known Diebold-Mariano test (Diebold and Mariano (1995)) rely on time-series averages. While these tests have proven useful in many economic applications, an important limitation of their usage is that sample sizes (T) are often short and so their statistical power can be quite low.[2] Conversely, in situations with long samples, non-stationarities in the underlying data generating process becomes an issue for inference. Moreover, new time-series observations arrive only slowly when outcomes are measured at a monthly, quarterly, or annual frequency, reducing the usefulness of real-time comparisons of predictive accuracy. These points highlight shortcomings of inference on predictive accuracy based on time-series averages.

In contrast, individual forecasting models can often be used to generate hundreds or even thousands of cross-sectional forecasts each period, as in the case of forecasts for individual customers, market places, product categories, or firms. Data with small T and large n can be used to compare the accuracy of pairs of forecasts in a particular time period or over a short period of time. Conducting such tests requires, however, an understanding of the assumptions under which it is possible to establish the distribution of cross-sectional averages underlying the test statistics. Most obviously, the loss differentials cannot be too strongly cross-sectionally dependent–otherwise a CLT will not apply to the cross-sectional test statistics.

We next develop a framework and a set of tests that allow us to conduct inference about relative predictive accuracy on single cross-sections.

[2] This is particularly relevant for microeconomic applications that often rely on short surveys, see, e.g., Giacomini et al. (2019) and Liu et al. (2019).

2.1 Setup

Let y_{it+h} denote the realized value of unit i at time $t+h$, where $i = 1, ..., n$ refers to the cross-sectional dimension and $t + h$ refers to the "target date", i.e., the point in time at which we observe the outcome. Further, suppose we observe the h-step-ahead forecast of y_{it+h} generated conditional on information available to the forecaster at time t. We denote these by $\hat{y}_{it+h|t,m}$, where $m = 1, ..., M$ indexes the individual forecasts (e.g., forecasting models) and $h \geq 0$ is the forecast horizon.

To compare the predictive accuracy of different forecasts we use a loss function that quantifies the cost of different forecast errors. Following Diebold and Mariano (1995), define the loss associated with forecast m as $L_{it+h|t,m} = L(y_{it+h}, \hat{y}_{it+h|t,m})$. Consistent with most empirical work, we assume that the loss is a quadratic function of the forecast error, $e_{it+h,m} = y_{it+h} - \hat{y}_{it+h|t,m}$, and thus takes the form[3]

$$L(y_{it+h}, \hat{y}_{it+h|t,m}) \equiv L_{it+h|t,m} = e^2_{it+h,m}. \tag{1}$$

Similarly, the squared-error loss differential between forecasts m_1 and m_2 for unit i at time $t + h$ is given by (dropping the reference to $m_1$and m_2)

$$\Delta L_{i,t+h|t} = e^2_{it+h,m_1} - e^2_{it+h,m_2}. \tag{2}$$

Following Diebold and Mariano (1995) and Giacomini and White (2006), we treat the forecasts as given and make high-level assumptions on the distribution of the forecast errors or, more generally, the losses $L_{it+h|t}$. Hence, we do not consider the effect of estimation error on the distribution of the test statistics which we derive.[4]

To keep our analysis simple, we focus on pair-wise comparisons of forecasting performance ($M = 2$). Often, empirical researchers have access to a large number of forecasts, e.g. from

[3]See Elliott et al. (2005) for a more general loss function that nests squared error loss.

[4]Estimation error and its effect on tests for equal predictive accuracy features prominently in the analysis of West (1996), Clark and McCracken (2001), McCracken (2007), and Hansen and Timmermann (2015).

surveys with large numbers of participants, from different forecasting models, or even from several cross-sections spanning different time periods. This introduces a multiple hypothesis testing problem when analyzing outcomes of several (pair-wise) test statistics. Dealing with this issue is beyond the scope of the present paper, but Qu et al. (2019) propose a Sup test procedure that allows for multiple comparisons while controlling the family-wise error rate.

2.2 Factor Structure

To capture cross-sectional dependencies in forecast errors, suppose we can decompose the forecast error of model m, $e_{i,t+h,m} = y_{i,t+h} - \hat{y}_{i,t+h|t,m}$, into a common component, f_{t+h}, with factor loadings λ_{im}, and an idiosyncratic component, $u_{it+h,m}$, so that, for $m = 1, 2$,

$$e_{i,t+h,m} = \lambda'_{im} f_{t+h} + u_{i,t+h,m}. \tag{3}$$

Under this setup, forecast errors are allowed to be affected by the same common factors, f_{t+h}, but we allow for differences in the factor loadings (λ_{im}) across units, i, and forecasts, m. Factor loadings, λ_{im}, can be either random or fixed as we make clear in the analysis below.

The assumed factor structure in (3) is typically well-motivated in economic forecast applications. Outcomes of economic variables such as GDP growth and inflation are likely to contain an important common unpredictable component reflecting large unanticipated supply shocks (e.g., commodity price shocks) or crises in financial markets. Common factors can be either global or regional in nature and are likely to have a very different impact on, e.g., advanced versus developing economies. The presence of common and idiosyncratic shocks is also consistent with macroeconomic models such as Mackowiak and Wiederholt (2009). A distinct advantage of the setup, which we demonstrate in our empirical analysis, is that the presence of common factors in the forecast errors is empirically testable through simple econometric tests.

We next consider how to conduct cross-sectional tests of equal predictive accuracy using the squared error loss function in (2) and the factor structure in (3).

2.3 Null Hypotheses

The assumed common factor structure in (3) introduces a common component that does not disappear asymptotically even as $n \to \infty$. To address this issue, we consider two different approaches for testing the null of equal predictive accuracy in a single cross-section.

First, we can test the unconditional null that the cross-sectional average loss differential at time $t+h$, $\overline{\Delta L}_{t+h} = n^{-1} \sum_{i=1}^{n} \Delta L_{i,t+h|t}$, equals zero in expectation:

$$H_{0,t+h}^{unc}: \; E(\overline{\Delta L}_{t+h}) = 0. \tag{4}$$

While the forecasts are only expected to be equally accurate at a single point in time, $t+h$, differences in predictive accuracy at that time are hypothesized to balance out across units, $i = 1, ..., n$. As we show below, this requires that the common factor component that introduces dependence in forecast errors cancels out in the loss differentials.

Second, we can test whether two forecasts are expected to be equally accurate, at time $t+h$, *conditional* on a particular outcome of the factor realizations, f_{t+h}, and factor loadings $\{\lambda_{i1}, \lambda_{i2}\}_{i=1}^{n}$ so that, for $\mathcal{F} = \sigma(f_{t+h}, \{\lambda_{i1}, \lambda_{i2}\}_{i=1}^{n})$,

$$H_{0,t+h}^{cond}: \; E(\overline{\Delta L}_{t+h} \mid \mathcal{F}) = 0. \tag{5}$$

This approach is valid provided that, conditional on the realized factor, a cross-sectional CLT applies to the idiosyncratic error components.

The conditional null in (5) is different from the unconditional null in (4) but is often of separate economic interest. For example, we can use (5) to test whether, conditional on the unusual realizations of the factors that occurred during the Global Financial Crisis, the accuracy of a set of alternative forecasts was the same. Or, as the complement to this, we can test whether the forecasts were equally accurate during more "normal" years.

If, in fact, factor realizations were the main driver of differences in the predictive accuracy of a pair of forecasts, we can imagine situations in which we reject the null in (4) without rejecting (5). Conversely, two forecasts could be equally accurate "on average" in a given period because one forecast is more strongly affected by shocks to the common factors and less affected by idiosyncratic error shocks, while the reverse holds for the other forecast and the effects balance out. In this case, we do not reject the null in (4), whereas the conditional null in (5) is rejected.

We next discuss settings under which the hypotheses in (5) and (4) hold along with how they can be tested.

2.4 Homogeneous Factor Loadings

Suppose loadings on the common factors affecting the individual forecast errors in (3) are the same across the two forecasts so $\lambda_{i1} = \lambda_{i2} = \lambda_i$. Under quadratic error loss,

$$\Delta L_{i,t+h|t} = (u_{i,t+h,1}^2 - u_{i,t+h,2}^2) + 2(u_{i,t+h,1} - u_{i,t+h,2})\lambda_i' f_{t+h}. \tag{6}$$

Common unpredictable shocks that are not picked up by any of the forecasts can be thought of as satisfying the assumption of homogeneous factor loadings since they can have a different effect on different units ($\lambda_{i1} \neq \lambda_{j1}$ for $i \neq j$), but will affect the forecasts in the same way($\lambda_{i1}' = \lambda_{i2}'$ for all i). These shocks will, therefore, cancel out from the forecast error differentials. For example, if the effects of a major event such as the Global Financial Crisis were unanticipated by both forecasts and affected them by the same amount, they cancel out from the loss differential.

Under homogeneous factor loadings, the cross-sectional dependence arising from the forecasts' exposure to the common factors, f_{t+h}, does not play an important role in deriving the asymptotics of tests of the null in (4) since $\lambda_i' f_{t+h}$ in (6) is multiplied by $(u_{i,t+h,1} - u_{i,t+h,2})$.

This is assured under the following assumption which requires (conditionally) independent idiosyncratic errors as well as a Lyapounov condition:

Assumption 1. *Suppose that the loadings are homogeneous,* $\lambda_{i1} = \lambda_{i2} = \lambda_i$ *for* $i = 1, ..., n$. *Conditional on* $\mathcal{F} = \sigma(f_{t+h}, \{\lambda_{i1}, \lambda_{i2}\}_{i=1}^n)$, $\{(u_{i,t+h,1}, u_{i,t+h,2})\}_{i=1}^n$ *is independent across* i *with mean zero and bounded* $(4+\delta)$ *moments for some* $\delta > 0$. *Moreover,* $\min_{1 \le i \le n} \mathrm{Var}[(u_{i,t+h,1} - u_{i,t+h,2}) \mid \mathcal{F}] \ge c$ *for some constant* $c > 0$ *and*

$$\frac{\left(\sum_{i=1}^n |\lambda_i' f_{t+h}|^{2+\delta}\right)^{1/(2+\delta)}}{\left(\sum_{i=1}^n |\lambda_i' f_{t+h}|^2\right)^{1/2}} = o_P(1).$$

To test the null of equal expected loss for the cross-sectional average in (4), consider the test statistic

$$Q_{t+h} = \frac{n^{1/2}\overline{\Delta L}_{t+h|t}}{\sqrt{n^{-1}\sum_{i=1}^n \left(\Delta L_{i,t+h|t}\right)^2}}. \tag{7}$$

Under the assumption of pair-wise homogeneous factor loadings, (6) shows that testing the null of equal predictive accuracy in period $t+h$ amounts to testing that $E(u_{i,t+h,1}^2 - u_{i,t+h,2}^2) = 0$. This is easily accomplished under Assumption 1 which ensures independence across i for $(u_{i,t+h,1}, u_{i,t+h,2})$ so that asymptotic normality can be established for Q_{t+h} in (7) as we next show:[5]

Theorem 1. *Suppose Assumption 1 holds. Then under the null of equal expected cross-sectional predictive accuracy,* $H_{0,t+h}^{unc} : E(\overline{\Delta L}_{t+h}) = 0$, *we have*

$$\limsup_{n\to\infty} P\left(|Q_{t+h}| > z_{1-\alpha/2}\right) \le \alpha,$$

where $z_{1-\alpha/2}$ *is the* $(1-\alpha/2)$ *quantile of a* $N(0,1)$ *variable.*

Theorem 1 shows that homogeneous factor loadings lead to a simple test of the null of equal expected loss for the pooled average using data only on a single cross-section. Moreover,

[5]Alternatively, we can test this null under assumptions of stationarity which allows us to exploit time-series variation in the factors.

the test statistic follows a Gaussian distribution in large cross-sections.

For now, we do not go into details of how the assumption of homogeneous loadings can be tested. However, as we show below, our approach for testing the null in (4) remains valid as long as $n^{-1}\sum_{i=1}^{n}[(\lambda'_{i,1}f_{t+h})^2 - (\lambda'_{i,2}f_{t+h})^2] = 0$. Moreover, this condition can be tested empirically and we propose ways to do so later on.

2.5 Heterogeneous Factor Loadings

Next, consider the case with heterogeneous factor loadings for the forecast errors, i.e., $\lambda_{i,1} \neq \lambda_{i,2}$. For this case, the loss differential in (6) is generalized to

$$\begin{aligned}\Delta L_{i,t+h|t} = & \left[(\lambda'_{i,1}f_{t+h})^2 - (\lambda'_{i,2}f_{t+h})^2\right] \\ & + \left[u^2_{i,t+h,1} - u^2_{i,t+h,2} + 2(\lambda'_{i,1}f_{t+h}u_{i,t+h,1} - \lambda'_{i,2}f_{t+h}u_{i,t+h,2})\right]. \end{aligned} \tag{8}$$

When the factor loadings differ for the forecasts, equation (8) shows that the relative predictive accuracy in period $t+h$ contains a systematic component, $E\left[(\lambda'_{i,1}f_{t+h})^2 - (\lambda'_{i,2}f_{t+h})^2\right]$. Even if f_{t+h} is independent of the factor loadings, $\{(\lambda_{i,1}, \lambda_{i,2})\}_{i=1}^{n}$, and these loadings are independent across i, $n^{-1/2}\sum_{i=1}^{n}\left[(\lambda'_{i,1}f_{t+h})^2 - (\lambda'_{i,2}f_{t+h})^2\right]$ is asymptotically normal only conditional on f_{t+h}. This suggests conducting a test of equal expected predictive accuracy conditional on the factor realization as is done in (5).

To test the conditional null in (5), let $\mathcal{F} = \sigma(f_{t+h}, \{\lambda_{i1}, \lambda_{i2}\}_{i=1}^{n})$ and assume that $E(u_{i,t+h,1} \mid \mathcal{F}) = E(u_{i,t+h,2} \mid \mathcal{F}) = 0$. Define

$$\xi_{i,t+h} = \left(u^2_{i,t+h,1} - u^2_{i,t+h,2}\right) - E\left(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F}\right) + 2(\lambda'_{i,1}f_{t+h}u_{i,t+h,1} - \lambda'_{i,2}f_{t+h}u_{i,+ht,2}).$$

Using equation (8), we have

$$\overline{\Delta L}_{t+h} - E(\overline{\Delta L}_{t+h} \mid \mathcal{F}) = n^{-1}\sum_{i=1}^{n}\xi_{i,t+h}. \tag{9}$$

The ideal variance estimate for the object in (9) is $n^{-1}\sum_{i=1}^{n}\xi_{i,t+h}^2$. However, at the unit level, we only observe $e_{i,t+h,m}$ and hence are restricted to computing $n^{-1}\sum_{i=1}^{n}(\Delta L_{i,t+h|t}-\overline{\Delta L}_{t+h})^2$. Consider the following test statistic

$$\tilde{Q}_{t+h}=\frac{n^{1/2}\overline{\Delta L}_{t+h}}{\sqrt{n^{-1}\sum_{i=1}^{n}(\Delta L_{i,t+h|t}-\overline{\Delta L}_{t+h})^2}}. \tag{10}$$

To establish properties of the test statistic in (10), we need a set of regularity conditions which we summarize in the following assumption:

Assumption 2. *Conditional on $\mathcal{F}=(f_{t+h},\{\lambda_{i1},\lambda_{i2}\}_{i=1}^{n})$, $\{(u_{i,t+h,1},u_{i,t+h,2})\}_{i=1}^{n}$ is independent across i with mean zero and bounded $(4+\delta)$ moments for some $\delta>0$. Moreover, $\min_{1\le i\le n}\mathrm{Var}[\xi_{i,t+h}\mid\mathcal{F}]\ge c$ for some constant $c>0$.*

Using this assumption, we can now test the null $E(\overline{\Delta L}_{t+h}\mid\mathcal{F})=0$ or, equivalently, establish a confidence interval for $E(\overline{\Delta L}_{t+h}\mid\mathcal{F})$:

Theorem 2. *Suppose Assumption 2 holds. Then, under the conditional null $H_{0,t+h}^{cond}$: $E(\overline{\Delta L}_{t+h}\mid\mathcal{F})=0$, the following result holds for the test statistic in (10)*

$$\limsup_{n\to\infty}P\left(|\tilde{Q}_{t+h}|>z_{1-\alpha/2}\right)\le\alpha,$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of a $N(0,1)$ variable.

Results based on the test statistic in (10) can be interpreted in two ways. First, as explained above, they can be viewed as tests of the conditional null $E(\overline{\Delta L}_{t+h}\mid\mathcal{F})=0$. Second, if we assume that the factor loadings $\{(\lambda_{i,1},\lambda_{i,2})\}_{i=1}^{n}$ are random, independent across i and independent of f_{t+h}, we can use the test statistic in (10) to test $E(\overline{\Delta L}_{t+h}\mid f_{t+h})=0$ without also conditioning on the factor loadings ($\lambda_{i,1}$ and $\lambda_{i,2}$). Testing the latter hypothesis introduces an additional term in the numerator of (10)

$$E(\overline{\Delta L}_{t+h}\mid\mathcal{F})-E(\overline{\Delta L}_{t+h|t}\mid f_{t+h})$$

$$= f'_{t+h}\left(n^{-1}\sum_{i=1}^{n}\left[\lambda_{i,1}\lambda'_{i,1} - \lambda_{i,2}\lambda'_{i,2} - E(\lambda_{i,1}\lambda'_{i,1} - \lambda_{i,2}\lambda'_{i,2})\right]\right) f_{t+h}.$$

However, the denominator in (10) still overestimates the variance of the numerator of the test statistic under the null. As a result, Theorem 2 remains valid for testing the null $E(\overline{\Delta L}_{t+h} \mid f_{t+h}) = 0$ and the critical values remain the same.

Under either interpretation, it follows from (8) that the variance estimate in (10) is conservative. Under the first interpretation, this follows because the variance estimate takes into account variation in the factor structure and in $E(u^2_{i,t+h,1} - u^2_{i,t+h,2})$. Under the second interpretation, the variance estimate still includes cross-sectional variations in $E(u^2_{i,t+h,1} - u^2_{i,t+h,2})$. This seems unavoidable without introducing additional modeling assumptions that impose structure on this variation.[6]

2.6 Other loss functions

In practice, applied researchers might consider loss functions other than the squared error loss in (1), including linex, absolute error or piece-wise linear loss, see, e.g., Elliott et al. (2005). Fortunately, the methodology in Section 2.5 can readily be extended to such loss functions.

To see this, suppose we replace Assumption 2 with the assumption that, conditional on $\mathcal{F}$, $\{\Delta L_{i,t+h|t}\}_{i=1}^{n}$ is independent, where $\Delta L_{i,t+h|t} = L(y_{it+h}, \hat{y}_{it+h|t,1}) - L(y_{it+h}, \hat{y}_{it+h|t,2})$ for a general loss function $L(\cdot,\cdot)$. Under the conditional null hypothesis $E(\overline{\Delta L}_{t+h} \mid \mathcal{F}) = 0$, $\overline{\Delta L}_{t+h} - E(\overline{\Delta L}_{t+h} \mid \mathcal{F})$ will be the average of terms that, conditional on $\mathcal{F}$, have mean zero and are independent. Therefore, with moment conditions similar to those in Assumption 2, Theorem 2 remains valid. In Section 5, we demonstrate this point using Monte Carlo simulations for the test of equal conditionally expected loss applied to the linex loss function. We find results that are very similar to those obtained under squared error loss.

[6]Essentially, we have a CLT for independent but non-identically distributed variables, $\Delta L_{i,t+h|t} - E\left[\Delta L_{i,t+h|t} | f_{t+h}\right]$, but the exact variance is difficult to estimate because $E\left[\Delta L_{i,t+h|t} | f_{t+h}\right]$ cannot be estimated from the observed data.

For the unconditional test, we can consider a linear factor structure as a series approximation. For example, suppose that $e_{i,t+h,m} = \lambda'_{im} f_{t+h} + u_{i,t+h,m}$ and $L_{i,t+h,m} = \phi(e_{i,t+h,m})$ for some function $\phi(\cdot)$. Provided that $\phi(\cdot)$ is smooth enough and $e_{i,t+h,m}$ is bounded, standard approximation results can be used to give a polynomial approximation $\phi(x) \approx \sum_{j=0}^{k} a_j x^j$, where k grows slowly with the sample size. Because $(\lambda'_{im} f_{t+h} + u_{i,t+h,m})^j$ contains powers of $\lambda'_{im} f_{t+h}$, polynomials of factors and factor loadings become new factors in an augmented linear factor structure. Clearly, the details of this approach (e.g., approximation rate and strong factor conditions) require serious theoretical analysis, which we leave for future research.

Cross-sectional comparisons of forecast errors sometimes involve variables that are measured in very different units. This can mean that the comparisons are dominated by a few variables, possibly impairing the finite-sample behavior of the test statistics. To address this point, one can use squared percentage errors which tend to be more comparable across variables. Alternatively, individual variables' forecast errors can be scaled by their standard errors prior to calculating the test statistics.

3 Decomposing Differences in Forecasting Performance

Equation (3) decomposes the forecast errors into a common factor component and an uncorrelated idiosyncratic error component. In some economic applications, it is important to be able to attribute differences in forecasting performance to these two sources. For example, Mackowiak and Wiederholt (2009) develop a rational inattention model in which firms acquire and process information subject to a constraint on their total attention budget. Consistent with the setup in (3), Mackowiak and Wiederholt (2009) partition firms' information set into signals about a common (aggregate) factor and an idiosyncratic term. The constraint on each forecaster's attention introduces a trade-off between reducing the uncertainty about the common factor versus reducing the variance of the idiosyncratic error. Similarly, the finance literature on performance of investment managers distinguishes between generalists who possess market timing skills that require an ability to predict pervasive (common) factors affecting a broad set of asset returns versus stock pickers with security selection skills which require specialist firm-level knowledge akin to more precise signals on the idiosyncratic error terms (see, e.g., Blake et al. (2013)).

The importance of these types of skills is likely to vary over time as a result of common factor volatility being higher during recessions or in financial crises (favoring market timers) and lower during expansions and calmer periods (favoring stock pickers), see, e.g., Kacperczyk et al. (2014). By conducting tests on individual cross-sections, our approach can help identify periods in which forecasters with a comparative advantage at predicting the common factors (generalists) perform relatively better than the forecasters who focus instead on the idiosyncratic error component (specialists).

Decomposing forecast errors into common factors and uncorrelated idiosyncratic terms is also important in applications of forecast combination since these terms matter for calculating optimal combination weights which depend on both the overall error variance and on the covariance between forecast errors. The larger the contribution to forecast errors from the common factors and the more homogeneous the factor loadings are, the closer the optimal combination weights will be to equal-weighting. Related to this, the scope for achieving gains in predictive accuracy from forecast combination is likely to be highest during times when the correlation in forecast errors is weakest, i.e., less driven by common factors with similar loadings and more by idiosyncratic errors.

We next discuss how to conduct inference on the squared conditional bias and idiosyncratic variance components.

3.1 Decomposing the Conditional Squared Error Loss

Using equation (8), we can express the (cross-sectional) average conditional squared error loss difference as the sum of the average difference in squared conditional bias and the average difference in the conditional idiosyncratic error variance:

$$\underbrace{n^{-1}\sum_{i=1}^{n} E\left(\Delta L_{i,t+h|t} \mid \mathcal{F}\right)}_{E(\overline{\Delta L}_{t+h} \mid \mathcal{F})} = \underbrace{n^{-1}\sum_{i=1}^{n}\left[(\lambda_{i,1}' f_{t+h})^2 - (\lambda_{i,2}' f_{t+h})^2\right]}_{bias^2_{t+h}} + \underbrace{n^{-1}\sum_{i=1}^{n} E\left(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F}\right)}_{E\left(\Delta u^2_{t+h} \mid \mathcal{F}\right)}. \tag{11}$$

The terms on the right hand side of the decomposition in (11) are unobserved. However, note that

$$\overline{\Delta L}_{t+h} - bias^2_{t+h} = \overline{\Delta u}^2_{t+h} + \frac{2}{n}\sum_{i=1}^{n}\left[\lambda_{i,1}' f_{t+h} u_{i,t+h,1} - \lambda_{i,2}' f_{t+h} u_{i,t+h,2}\right], \tag{12}$$

where $\overline{\Delta u}^2_{t+h} = n^{-1}\sum_{i=1}^{n}\left(u^2_{i,t+h,1} - u^2_{i,t+h,2}\right)$. Provided that n is relatively large so the last term on the right side of (12) is small, the bias-adjusted average loss differential on the left hand side of (12) can be expected to be a good estimate of the difference in the two forecasts' idiosyncratic variance at time t, $E\left(\Delta u^2_{t+h} \mid \mathcal{F}\right)$.[7]

We next discuss three strategies for computing $(\lambda_{i,1}' f_{t+h})^2 - (\lambda_{i,2}' f_{t+h})^2$. The first exploits clusters in factor loadings and so is applicable when factor loadings are homogeneous within certain groups of units. This approach can be computed on a single cross-section and poses no limit on the number of factors affecting the forecast errors but requires that clusters can be identified within which there is little or no heterogeneity in the factor loadings. The second approach uses the common correlated effects (CCE) method of Pesaran (2006) and so requires the availability of panel data to estimate factor loadings from time series data. This approach does not impose tight restrictions on factor loadings but, in practice, limits the number of common factors driving the forecast errors. The third approach, principal components (PCA), again requires the availability of panel data and is similar to the CCE approach. However, it does not impose tight bounds on the number of common factors in the forecast error differentials.

[7] Of course, we do not directly observe the idiosyncratic errors and factors. However, since $\overline{\Delta L}_{t+h}$ is observed, from (12) we only need to estimate the factor-induced squared bias term, $bias^2_{t+h}$.

3.2 Clustering in Factor Loadings

It is common in empirical applications to have data on units that share certain observable characteristics or features which make them more similar than randomly selected units. For example, advanced economies may react in a broadly similar way to supply shocks which, in turn, affect emerging or developing economies very differently. Or, the effect of an interest rate increase on the default probability of credit card holders may be quite different across high, medium, and low income households, yet be broadly similar within these three categories.

In this section we develop a class of estimators using the identifying assumption that clusters of cross-sectional units share the same factor loadings, while allowing factor loadings to differ across clusters. Formally, suppose that a set of K clusters $\bigcup_{k=1}^{K} H_k = \{1, ..., n\}$ form a partition of all n units so that each unit belongs to a unique cluster, H_k, i.e., $H_j \cap H_l = \emptyset$ with $n_k = |H_k|$ elements in the kth cluster. We assume that the cluster membership for each unit is known ex ante and so is not determined endogenously from the data. Moreover, suppose that the factor loadings $(\lambda_{i,1}, \lambda_{i,2})$ can differ across clusters $(\lambda_{i,1}, \lambda_{i,2}) \neq (\lambda_{j1}, \lambda_{j2})$ for $i \in H_k$ and $j \in H_l$, but are homogeneous within clusters

$$(\lambda_{i,1}, \lambda_{i,2}) = (\lambda_{1,(k)}, \lambda_{2,(k)}) \text{ for all } i \in H_k. \tag{13}$$

3.2.1 Testing Equal Idiosyncratic Error Variances

We first discuss how to test the conditional null of equal average idiosyncratic error variance for the two forecasts given $\mathcal{F}$ for all units in cluster k:

$$H_0^{idio}: \; n_k^{-1} \sum_{i \in H_k} E(u_{i,t+h,1}^2 - u_{i,t+h,2}^2 \mid \mathcal{F}) = 0. \tag{14}$$

To test this null, we need to construct an estimate of the idiosyncratic variance within each cluster. To see how group patterns in factor loadings allow us to identify the idiosyncratic variance component, $\overline{\Delta u}^2_{t+h}$, define the errors from the two forecasts, averaged within each cluster, as

$$\bar{e}_{1,k,t+h} \equiv n_k^{-1} \sum_{i \in H_k} \left(y_{i,t+h} - \hat{y}_{i,t+h|t,1} \right) = \lambda'_{1,(k)} f_{t+h} + n_k^{-1} \sum_{i \in H_k} u_{i,t+h,1},$$

and

$$\bar{e}_{2,k,t+h} \equiv n_k^{-1} \sum_{i \in H_k} \left(y_{i,t+h} - \hat{y}_{i,t+h|t,2} \right) = \lambda'_{2,(k)} f_{t+h} + n_k^{-1} \sum_{i \in H_k} u_{i,t+h,2}.$$

Squaring these within-cluster average forecast errors, we have

$$\bar{e}^2_{1,k,t+h} - \bar{e}^2_{2,k,t+h} = (\lambda'_{1,(k)} f_{t+h})^2 - (\lambda'_{2,(k)} f_{t+h})^2 + \left(n_k^{-1} \sum_{i \in H_k} u_{i,t+h,1} \right)^2 - \left(n_k^{-1} \sum_{i \in H_k} u_{i,t+h,2} \right)^2 + 2\lambda'_{1,(k)} f_{t+h} n_k^{-1} \sum_{i \in H_k} u_{i,t+h,1} - 2\lambda'_{2,(k)} f_{t+h} n_k^{-1} \sum_{i \in H_k} u_{i,t+h,2}. \quad (15)$$

Define $\overline{\Delta L}_{t+h,k} \equiv n_k^{-1} \sum_{i \in H_k} \Delta L_{i,t+h}$ and let $\overline{\Delta u}^2_{t+h,k}$ be the average loss differential for cluster k adjusted for the difference $(\bar{e}^2_{1,k,t+h} - \bar{e}^2_{1,k,t+h})$:

$$\overline{\Delta u}^2_{t+h,k} = \overline{\Delta L}_{t+h,k} - (\bar{e}^2_{1,k,t+h} - \bar{e}^2_{1,k,t+h}). \quad (16)$$

This suggests using the following statistic to test H_0^{idio} in (14):

$$S_k = \frac{\sqrt{n_k} \overline{\Delta u}^2_{t+h,k}}{\sqrt{n_k^{-1} \sum_{i \in H_k} (\Delta L_{i,t+h} - \overline{\Delta L}_{t+h,k})^2}}. \quad (17)$$

Theorem 3. *Suppose Assumption 2 holds. Then under the null hypothesis* H_0^{idio} : $n_k^{-1} \sum_{i \in H_k} E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F}) = 0$, *we have*

$$\limsup_{n_k \to \infty} P\left(|S_k| > z_{1-\alpha/2} \right) \leq \alpha,$$

where $z_{1-\alpha/2}$ *is the* $(1 - \alpha/2)$ *quantile of a* $N(0, 1)$ *variable.*

Alternatively, we can test the weaker null of equal expected squared idiosyncratic forecast errors holding on average, i.e., across all units though not necessarily within each cluster:

$$H_0^{idio-av}:\ n^{-1}\sum_{i=1}^{n} E(u_{i,t+h,1}^2 - u_{i,t+h,2}^2 \mid \mathcal{F}) = 0. \tag{18}$$

To this end, let $\overline{\Delta u}_{t+h}^2 = \sum_{k=1}^{K}\frac{n_k}{n}\overline{\Delta u}_{t+h,k}^2$ be the cluster-weighted average difference in squared idiosyncratic forecast errors, and consider the test statistic

$$S_c = \frac{\sqrt{n}\overline{\Delta u}_{t+h}^2}{\sqrt{n^{-1}\sum_{k=1}^{K}\sum_{i\in H_k}(\Delta L_{i,t+h} - \overline{\Delta L}_{t+h,k})^2}}. \tag{19}$$

We use S_c to test the null in (18) of equal average idiosyncratic forecast error variance:

Corollary 1. *Suppose Assumption 2 holds and assume that $\lim_{n\to\infty} n_k/n > 0$ for all $1 \leq k \leq K$. Then under the null $H_0^{idio-av}:\ n^{-1}\sum_{i=1}^{n} E(u_{i,t+h,1}^2 - u_{i,t+h,2}^2 \mid \mathcal{F}) = 0,$ we have*

$$\limsup_{n\to\infty} P\left(|S_c| > z_{1-\alpha/2}\right) \leq \alpha,$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of a $N(0,1)$ variable.

Using Corollary 1, we can compute a $1-\alpha$ confidence interval for the squared idiosyncratic forecast errors $\overline{\Delta u}_{t+h}^2$ as

$$\overline{\Delta u}_{t+h}^2 \pm \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{n^{-1}\sum_{k=1}^{K}\sum_{i\in H_k}(\Delta L_{i,t+h} - \overline{\Delta L}_{t+h,k})^2}. \tag{20}$$

3.2.2 Testing Equal Squared Biases

Next, consider the squared bias component of the expected loss differential in (11). Under the assumed homogeneous factor loadings within clusters in (13), we have

$$n^{-1}\sum_{i=1}^{n}\left[(\lambda_{i,1}' f_{t+h})^2 - (\lambda_{i,2}' f_{t+h})^2\right] = \sum_{k=1}^{K}\frac{n_k}{n}\left((\lambda_{1,(k)}' f_{t+h})^2 - (\lambda_{2,(k)}' f_{t+h})^2\right).$$

We can estimate $(\lambda'_{1,(k)} f_{t+h})^2 - (\lambda'_{2,(k)} f_{t+h})^2$ by $\bar{e}^2_{1,k,t+h} - \bar{e}^2_{2,k,t+h}$. By (15), we have

$$\begin{aligned}\bar{e}^2_{1,k,t+h} - \bar{e}^2_{2,k,t+h} &= (\lambda'_{1,(k)} f_{t+h})^2 - (\lambda'_{2,(k)} f_{t+h})^2 \\ &\quad + 2\lambda'_{1,(k)} f_{t+h} n_k^{-1} \sum_{i \in H_k} u_{i,t+h,1} - 2\lambda'_{2,(k)} f_{t+h} n_k^{-1} \sum_{i \in H_k} u_{i,t+h,2} + O_P(n_k^{-1}).\end{aligned}$$

To test the null of equal squared bias, we use the following test statistic:

$$B_{n,1} = \frac{\sqrt{n} \sum_{k=1}^K \frac{n_k}{n} (\bar{e}^2_{1,k,t+h} - \bar{e}^2_{2,k,t+h})}{2\sqrt{n^{-1} \sum_{k=1}^K \sum_{i \in H_k} (\bar{e}_{1,k,t+h} \hat{u}_{i,t+h,1} - \bar{e}_{2,k,t+h} \hat{u}_{i,t+h,2})^2}}, \tag{21}$$

where $\hat{u}_{i,t+h,1} = y_{i,t+h} - \hat{y}_{i,t+h|t,m_1} - \bar{e}_{1,k,t+h}$ and $\hat{u}_{i,t+h,2} = y_{i,t+h} - \hat{y}_{i,t+h|t,m_2} - \bar{e}_{2,k,t+h}$. We can show that $B_{n,1}\left(n^{-1} \sum_{i=1}^n \left[(\lambda'_{i,1} f_{t+h})^2 - (\lambda'_{i,2} f_{t+h})^2\right]\right) \to^d N(0,1)$, and so:

Theorem 4. *Suppose Assumptions 1 holds and assume that* $\lim_{n\to\infty} n_k/n > 0$ *for all* $1 \leq k \leq K$. *Then under* $H_0:\ n^{-1} \sum_{i=1}^n \left[(\lambda'_{i,1} f_{t+h})^2 - (\lambda'_{i,2} f_{t+h})^2\right] = 0$, *we have*

$$\limsup_{n\to\infty} P\left(|B_{n,1}| > z_{1-\alpha/2}\right) \leq \alpha.$$

The null of equal squared bias relates to our earlier discussion of homogeneous versus heterogeneous factor loadings: If factor loadings are the same across two sets of forecasts, their squared bias differential should also be close to zero.

Theorem 4 yields a $1-\alpha$ confidence interval for $n^{-1} \sum_{i=1}^n \left[(\lambda'_{i,1} f_{t+h})^2 - (\lambda'_{i,2} f_{t+h})^2\right]$:

$$\sum_{k=1}^K \frac{n_k}{n} (\bar{e}^2_{1,k,t+h} - \bar{e}^2_{2,k,t+h}) \pm \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{n^{-1} \sum_{k=1}^K \sum_{i \in H_k} (\bar{e}_{1,k,t+h} \hat{u}_{i,t+h,1} - \bar{e}_{2,k,t+h} \hat{u}_{i,t+h,2})^2}. \tag{22}$$

Note that because $B_{n,1}\left(n^{-1} \sum_{i=1}^n \left[(\lambda'_{i,1} f_{t+h})^2 - (\lambda'_{i,2} f_{t+h})^2\right]\right) \to^d N(0,1)$, the confidence interval is asymptotically exact.

3.3 Factor Structure Estimated by CCE

In many empirical applications, a cluster structure may not be suitable either because units are not easily assigned to individual clusters or because factor loadings are not homogeneous within clusters. For such applications, a more traditional factor setting may be more appropriate. To this end, suppose we observe a panel of forecast errors $\{e_{i,s+h,m}\}_{1\leq i\leq n,\ 1\leq s\leq T}$ generated according to the factor model in (3), $e_{i,s+h,m} = \lambda_{i,m}' f_{s+h} + u_{i,s+h,m}$, where $m = 1, 2$, $\lambda_{i,m} \in \mathbb{R}^{r\times v}$ and $f_{s+h} \in \mathbb{R}^r$ with $v \geq r$, so the number of observables, v, is at least equal to the number of factors, r. The requirement that $v \geq r$ implies that if we do not include observables other than the two sets of forecast errors, we can allow for at most two factors. Conversely, including more observable variables that are driven by the same factors lets us relax this restriction and allow for additional factors.

3.3.1 Difference in Idiosyncratic Error Variances

Let $e_{i,s+h} = (e_{i,s+h,1}, e_{i,s+h,2})' \in \mathbb{R}^2$ and $u_{i,s+h} = (u_{i,s+h,1}, u_{i,s+h,2})' \in \mathbb{R}^2$ be 2×1 vectors of forecast errors and idiosyncratic residuals and define the cross-sectional averages $\bar{e}_{s+h} = n^{-1}\sum_{i=1}^n e_{i,s+h}$, $\bar{u}_{s+h} = n^{-1}\sum_{i=1}^n u_{i,s+h}$ and $\bar{\lambda} = n^{-1}\sum_{i=1}^n \lambda_i$ with $\lambda_i = (\lambda_{i,1}, \lambda_{i,2}) \in \mathbb{R}^{r\times 2}$. Assuming that we can invoke a CLT for the cross-sectional average of the idiosyncratic shocks, $\bar{u}_{s+h}$ will be small and $\bar{e}_{s+h} \approx \bar{\lambda}' f_{s+h}$ can be used as a proxy for the unobserved factors. This is the common correlated effects (CCE) idea proposed in Pesaran (2006). In turn, we can estimate the individual factor loadings, λ_{im}, from a time-series regression

$$\hat{\lambda}_i' = \left(\sum_{s=1}^T e_{i,s+h}\bar{e}_{s+h}'\right)\left(\sum_{s=1}^T \bar{e}_{s+h}\bar{e}_{s+h}'\right)^{-1}.$$

Let $\lambda_{i,1}$ denote the first column of λ_i, with similar notations used for $\hat{\lambda}_{i,1}$ and $\hat{\lambda}_{i,2}$. Consider the following regularity conditions:

Assumption 3. *The following conditions hold for $m = 1, 2$:*

(1) the smallest eigenvalue of $\bar{\lambda}\bar{\lambda}'$ is bounded away from zero.

(2) conditional on $\{f_{s+h}\}_{s+h=1}^{T}$ *and* $\{\lambda_i\}_{i=1}^{n}$, $\{u_{i,t+h,m}\}_{i=1}^{n}$ *has mean zero with bounded variance and is independent across* i.

The first part of Assumption 3 implies that the number of factors cannot exceed the dimension of $e_{i,s+h}$ since otherwise the smallest eigenvalue of $\bar{\lambda}\bar{\lambda}'$ is zero. We also impose additional regularity conditions. These are part of Assumptions A, B and C in Bai (2003) and are routinely imposed in factor analysis.

Assumption 4. *The following conditions hold for* $m = 1, 2$*:*

(1) $n^{-1}\sum_{i=1}^{n}\lambda_{i,m}\lambda_{i,m}'$ *and* $Ef_{s+h}f_{s+h}'$ *have eigenvalues bounded away from zero and infinity.*

(2) $\sum_{s+h=1}^{T}\sum_{i=1}^{n}\lambda_{i,m}u_{i,s+h,m}f_{s+h}' = O_P(\sqrt{nT})$.

(3) There exists a constant $M > 0$ *such that* $\|\gamma_n(s,\tau)\| \leq M$ *and* $T^{-1}\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\|\gamma_n(s,\tau)\| \leq M$, *where* $\gamma_n(s,\tau) = n^{-1}\sum_{i=1}^{n}Eu_{i,s+h}u_{i,\tau+h}'$.

(4) $n/T^2 = o(1)$.

Using Assumption 3 and 4, we can characterize the difference between the average squared forecast errors and the average squared factor values, both weighted by the factor loadings, λ_i' :

Lemma 1. *Under Assumptions 3 and 4, we have*

$$n^{-1/2}\sum_{i=1}^{n}[(\hat{\lambda}_{i,1}'\bar{e}_{t+h})^2 - (\lambda_{i,1}'f_{t+h})^2] = 2n^{-1/2}\bar{u}_{t+h}'\bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\left(\sum_{i=1}^{n}\lambda_{i,1}\lambda_{i,1}'\right)f_{t+h} + o_P(1).$$

Next, consider the null that the difference in the squared idiosyncratic variance component of the forecast errors equals zero:

$$H_0: \; n^{-1}\sum_{i=1}^{n}E(u_{i,t+h,1}^2 - u_{i,t+h,2}^2 \mid \mathcal{F}) = 0. \tag{23}$$

To test this null, we use the following test statistic

$$S_{cce} = \frac{\sqrt{n}\overline{\Delta\hat{u}}_{t+h}^2}{\sqrt{n^{-1}\sum_{i=1}^{n}(\Delta L_{i,t+h|t} - [(\hat{\lambda}_{i,1}'\bar{e}_{t+h})^2 - (\hat{\lambda}_{i,2}'\bar{e}_{t+h})^2] - \hat{c}_{t+h} + \hat{u}_{i,t+h}'\hat{D}_{t+h})^2}}, \tag{24}$$

where $\hat{c}_{t+h} = n^{-1}\sum_{i=1}^{n}(\Delta L_{i,t+h|t} - [(\hat{\lambda}'_{i,1}\bar{e}_{t+h})^2 - (\hat{\lambda}'_{i,2}\bar{e}_{t+h})^2] + \hat{u}'_{i,t+h}\hat{D}_{t+h})$,

$$\overline{\Delta\hat{u}}^2_{t+h} = n^{-1}\sum_{i=1}^{n}\Delta L_{i,t+h} - n^{-1}\sum_{i=1}^{n}[(\hat{\lambda}'_{i,1}\bar{e}_{t+h})^2 - (\hat{\lambda}'_{i,2}\bar{e}_{t+h})^2] \tag{25}$$

and

$$\hat{D}_{t+h} = n^{-1}\sum_{i=1}^{n}(\hat{\lambda}_{i,1}\hat{\lambda}'_{i,1} - \hat{\lambda}_{i,2}\hat{\lambda}'_{i,2})\bar{e}_{t+h}. \tag{26}$$

Using these definitions, we now have the following result:

Theorem 5. *Suppose that Assumptions 3 and 4 hold. Then under* $H_0 : n^{-1}\sum_{i=1}^{n} E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F}) = 0$,

$$S_{cce} \to^d N(0,1).$$

Using that S_{cce} follows a standard Gaussian distribution asymptotically, we can compute a $1-\alpha$ confidence interval for $n^{-1}\sum_{i=1}^{n} E(u^2_{i,t,1} - u^2_{i,t,2} \mid \mathcal{F})$ as

$$\overline{\Delta\hat{u}}^2_{t+h} \pm \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{n^{-1}\sum_{i=1}^{n}(\Delta L_{i,t+h|t} - [(\hat{\lambda}'_{i,1}\bar{e}_{t+h})^2 - (\lambda'_{i,2}\bar{e}_{t+h})^2] - \hat{c}_{t+h} + \hat{u}'_{i,t+h}\hat{D}_{t+h})^2} \tag{27}$$

3.3.2 Squared Bias Differences

Next, consider the squared bias component of the MSE loss differential. Define

$$D_{t+h} = \bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\left(n^{-1}\sum_{i=1}^{n}[\lambda_{i,1}\lambda'_{i,1} - \lambda_{i,2}\lambda'_{i,2}]\right) f_{t+h}.$$

Using

$$\sqrt{n}\left(n^{-1}\sum_{i=1}^{n}\left[(\hat{\lambda}'_{i,1}\bar{e}_{t+h})^2 - (\hat{\lambda}'_{i,2}\bar{e}_{t+h})^2\right] - n^{-1}\sum_{i=1}^{n}\left[(\lambda'_{i,1}f_{t+h})^2 - (\lambda'_{i,2}f_{t+h})^2\right]\right) = 2n^{1/2}\bar{u}'_{t+h}D_{t+h} + o_P(1),$$

it follows that $n^{-1}\sum_{i=1}^{n}\left[(\hat{\lambda}'_{i,1}\bar{e}_{t+h})^2 - (\hat{\lambda}'_{i,2}\bar{e}_{t+h})^2\right]$ is a $\sqrt{n}$-consistent estimator for the average difference in the squared bias differential, $n^{-1}\sum_{i=1}^{n}\left[(\lambda'_{i,1}f_{t+h})^2 - (\lambda'_{i,2}f_{t+h})^2\right]$, where the estimation error is asymptotically $2\bar{u}'_{t+h}D_{t+h}$. To construct tests for the squared bias difference, consider the following test statistic

$$B_{n,2} = \frac{n^{-1/2}\sum_{i=1}^{n}\left[(\hat{\lambda}_{i,1}'\bar{e}_{t+h})^2 - (\hat{\lambda}_{i,2}'\bar{e}_{t+h})^2\right]}{2\sqrt{n^{-1}\sum_{i=1}^{n}(\hat{u}_{i,t+h}'\hat{D}_{t+h})^2}}, \tag{28}$$

where, again, $\hat{u}_{i,t+h} = e_{i,t+h} - \hat{\lambda}_i'\bar{e}_{t+h}$. The following result characterizes the distribution of this statistic:

Theorem 6. *Suppose that Assumption 3 holds. Then under* H_0 : $n^{-1}\sum_{i=1}^{n}\left[(\lambda_{i,1}'f_{t+h})^2 - (\lambda_{i,2}'f_{t+h})^2\right] = 0$,

$$B_{n,2} \to^d N(0,1).$$

Using Theorem 6, we can construct a confidence interval for the average squared bias differential, $n^{-1}\sum_{i=1}^{n}\left[(\lambda_{i,1}'f_{t+h})^2 - (\lambda_{i,2}'f_{t+h})^2\right]$ as

$$n^{-1}\sum_{i=1}^{n}\left[(\hat{\lambda}_{i,1}'\bar{e}_{t+h})^2 - (\hat{\lambda}_{i,2}'\bar{e}_{t+h})^2\right] \pm \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{n^{-1}\sum_{i=1}^{n}(\hat{u}_{i,t+h}'\hat{D}_{t+h})^2}. \tag{29}$$

Again, this confidence interval is asymptotically exact.

Comparing (27) and (29), we note a difference in the asymptotics. Although both variance expressions have $\hat{u}_{i,t+h}'\hat{D}_{t+h}$, the former has the additional term $\Delta L_{i,t+h|t} - [(\hat{\lambda}_{i,1}'\bar{e}_{t+h})^2 - (\hat{\lambda}_{i,2}'\bar{e}_{t+h})^2] - \hat{c}_{t+h}$. This difference could well make a difference to the finite-sample performance of the two tests. For example, overfitting could result in a very small $\hat{u}_{i,t+h}$ and thus a small $\hat{u}_{i,t+h}'\hat{D}_{t+h}$. By including the extra term, tests associated with Theorem 5 might be more robust in small samples.

3.4 Factor Structure Estimated by PCA

An alternative to the CCE approach in Section 3.3 is to use principal components analysis (PCA) to extract the common factors. A notable advantage of the PCA approach is that, unlike the CCE approach, the number of observed forecast errors does not pose an upper bound on the number of factors. In practice, this means that we can allow for more factors under the PCA approach.

Define the difference in the idiosyncratic forecast error variance

$$\overline{\Delta\hat{u}}^2_{t+h} = n^{-1}\sum_{i=1}^{n} \Delta L_{i,t+h} - n^{-1}\sum_{i=1}^{n}\left[(\hat{\lambda}'_{i,1}\hat{f}_{t+h})^2 - (\hat{\lambda}'_{i,2}\hat{f}_{t+h})^2\right]. \tag{30}$$

As before, let $\hat{f}_{t+h}$ and $\hat{\lambda}_i$ be the estimated factors and factor loadings obtained using PCA estimation. Then we have the following results:

Lemma 2. *Under Assumptions A-F in Bai (2003), we have*

$$\begin{aligned}&\sqrt{n}\left[\overline{\Delta\hat{u}}^2_{t+h} - n^{-1}\sum_{i=1}^{n} E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F})\right] \\ &= n^{-1/2}\sum_{i=1}^{n}\left[(u^2_{i,t+h,1} - u^2_{i,t+h,2}) - E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F}) + 2(\lambda'_{i,1}f_{t+h}u_{i,t+h,1} - \lambda'_{i,2}f_{t+h}u_{i,t+h,2})\right] \\ &\quad + o_P(1).\end{aligned}$$

Notice that we no longer have a term involving $\hat{D}_{t+h}$. Depending on the distribution of the idiosyncratic term, the PCA approach might yield a more efficient estimator than the CCE approach since it does not require us to estimate this term.

From this point, all steps in the inference procedure are exactly the same as those in Section 3.3, except that $(\hat{\lambda}'_{i,1}\bar{e}_{t+h}, \hat{\lambda}'_{i,2}\bar{e}_{t+h})$ is replaced by the PCA estimate $(\hat{\lambda}'_{i,1}\hat{f}_{t+h}, \hat{\lambda}'_{i,2}\hat{f}_{t+h})$ and we set $\hat{D}_{t+h} = 0$. Specifically, in Equations (24), (25) and (27), we replace $(\hat{\lambda}'_{i,1}\bar{e}_{t+h}, \hat{\lambda}'_{i,2}\bar{e}_{t+h})$ with the PCA estimate $(\hat{\lambda}'_{i,1}\hat{f}_{t+h}, \hat{\lambda}'_{i,2}\hat{f}_{t+h})$ and set $\hat{D}_{t+h} = 0$. We also replace $B_{n,2}$ in (28) with the following

$$\tilde{B}_{n,2} = \frac{n^{-1/2}\sum_{i=1}^{n}\left[(\hat{\lambda}'_{i,1}\hat{f}_{t+h})^2 - (\hat{\lambda}'_{i,2}\hat{f}_{t+h})^2\right]}{2\sqrt{n^{-1}\sum_{i=1}^{n}(\hat{\lambda}'_{i,1}\hat{f}_{t+h}\hat{u}_{i,t+h,1} - \hat{\lambda}'_{i,2}\hat{f}_{t+h}\hat{u}_{i,t+h,2})^2}}, \tag{31}$$

where $\hat{u}_{i,t+h,m} = e_{i,t+h,m} - \lambda'_{i,m}f_{t+h}$.

4 Empirical Application to Earnings Forecasts

To illustrate the economic insights that can be gained from our new test statistics, we next conduct an empirical analysis that compares the accuracy of analysts' forecasts of quarterly earnings recorded across six large brokerage firms.

4.1 Data

Using data from the Institutional Brokers Estimate System (IBES), we examine forecasts of quarterly earnings per share (EPS) generated by analysts at six large brokerage firms, namely Merrill Lynch (MERRILL), JP Morgan Chase (JPMORGAN), Credit Suisse (FBOSTON), Goldman Sachs (GOLDMAN), Morgan Stanley (MORGAN) and Deutsche Bank (LAWRENCE). Analysts' forecasts are not always updated so frequently at long horizons, so we focus on forecasts generated at the two-month horizon to avoid issues caused by stale forecasts.[8]

Our quarterly data span the 20-year period from 2000Q1 to 2020Q1. Table 1 presents summary statistics on the number of firms covered by each brokerage firm (Panel A) as well as the average number of firms covered each quarter (Panel B). The total number of firms covered by the brokerage firms in at least one quarter ranges from 1,437 (Lawrence) to 1,825 (Merrill), while the average number of firm-level quarterly EPS estimates reported by the brokerage firms ranges from 239 (Lawrence) to 356 (Merrill).

In addition to inspecting the forecasting performance across all firms, we also use SIC codes to assign individual firms to five industry groupings chosen to match the Fama-French industry classification, namely Consumer, Manufacturing, High Tech, Health, and Other. Firm numbers are highest in the Other category, followed by High Tech, Manufacturing, Consumer, and Health.

[8]We calculate the forecast horizon using daily data on the announcement date (ANNDATS) and forecast period end date (FPEDATS).

4.2 Factor Structure in Errors and Loss Differentials

Table 2 presents results from testing for the presence of common factors in the EPS forecast errors for the six brokerage firms using the growth ratio (GR) and eigenvalue (ER) statistics of Ahn and Horenstein (2013) as well as the Onatski (2009) test (ED).[9] For three of the brokerage firms (Fboston, Goldman, and Merrill), the three tests identify a single common factor in the forecast errors, while for a fourth (Lawrence), two of the tests suggest a single common factor while the third (ED) uncovers three factors. For the remaining two brokerages, the tests identify either zero (JP Morgan) or two (Morgan) common factors.

Given these findings, we next inspect whether controlling for a common factor in the EPS forecast errors captures their correlation. To this end, Table 3 reports the average correlation between the forecast errors without controlling for a common factor (top row labeled 0) as well as after controlling for one or two common factors (second and third rows), along with values of the test statistic of Pesaran (2004). Under the null that the data are uncorrelated, this test statistic is asymptotically normally distributed. The average cross-sectional correlation in forecast errors ranges from 0.07 (Morgan) to 0.12 (Lawrence). Moreover, the underlying correlations are highly statistically significant with test statistics exceeding 25, indicating very strong evidence of cross-correlations in all brokerage firms' EPS forecast errors.[10]

Controlling for exposures to a single common factor, average correlations drop to a much narrower range from -0.01 to 0.03. While some of these test statistics remain statistically significant–notably for Morgan Stanley–the test statistics typically come down by more than an order of magnitude as does the average cross-correlation estimate. Accounting for a second common factor only has a marginal effect on average correlations and test statistics, except for Morgan Stanley whose average correlation declines from 0.03 to 0.01. We conclude

[9]We estimate the factors from the subset of firm-brokerage pairings with at least 40 quarterly observations, corresponding to half of the sample period.

[10]Cross-sectional regressions of EPS outcomes on brokerage forecasts yield predictive R^2-values in the range 0.5-1, with an average of 0.91.

that very little common variation remains in the forecast errors after accounting for a single common factor and, hence, that the setup in (3) appears to provide an accurate empirical characterization for our data.

We next explore evidence of heterogeneity in cluster loadings across industries. To the extent that industries differ in how sensitive their earnings are to the economic cycle, we might expect factor loadings to be clustered along industry lines with firms within a particular industry exhibiting more similar factor loadings than firms belonging to different industries. To see if this holds, we estimate a common factor model $\tilde{e}_{it+h} = \lambda_i f_{t+h} + \varepsilon_{it+h}$ on the standardized forecast errors ($\tilde{e}_{it+h}$) subject to the constraint $\sum_{i=1}^{N} \lambda_i^2 = 1$. Specifically, we first demean and scale the forecast errors so they have mean zero and unit standard deviation. Next, we estimate factors and factor loadings by PCA using the EM algorithm.

Table 4 shows the standard deviation of the estimated factor loadings across all firms (first column) as well as within the five industry clusters. Factor loadings that are more homogeneous within a particular industry than in the aggregate should give rise to smaller values of the standard deviations than in the first column. We see modest evidence of this: For all but one of the six brokerage firms, the standard deviation of the factor loadings is smaller in three of the five industries compared to in the aggregate. Similarly, for the Consumer, Manufacturing, and High Tech industries, the standard deviation of factor loadings is smaller than the standard deviation of factor loadings in the aggregate for four of the six brokerages. For the "Other" industry, there is typically higher heterogeneity in factor loadings than what we see in the aggregate, indicating that this industry group includes many heterogeneous firms.

4.3 Test Results

We next use our new cross-sectional tests of equal predictive accuracy to compare the EPS forecasts. With six brokerage firms, we can conduct a total of 15 pair-wise comparisons. To focus the discussion, we concentrate on four pairs, namely Morgan Stanley vs. Goldman, Morgan Stanley vs. Merrill, Goldman vs. Merrill, and Lawrence (Deutsche Bank) vs. Merrill.[11]

Figure 1 plots time-series of the quarterly values of the cross-sectional average test statistics for the null of equal predictive accuracy. We show separate lines for the test statistics assuming homogeneous factor loadings, (7), used to test the unconditional null in (4), and heterogeneous factor loadings, (10), used to test the conditional null in (5). In each panel, positive values of the test statistic indicate that the second forecaster is more accurate than the first forecaster, while negative values suggest the reverse.

The first point to note is that the two sets of test statistics in (7) and (10) are very similar even though they test different hypotheses and deal with factor-related shocks in different ways. This similarity arises because the tests only differ with respect to the centering of the terms in the denominator which turns out to be of little importance.

Next, consider the pairwise comparisons starting with Morgan Stanley vs. Goldman (top left panel). In most quarters during our sample, the test statistic is not statistically significant, the three exceptions being 2004Q4, 2012Q1 and 2020Q1 where Goldman's forecasts are significantly more accurate than Morgan Stanley's. Comparing Morgan Stanley vs. Merrill (top right corner), Merrill comes out on top in two quarters (2001Q3, 2018Q4). The pairwise comparison of Goldman vs. Merrill (bottom left) only shows one quarter (2004Q3) with significant underperformance for Merrill relative to Goldman, while Lawrence produces significantly more accurate earnings forecasts than Merrill (bottom right) in five quarters (2005Q4, 2011Q1, 2013Q4, 2014Q2 and 2017Q3) and only underperforms significantly during a single quarter (2007Q3).

An important point to bear in mind when interpreting these results is that we are inspecting multiple test statistics–81 in this case–which introduces a multiple hypothesis testing problem. While we do not deal with this issue here, Qu et al. (2019) develop a Sup-type bootstrap approach that evaluates the joint statistical significance of individual test statis-

[11] For each of the pairwise comparisons of firm-level EPS forecasts, our analysis imposes a requirement of at least five observations.

tics.

We conclude the following from these results. First, the empirical results are very robust to whether we assume homogeneous or heterogeneous factor loadings and test the null of equal cross-sectional average predictive accuracy unconditionally or conditional on the factors and factor loadings. Second, our results suggest that the brokerage firms produce short-term earnings forecasts that are equally accurate during the vast majority of quarters but also indicate that there are significant differences in predictive accuracy in a few periods.

4.4 Decomposition Results

Figure 2 presents a set of heat diagrams displaying the quarterly values of the cross-sectional tests statistics used to test the null of equal idiosyncratic variances (23) for pairs of brokerage firms. Each panel corresponds to a particular pair-wise comparison, using the four pairs from Figure 1. Red colors indicate quarters in which the first forecaster has a larger idiosyncratic error variance component than the second forecaster, while blue colors indicate the reverse. Asterisks mark quarters in which the test statistic is significant at the 5% level, using a two-sided test. Each diagram contains three rows showing results based on the PCA, CCE, and cluster approaches, respectively.

First consider the comparison of Morgan Stanley vs. Goldman (top panel). The test statistics fluctuate around zero in most quarters without being statistically significant. Using the PCA-based test we see find two quarters in which Morgan Stanley's idiosyncratic error variance was significantly higher than that of Goldman while for the CCE and cluster tests this holds in zero and one quarter, respectively. Given that we are considering 81 quarterly test statistics, this number of rejections is lower than what we would expect by random chance and so does not provide strong evidence that idiosyncratic error variances differ in any significant way across the two brokerage firms. Similar results hold for the Morgan Stanley vs. Merrill Lynch and Goldman vs. Merrill Lynch comparisons. The comparison of the idiosyncratic error variances of Lawrence versus Merrill Lynch (bottom panel) leads to more rejections of the null of equal accuracy–four in total–when based on the PCA method, with one and four rejections using the CCE and cluster methods, respectively.

In total, across the four pair-wise comparisons in Figure 2, the PCA test produces eleven rejections of the null, while the CCE and cluster-based methods record five and nine rejections. These findings suggest that there is little overall evidence of systematic differences in the magnitude of the idiosyncratic error variance component of analysts' EPS estimates.

Figure 3 shows the outcome of cross-sectional comparisons of the squared bias component in the errors of the four pairs of brokerage firms' EPS forecasts using the test statistics in (21), (28) and (31). Starting with the Morgan Stanley vs. Goldman comparison, we find nine rejections of the null of equal squared biases based on the PCA test, six rejections based on the CCE test, and a single rejection based on the cluster test. Rejection rates are lower for the three other pairwise comparisons, with five to six rejections for the PCA-based test, two to seven rejections for the CCE-based test, and zero or one rejections for the cluster test.

Overall, across the four pair-wise comparisons, we find 26 rejections of the null based on the PCA test, 22 rejections based on the CCE test, and only two rejections based on the cluster test. Hence, for the PCA and CCE-based tests, the rejection rate is somewhat higher than what we would expect by random chance from applying a test with a 5% size to 324 cross-sectional comparisons (16 rejections), while clearly this is not the case for the cluster-based test. Many of the rejections based on the PCA and CCE tests occur during the Global Financial crisis (2008-09). During this period, factor volatility is likely to have been higher than normal and so this could have boosted the power of the test for equal squared biases.

There are good theoretical reasons why the PCA approach appears to have better power in our empirical application. First, the cluster-based method is likely to be conservative since its asymptotic size is not exact; in fact, in Theorems 3 and 4, the asymptotic size is only bounded by the nominal size, rather than being equal to it. Second, the asymptotic variance of the estimates for the difference in squared biases and idiosyncratic variances can be smaller under PCA than under CCE because, as we noted after Lemma 2, the PCA estimator tends to be more efficient than the CCE, thus increasing its relative power.

The empirical results displayed in Figures 2 and 3 show notable differences across the tests of equal idiosyncratic variance versus equal squared biases. The Monte Carlo simulations reported in the next section suggest that the test for equal idiosyncratic error variances can be quite conservative which is consistent with the smaller number of rejections of the null for this test compared to the test for equal squared error bias.

The test statistics plotted in Figures 1-3 provide evidence on the statistical significance of differences in squared error loss, idiosyncratic variances, and squared biases. They do not show how much of the variation in differences in squared error loss is explained by differences in the idiosyncratic variance and squared bias components. To address this point, Table 5 reports the mean and variance of the contributions from these components, both measured relative to the total loss differential. Specifically, defining the cross-sectional sample moments

$$\overline{bias^2_{t+h}} = n^{-1} \sum_{i=1}^{n} \left[(\hat{\lambda}'_{i,1} \hat{f}_{t+h})^2 - (\hat{\lambda}'_{i,2} \hat{f}_{t+h})^2 \right],$$

$$\overline{\triangle u^2_{t+h}} = n^{-1} \sum_{i=1}^{n} \left(\hat{u}^2_{i,t+h,1} - \hat{u}^2_{i,t+h,2} \right),$$

the columns labeled mean ratio in Table 5 report the time-series averages

$$\frac{100}{81} \left(\frac{\sum_{t=2000Q1}^{2020Q1} \overline{\triangle u^2_{t+h}}}{\sum_{t=2000Q1}^{2020Q1} \overline{\triangle L}_{t+h}} \right), \qquad \frac{100}{81} \sum_{t=2000Q1}^{2020Q1} \left(\frac{\sum_{t=2000Q1}^{2020Q1} \overline{bias^2_{t+h}}}{\sum_{t=2000Q1}^{2020Q1} \overline{\triangle L}_{t+h}} \right),$$

for the idiosyncratic variance (top panel) and squared bias (bottom panel) components, respectively. As before, $\overline{\triangle L}_{t+h} = n^{-1} \sum_{i=1}^{n} \Delta L_{i,t+h|t}$. From (11), these measures can be positive or negative but sum to 100.

Similarly, columns labeled variance ratio report the following

$$\frac{Var(\overline{\triangle u^2_{t+h}})}{Var(\overline{\triangle L}_{t+h})}, \qquad \frac{Var(\overline{bias^2_{t+h}})}{Var(\overline{\triangle L}_{t+h})},$$

where

$$Var(\overline{\triangle L}_{t+h}) = \frac{1}{80}\sum_{t=2000Q1}^{2020Q1}\left(\overline{\triangle L}_{t+h} - \overline{\overline{\triangle L}}_{t+h}\right)^2,$$

$$Var(\overline{\triangle u^2_{t+h}}) = \frac{1}{80}\sum_{t=2000Q1}^{2020Q1}\left(\overline{\triangle u^2_{t+h}} - \overline{\overline{\triangle u^2_{t+h}}}\right)^2,$$

$$Var(\overline{bias^2_{t+h}}) = \frac{1}{80}\sum_{t=2000Q1}^{2020Q1}\left(\overline{bias^2_{t+h}} - \overline{\overline{\triangle bias^2_{t+h}}}\right)^2.$$

and $\overline{\overline{\triangle L}}_{t+h} = (1/81)\sum_{t=2000Q1}^{2020Q1}\overline{\triangle L}_{t+h}$, $\overline{\overline{\triangle u^2_{t+h}}} = (1/81)\sum_{t=2000Q1}^{2020Q1}\overline{\triangle u^2_{t+h}}$, and $\overline{\overline{\triangle bias^2_{t+h}}} = (1/81)\sum_{t=2000Q1}^{2020Q1}\overline{\triangle bias^2_{t+h}}$. These variance ratios do not sum to 100 because of the omitted covariance term.

The mean ratios of the four pairwise comparisons reported in Table 5 are generally notably higher for differences in the idiosyncratic variances than for differences in squared biases, with the former falling within ranges of 40-87%, 1-95%, and 85-96% for the PCA, CCE, and cluster methods, respectively. Variance ratios are also higher–typically by a large margin–for differences in the idiosyncratic variance component than for differences in the squared biases for all but one pairwise comparison (Morgan Stanley vs. Goldman, CCE method). Variation in the idiosyncratic variance component is thus generally substantially more important to explaining squared error loss differences between brokerage firms' earnings forecasts than variation in the squared bias term.

These results show that, on average, differences in idiosyncratic error variances account for far more of squared error loss differences in brokerage firms' EPS forecasts than the squared bias component. Differences in brokerage firms' quarterly EPS forecast accuracy therefore appear not so much to be driven by differences in their ability to predict common factors, i.e., their skills as "generalists". Rather, differences in predictive accuracy tend to be driven by differences in brokerage firms' ability to reduce uncertainty about the idiosyncratic component of EPS as this relates to their specialist knowledge of individual firm performance. The main exception to this finding occurs around the Global Financial Crisis (2008-09) during which the squared bias term becomes more important in explaining differences in squared-error losses across brokerages, particularly for the PCA-based test.

5 Monte Carlo Simulations

Our final section reports the outcome of a set of Monte Carlo simulations which address the finite-sample properties of our tests.

5.1 Setup

Our baseline simulations use a simple setup designed to satisfy the assumptions of the three different estimation procedures (clustering, CCE and PCA) which allows us to more directly compare their performance. First, we generate factors $f_{1,t}$ and $f_{2,t}$ as i.i.d variables from the standard normal distribution. Next, we compute realized outcomes as $y_{it+h} = f_{1,t}+f_{2,t}+\varepsilon_{it+h}$, while forecasts are generated as $\hat{y}_{it+h|t,1} = f_{1,t} + \xi_{it+h,1}$ and $\hat{y}_{it+h|t,2} = f_{2,t} + \xi_{it+h,2}$, where ε_{it+h}, $\xi_{it+h,1}$ and $\xi_{it+h,2}$ are mutually independent i.i.d. $N(0, \sigma^2)$ draws. We calibrate σ^2 to yield a value for the predictive power ρ^2 in a certain range, where for $m \in \{1, 2\}$,

$$\rho^2 = 1 - \frac{E(y_{it+h} - \hat{y}_{it+h|t,m})^2}{Ey_{it+h}^2}.$$

Because $\rho^2 = 1/(2 + \sigma^2)$, $\rho^2 \in (0, 1/2)$. We set $n \in \{10, 25, 50, 100, 200, 1000\}$ and $\rho^2 \in \{0.05, 0.1, 0.2, 0.25, 0.3, 0.4, 0.45\}$ as well as $T = 80$. All results are based on 2000 random samples.

Initially we consider the performance in a typical time period ($t = 3$) and note that results for other time periods would be similar, given the i.i.d. setting. Section 5.4.2 introduces breaks to the data generating process and so considers both pre- and post-break performance.

Because we are testing a random hypothesis, the hypothesized value is not zero but a random quantity that depends on the realization of the factors and factor loadings. For this reason, and to simplify the presentation of size and power results, we invert our test statistics to form 95% confidence intervals for $E(\overline{\Delta L}_{t+h} \mid \mathcal{F})$ and report the coverage probabilities.

5.2 Baseline results

Table 6 reports results on the procedure for conducting inference on $E(\overline{\Delta L}_{t+h} \mid \mathcal{F}) = 0$ described in Section 2.5. Coverage probabilities are generally quite accurate although there is some undercoverage for very small values of n, suggesting that the test might slightly overreject in such cases.

Next, we invert the procedures described in Section 3 to construct 95% confidence intervals for the squared error loss decompositions based on the clustering, CCE, and PCA methods. Table 7 reports results for the average difference in the squared bias component $n^{-1}\sum_{i=1}^{n}\left[(\lambda_{i,1}' f_{t+h})^2 - (\lambda_{i,2}' f_{t+h})^2\right]$ while results for the average difference in the idiosyncratic variance component $n^{-1}\sum E(u_{i,t+h,1}^2 - u_{i,t+h,2}^2 \mid \mathcal{F})$ are reported in Table 8.

For the tests applied to the squared bias terms (Table 7), the coverage probability generally improves with the sample size n, with exception of the PCA method when ρ^2 is very small.[12] In larger samples, the coverage probability for the difference in squared bias is relatively stable as a function of ρ^2, while for smaller values of n, the tests are under-sized for small values of ρ^2 and oversized for large values of ρ^2. Conversely, the confidence intervals for the difference in variances (Table 8) tend to be more conservative when ρ^2 is large, with coverage probabilities exceeding 99%. To understand this finding, note that $\rho^2 = 1/(2+\sigma^2)$ and inference on the difference in variance relies on variation in $u_{it+h|m}$ which is $2\sigma^2$ in this case. Large values of ρ^2 are therefore associated with smaller variation in $u_{it+h|m}$ and so the higher-order terms in the asymptotic expansion tend to be more pronounced which means

[12] A reason for this finding is that for $n = 1000$ and $T = 80$, the two dimensions of the sample size are not very balanced and the accuracy of the PCA method is determined by $\min\{n, T\}$.

that the first-order asymptotic approximation underlying the inference procedure is generally less accurate.

Overall, the procedure for testing differences in idiosyncratic variances has better coverage than its counterpart for testing differences in the squared bias component. This might be explained by the greater robustness of the test for equal idiosyncratic variances highlighted earlier. Moreover, the size distortion results suggest that tests for equal squared biases are likely to have more power than tests for equal idiosyncratic error variances.

5.3 Decompositions with heterogeneous factor loadings

We next consider various extensions to the baseline simulation setup. To keep the presentation short, all results are reported in a set of Appendix tables.

5.3.1 Heterogeneous factor loadings across Clusters

Our first extension allows factor loadings to have a cluster structure. Specifically, we partition the cross-section of n units into five equal-sized clusters and set $\hat{y}_{it+h|t,1} = f_{1,t}\lambda_{k(i)} + \xi_{it+h,1}$, where $\lambda_{k(i)} \in \{0, 0.5, 1, 1.5, 2\}$ and $k(i)$ is the cluster that contains unit i. Similarly, we set $\hat{y}_{it+h|t,2} = f_{2,t}\lambda_{k(i)} + \xi_{it+h,2}$ with $\lambda_{k(i)} \in \{0, 0.5, 1, 1.5, 2\}$. Since each cluster contains $n/5$ units, the clusters are very small for the smallest values of n, i.e., only two and five units per cluster for $n = 10$ and $n = 25$, respectively.

Results from this setup are reported in Appendix Tables A1 and A2. For inference on differences in the squared bias (Table A1), the clustering method has a substantial under-coverage for small values of n but performs notably better with larger sample sizes. This is as expected since the clustering method uses the cluster-wise average to estimate the factor structure and the size of each cluster is $n/5$. The CCE method mostly has sufficient coverage probability while the PCA method tends to be very conservative with coverage probabilities at or above 99%. For inference on differences in the error variance (Table A2), all three methods perform reasonably well across various sample sizes, although the clustering and CCE approaches tend to be somewhat conservative while, conversely, the PCA method overrejects if n is very small ($n = 10$).

5.3.2 General heterogeneous factor loadings

Our second extension applies a more general setting in which factor loadings are neither constant, nor have a cluster structure as we generate factor loadings as the absolute value of a standard normal distribution, i.e., $\lambda_{1,i}$ and $\lambda_{2,i}$ are i.i.d $|N(0,1)|$. Using absolute values ensures that $E(\lambda_{1,i}) = E(\lambda_{2,i})$ is positive as required by the CCE method (Assumption 3). Conversely, the clustering method is no longer valid in this setting and so we omit results for this method. Using these heterogeneous factor loadings, we set $y_{it+h} = \lambda_{1,i} f_{1,t} + \lambda_{2,i} f_{2,t} + \varepsilon_{it+h}$ and generate forecasts as $\hat{y}_{it+h|t,1} = \lambda_{1,i} f_{1,t} + \xi_{it+h,1}$ and $\hat{y}_{it+h|t,2} = \lambda_{2,i} f_{2,t} + \xi_{it+h,2}$, where ε_{it+h}, $\xi_{it+h|1}$ and $\xi_{it+h|2}$ are again drawn independently with mean zero and variance σ^2. Appendix Table A3 shows that the coverage of the CCE method is often better than that of the PCA method for inference on differences in the squared bias with the latter having issues with undercoverage for small values of n; both methods provide sufficient overall coverage but tend to be conservative for inference on differences in variances, particularly when ρ^2 is large.

5.4 Variation in the factor structure

5.4.1 Three factors

The key reason for using the PCA method is that once the number of factors exceeds two, PCA is the only valid method for handling the general case with heterogeneous factor loadings.[13] We illustrate this point in a setting with three factors as we set $y_{it+h} = \lambda_{1,i} f_{1,t} + \lambda_{2,i} f_{2,t} + \lambda_{3,i} f_{3,t} + \varepsilon_{it+h}$ and generate the forecasts as $\hat{y}_{it+h|t,1} = \lambda_{1,i} f_{1,t} + \xi_{it+h,1}$ and $\hat{y}_{it+h|t,2} = \lambda_{2,i} f_{2,t} + \xi_{it+h,2}$, where all variables (including all factors and factor loadings)

[13] Another reason for using the PCA method is that it remains asymptotically valid even if $E(\lambda_i) = 0$, whereas the CCE method would fail in this setting.

are generated as before.

As shown in Appendix Table A4, the 95% coverage probability for the CCE method can be as low as 40% for comparing the squared bias and as low as 59% for comparing variances when $n = 1000$ and $\rho^2 = 0.45$. This phenomenon arises because we only observe two variables (two forecast errors) and the CCE method can handle at most two factors in our setup. With more than two factors, the CCE method does not guarantee consistent estimation of the factor structure. Since we are studying the average across n units, the problem becomes more pronounced as n increases.

5.4.2 Breaks in the number of factors

We next consider a setting in which the number of factors changes as represented by a discrete break to the factor structure: $y_{it+h} = \lambda_{1,i} f_{1,t} + \lambda_{2,i} f_{2,t} + \lambda_{3,i} f_{3,t} \mathbf{1}_{\{t>T/2\}} + \varepsilon_{it+h}$. In this model, the third factor ($f_{3,t}$) only shows up in the second half of the sample. All other details remain the same. Instability in the number of factors is empirically plausible and has been studied in Cheng et al. (2016).

Appendix Table A5 reports coverage probabilities for 95% confidence intervals based on the PCA and CCE methods. We consider two time periods: one before the break ($t = 3$), the other after the break ($t = T - 3$).[14] Overall, the PCA method maintains sufficient coverage probability while the CCE method can suffer from severe undercoverage. Again, the reason is that when there are three factors, CCE cannot consistently estimate the factor structure from two observed variables. The performance of the PCA approach is similar before and after the break. Conversely, the CCE method performs worse after the break than before, most likely because there are only two factors before the break, consistent with a setting in which the CCE approach is valid.

[14] We conduct the PCA analysis for the full sample using three factors because there are three spiked eigenvalues in the data matrix for the full sample.

5.5 Linex Loss

Table A6 reports 95% confidence intervals for testing the null of equal conditional expected loss (5) using linex loss:

$$L(e_{it+h}) = \frac{1}{a^2}\left[\exp(ae_{it+h}) - ae_{it+h} - 1\right] \tag{32}$$

where $a = 1$. The data generating process is identical to that in the baseline case used to construct Table 6. Coverage probabilities are very similar to those in Table 6, with a slight undercoverage for small values of n and coverage approximating 95% as n grows larger.

5.6 Conditional heteroskedasticity

We now conduct a set of simulations in which the data generating process allows for conditional heteroskedasticity modeled through a simple ARCH process of the form:

$$f_t = \sigma_t \varepsilon_t,$$

where $\sigma_t^2 = (1 - r) + rf_{t-1}^2$ with $r = 0.5$. Notice that $E\sigma_t^2 = 1$.

Results are reported in Appendix Tables A7 and A8. Compared to the baseline setup in Tables 7 and 8, the results do not change in any material ways, showing that conditional heteroskedasticity in the innovations of the data generating process need not have a material effect on the performance of our cross-sectional tests for equal predictive accuracy.

5.7 Relation to empirical results

In our empirical analysis, the PCA and CCE methods lead to notably more rejections of the null of equal squared biases than the clustering method which rarely rejects the null. To help explain these results, we slightly modify the simulation setup so as to match the high cross-sectional R^2 values found in our application (0.9 on average) and allow for broad heterogeneity in factor loadings. We accomplish this by adding a third factor to the model and letting factor loadings be random: $y_{it+h} = \lambda_{i1} f_{1,t} + \lambda_{i2} f_{2,t} + \lambda_{i3} f_{3,t} + \varepsilon_{it+h}$, where $f_{1,t}, f_{2,t}, f_{3,t} \sim iidN(0,1)$, $\lambda_{i1} \sim iidN(0,V)$, $\lambda_{i2}, \lambda_{i3} \sim iidN(1,1)$, and $\varepsilon_{i,t+h} \sim iidN(0,\sigma^2)$. The two forecasts are generated as $\hat{y}_{it+h|t,1} = \lambda_{i1} f_{1,t} + \lambda_{i2} f_{2,t}$ and $\hat{y}_{it+h|t,2} = \lambda_{i1} f_{1,t} + \lambda_{i3} f_{3,t}$, respectively, with forecast errors $e_{it+h,1} = \lambda_{i3} f_{3,t} + \varepsilon_{it+h}$ and $e_{it+h,2} = \lambda_{i2} f_{2,t} + \varepsilon_{it+h}$.

Normally distributed factor loadings is likely to cause the biggest problem for the clustering method which approximates heterogeneity in factor loadings by means of a small number of discrete values. Such heterogeneity is particularly important if the fraction of the variation in forecast errors explained by the omitted factors is large. Here, this is given by $\rho_e^2 = 2/(2+\sigma^2)$ and we vary σ to obtain a range of values $\rho_e^2 \in \{0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. To mimic the cross-sectional R^2, note that for each t, $1 - E[(y_{it+h} - \hat{y}_{it+h|t,1})^2 | f_{1,t}, f_{2,t}, f_{3,t}]/E[y_{it+h}^2 | f_{1,t}, f_{2,t}, f_{3,t}] = (Vf_{1,t}^2 + 2f_{2,t}^2)/(Vf_{1,t}^2 + 2f_{2,t}^2 + 2f_{3,t}^2 + \sigma^2)$. Matching the empirical evidence, we set this number to 0.9 for $f_{1,t}^2 = f_{2,t}^2 = f_{3,t}^3 = 1$ by choosing an appropriate value of V.

Simulation results that use this setup are reported in Appendix Table A9. When the omitted factors matter less for the variation in forecast errors (small ρ_e^2), the PCA and clustering methods generate tighter confidence intervals than the CCE approach. However, as the common factors gain in importance (high ρ_e^2), the PCA approach produces notably narrower confidence intervals than the clustering method, with the CCE approach in the middle. This is consistent with the clustering approach having weaker power and so helps explain the far lower rejection rate observed empirically for this estimator for the equality of squared bias tests in situations with substantial cross-sectional heterogeneity in factor loadings.

6 Conclusion

This paper develops new methods for testing the null of equal predictive accuracy on a single cross-section containing pairs of forecasts of multiple outcome variables. In settings where the cross-sectional dependence in forecast errors can be captured by a common factor structure, we show that it is possible to conduct formal inference about equal predictive accuracy and develop a set of test statistics. In particular, we show that the null of equal predictive accuracy can be conducted in settings with a large cross-sectional dimension if either (i) factor loadings are homogeneous across units so that the effect of common factors on forecast errors cancels out in squared error loss differentials; or (ii) we condition on factor realizations and conduct a test of equal predictive accuracy, given these factors.

We illustrate our tests in an empirical application that compares the accuracy of analyst short-term earnings forecasts across six brokerage firms, using a sample covering hundreds of individual firms. While our cross-sectional tests fail to reject the null of equal predictive accuracy for most quarters, we do identify individual quarters with significant differences among pairs of brokers. Moreover, our empirical results suggest that differences in the variances of the idiosyncratic error component tend to be more important than differences in squared biases for explaining variation in differences in brokerage firms' earnings per share squared-error loss performance.

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

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

Blake, D., Rossi, A. G., Timmermann, A., Tonks, I., and Wermers, R. (2013). Decentralized investment management: Evidence from the pension fund industry. *The Journal of Finance*, 68(3):1133–1178.

Chen, X., Shao, Q.-M., and Wu, W. B. (2016). Supplement to "self-normalized cramér-type moderate deviations under dependence". *The Annals of Statistics*.

Cheng, X., Liao, Z., and Schorfheide, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies*, 83(4):1511–1543.

Chong, Y. Y. and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *The Review of Economic Studies*, 53(4):671–690.

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

Davies, A. and Lahiri, K. (1995). A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics*, 68(1):205–227.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, pages 253–263.

Elliott, G., Komunjer, I., and Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies*, 72(4):1107–1125.

Giacomini, R., Lee, S., and Sarpietro, S. (2019). Microforecasting with individual forecast selection. *Unpublished working paper, UCL.*

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

Hansen, P. R. and Timmermann, A. (2015). Equivalence between out-of-sample forecast comparisons and wald statistics. *Econometrica*, 83(6):2485–2505.

Kacperczyk, M., Nieuwerburgh, S. V., and Veldkamp, L. (2014). Time-varying fund manager skill. *The Journal of Finance*, 69(4):1455–1484.

Keane, M. P. and Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. *American Economic Review*, 80(4):714–735.

Liu, L., Moon, H. R., and Schorfheide, F. (2018). Forecasting with dynamic panel data models. *Unpublished working paper, University of Pennsylvania.*

Liu, L., Moon, H. R., and Schorfheide, F. (2019). Forecasting with a panel tobit model. *Unpublished working paper, University of Pennsylvania.*

Mackowiak, B. and Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, 99(3):769–803.

McCracken, M. W. (2007). Asymptotics for out of sample tests of granger causality. *Journal of Econometrics*, 140(2):719–752.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479.

Patton, A. J. and Timmermann, A. (2012). Forecast rationality tests based on multi-horizon bounds. *Journal of Business & Economic Statistics*, 30(1):1–40.

Peña, V. H., Lai, T. L., and Shao, Q.-M. (2008). *Self-normalized processes: Limit theory and Statistical Applications.* Springer Science & Business Media.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Qu, R., Timmermann, A., and Zhu, Y. (2019). Do any economists have superior forecasting skills? *Working paper.*

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, pages 1067–1084.

A Proofs

This section presents proofs of the theoretical results in the main body of our paper.

A.1 Theorem 1

Proof. Using (6), we have

$$\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t}) = (u_{i,t+h,1}^2 - u_{i,t+h,2}^2) - E[u_{i,t+h,1}^2 - u_{i,t+h,2}^2] + 2(u_{i,t+h,1} - u_{i,t+h,2})\lambda_i' f_{t+h}.$$

Hence, conditional on $\mathcal{F}$, $\{\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t})\}_{i=1}^n$ is independent across i with mean zero. By Assumption 1, the sequence $\{\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t})\}_{i=1}^{n_t}$ conditional on $\mathcal{F}$ satisfies the Lyapunov condition. Hence, a standard argument yields

$$\frac{n^{-1/2}\sum_{i=1}^n[\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t})]}{\sqrt{n^{-1}\sum_{i=1}^n[\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t})]^2}} \xrightarrow{d} N(0,1).$$

Under the null that $n^{-1}\sum_{i=1}^n E(\Delta L_{i,t+h|t}) = 0$, we have

$$\frac{n^{-1/2}\sum_{i=1}^n \Delta L_{i,t+h|t}}{\sqrt{n^{-1}\sum_{i=1}^n[\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t})]^2}} \xrightarrow{d} N(0,1).$$

The result now follows by noticing that $n^{-1}\sum_{i=1}^n[\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t})]^2 \leq n^{-1}\sum_{i=1}^n(\Delta L_{i,t+h|t})^2$. □

A.2 Theorem 2

Proof. Using (8), we have

$$\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t} \mid \mathcal{F}) = (u_{i,t+h,1}^2 - u_{i,t+h,2}^2) - E[u_{i,t+h,1}^2 - u_{i,t+h,2}^2 \mid \mathcal{F}] + 2(u_{i,t+h,1} - u_{i,t+h,2})\lambda_i' f_{t+h}.$$

Hence, conditional on $\mathcal{F}$, $\{\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t} \mid \mathcal{F})\}_{i=1}^n$ is independent across i with mean zero. By Assumption 2, the sequence $\{\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t} \mid \mathcal{F})\}_{i=1}^{n_t}$ conditional on $\mathcal{F}$ satisfies the Lyapunov condition. Hence, a standard argument yields

$$\frac{n^{-1/2}\sum_{i=1}^n [\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t} \mid \mathcal{F})]}{\sqrt{n^{-1}\sum_{i=1}^n [\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t} \mid \mathcal{F})]^2}} \xrightarrow{d} N(0,1).$$

Under the null that $n^{-1}\sum_{i=1}^n E(\Delta L_{i,t+h|t} \mid \mathcal{F}) = 0$, we have

$$\frac{n^{-1/2}\sum_{i=1}^n \Delta L_{i,t+h|t}}{\sqrt{n^{-1}\sum_{i=1}^n [\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t} \mid \mathcal{F})]^2}} \xrightarrow{d} N(0,1).$$

The result now follows by noticing that $n^{-1}\sum_{i=1}^n [\Delta L_{i,t+h|t} - E(\Delta L_{i,t+h|t} \mid \mathcal{F})]^2 \leq n^{-1}\sum_{i=1}^n (\Delta L_{i,t+h|t})^2$. □

A.3 Theorem 3

Proof. Start by noticing that

$$\begin{aligned}
&\sqrt{n_k}\left(\overline{\Delta u}_{t+h,k}^2 - n_k^{-1}\sum_{i\in H_k} E(u_{i,t+h,1}^2 - u_{i,t+h,2}^2 \mid \mathcal{F})\right) \\
&= n_k^{-1/2}\sum_{i\in H_k}\left[(u_{i,t+h,1}^2 - u_{i,t+h,2}^2) - E(u_{i,t+h,1}^2 - u_{i,t+h,2}^2 \mid \mathcal{F}) + 2(\lambda_1' f_{t+h} u_{i,t+h,1} - \lambda_2' f_{t+h} u_{i,t+h,2})\right] \\
&\quad + n_k^{-1/2}\left(n_k^{-1}\sum_{i\in H_k} u_{i,t+h,2}\right)^2 - n_k^{-1/2}\left(n_k^{-1}\sum_{i\in H_k} u_{i,t+h,1}\right)^2.
\end{aligned}$$

By a CLT,

$$n_k^{-1/2}\left(n_k^{-1}\sum_{i\in H_k} u_{i,t+h,2}\right)^2 - n_k^{-1/2}\left(n_k^{-1}\sum_{i\in H_k} u_{i,t+h,1}\right)^2 = O_P(n_k^{-3/2}) = o_P(1).$$

Therefore, $\overline{\Delta u}^2_{t+h,k}$ is a $\sqrt{n_k}$-consistent estimator for $n_k^{-1}\sum_{i\in H_k} E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F})$. By the same CLT argument, it follows that

$$\begin{aligned}
&\sqrt{n_k}\left(\overline{\Delta u}^2_{t+h,k} - n_k^{-1}\sum_{i\in H_k} E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F})\right) \\
&= n_k^{-1/2}\sum_{i\in H_k}\left[(u^2_{i,t+h,1} - u^2_{i,t+h,2}) - E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F}) + 2(\lambda_1' f_{t+h} u_{i,t+h,1} - \lambda_2' f_{t+h} u_{i,t+h,2})\right] + o_P(1)
\end{aligned}$$

is asymptotically normal and that the variance of $\sqrt{n_k}\left(\overline{\Delta u}^2_{t+h,k} - n_k^{-1}\sum_{i\in H_k} E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F})\right)$ can be estimated by

$$\hat{V} := n_k^{-1}\sum_{i\in H_k}\left[(u^2_{i,t+h,1} - u^2_{i,t+h,2}) - E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F}) + 2(\lambda_1' f_{t+h} u_{i,t+h,1} - \lambda_2' f_{t+h} u_{i,t+h,2})\right]^2.$$

Recall our estimate $\bar{V} := n_k^{-1}\sum_{i\in H_k}(\Delta L_{i,t+h} - \overline{\Delta L}_{t+h,k})^2$ with $\overline{\Delta L}_{t+h,k} = n_k^{-1}\sum_{i\in H_k}\Delta L_{i,t+h}$. It remains to show that $\hat{V} = \bar{V} + o_P(1)$. By (8) and (9), we have

$$\begin{aligned}
&\Delta L_{i,t+h} - \overline{\Delta L}_{t+h,k} \\
&= \left[u^2_{i,t+h,1} - u^2_{i,t+h,2} + 2(\lambda_{i,1}' f_{t+h} u_{i,t+h,1} - \lambda_{i,2}' f_{t+h} u_{i,t+h,2})\right] \\
&\quad - n_k^{-1}\sum_{j\in H_k}\left[\left(u^2_{j,t+h,1} - u^2_{j,t+h,2}\right) + 2(\lambda_{j,1}' f_{t+h} u_{j,t+h,1} - \lambda_{j,2}' f_{t+h} u_{j,+ht,2})\right] \\
&= (u^2_{i,t+h,1} - u^2_{i,t+h,2}) - E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F}) + 2(\lambda_1' f_{t+h} u_{i,t+h,1} - \lambda_2' f_{t+h} u_{i,t+h,2}) + h_{n,1} + h_{n,2},
\end{aligned}$$

where $h_{n,1} = E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F}) - n_k^{-1}\sum_{j\in H_k}(u^2_{j,t+h,1} - u^2_{j,t+h,2})$ and $h_{n,2} = -2n_k^{-1}\sum_{j\in H_k}(\lambda_{j,1}' f_{t+h} u_{j,t+h,1} - \lambda_{j,2}' f_{t+h} u_{j,+ht,2})$. Clearly, by a LLN, $h_{n,1} = o_P(1)$ and $h_{n,2} = o_P(1)$. By the elementary inequality $\left|\sqrt{\sum(a_i+b_i)^2} - \sqrt{\sum a_i^2}\right| \le \sqrt{\sum b_i^2}$, we have that

$$\left|\sqrt{\hat{V}} - \sqrt{\bar{V}}\right| \leq \sqrt{n_k^{-1} \sum_{i \in H_k} (h_{n,1} + h_{n,2})^2} = |h_{n,1} + h_{n,2}| = o_P(1).$$

Thus, $\hat{V} = \bar{V} + o_P(1)$. The proof is complete. □

A.4 Corollary 1

Proof. The result follows once we notice that

$$\begin{aligned}
&\sqrt{n}\left(\sum_{k=1}^{K} \frac{n_k}{n} \overline{\Delta u}^2_{t+h,k} - n^{-1} \sum_{i=1}^{n} E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F})\right) \\
&= \sqrt{n} \sum_{k=1}^{K} \frac{n_k}{n} \left(\overline{\Delta u}^2_{t+h,k} - n_k^{-1} \sum_{i \in H_k} E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F})\right) \\
&= \sum_{k=1}^{K} \frac{n_k}{\sqrt{n}} \left\{ n_k^{-1} \sum_{i \in H_k} \left[(u^2_{i,t+h,1} - u^2_{i,t+h,2}) - E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F})\right] + O_P(n_k^{-1}) \right\} \\
&= n^{-1/2} \sum_{i=1}^{n} \left[(u^2_{i,t+h,1} - u^2_{i,t+h,2}) - E(u^2_{i,t+h,1} - u^2_{i,t+h,2} \mid \mathcal{F})\right] + O_P(n^{-1/2}).
\end{aligned}$$

□

A.5 Theorem 4

Proof. By equation (15), we have

$$\begin{aligned}
&\sqrt{n}\left[\sum_{k=1}^{K} \frac{n_k}{n} (\overline{e}^2_{1,k,t+h} - \overline{e}^2_{2,k,t+h}) - \sum_{k=1}^{K} \frac{n_k}{n} \left((\lambda'_{1,(k)} f_{t+h})^2 - (\lambda'_{2,(k)} f_{t+h})^2\right)\right] \\
&= n^{-1/2} \sum_{k=1}^{K} n_k \left\{ \left(n_k^{-1} \sum_{i \in H_k} u_{i,t+h,1}\right)^2 - \left(n_k^{-1} \sum_{i \in H_k} u_{i,t+h,2}\right)^2 \right\} \\
&\quad + 2n^{-1/2} \sum_{i=1}^{n} \left(\lambda'_{i,1} f_{t+h} u_{i,t+h,1} - \lambda'_{i,2} f_{t+h} u_{i,t+h,2}\right).
\end{aligned}$$

Again as in the proof of Theorem 3, we can show that

$n^{-1/2}\sum_{k=1}^{K} n_k \left\{\left(n_k^{-1}\sum_{i\in H_k} u_{i,t+h,1}\right)^2 - \left(n_k^{-1}\sum_{i\in H_k} u_{i,t+h,2}\right)^2\right\} = o_P(1)$, and so

$$\begin{aligned}\sqrt{n}&\left[\sum_{k=1}^{K}\frac{n_k}{n}(\bar{e}^2_{1,k,t+h}-\bar{e}^2_{2,k,t+h}) - n^{-1}\sum_{i=1}^{n}\left[(\lambda'_{i,1}f_{t+h})^2-(\lambda'_{i,2}f_{t+h})^2\right]\right]\\ &= 2n^{-1/2}\sum_{i=1}^{n}\left(\lambda'_{i,1}f_{t+h}u_{i,+ht,1}-\lambda'_{i,2}f_{t+h}u_{i,t+h,2}\right)+o_P(1).\end{aligned}$$

The rest of the proof follows by a CLT as in the proof of Theorem 3. □

A.6 Lemma 1

Proof. Since we can write $f_{s+h} = (\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}(\bar{e}_{s+h}-\bar{u}_{s+h})$, we have $e_{i,s+h} = \lambda'_i(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\bar{e}_{s+h} + u_{i,s+h} - \lambda'_i(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\bar{u}_{s+h}$. It is not difficult to see that

$$\begin{aligned}\hat{\lambda}'_i &= \left(\sum_{s+h=1}^{T}[\lambda'_i(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\bar{e}_{s+h}+u_{i,s+h}-\lambda'_i(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\bar{u}_{s+h}]\bar{e}'_{s+h}\right)\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{e}'_{s+h}\right)^{-1}\\ &= \lambda'_i(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda} + \left(\sum_{s+h=1}^{T}u_{i,s+h}\bar{e}'_{s+h}\right)\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{e}'_{s+h}\right)^{-1}\\ &\quad - \lambda'_i(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\left(\sum_{s+h=1}^{T}\bar{u}_{s+h}\bar{e}'_{s+h}\right)\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{e}'_{s+h}\right)^{-1}\end{aligned}$$

and thus

$$\hat{\lambda}'_i\bar{e}_{t+h} = \lambda'_i f_{t+h} + \xi_{i,t+h} + \varepsilon_{i,t+h} + \zeta_{i,t+h},$$

where $\xi_{i,t+h} = \lambda'_i(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\bar{u}_{t+h}$, $\varepsilon_{i,t+h} = \left(\sum_{s+h=1}^{T}u_{i,s+h}\bar{e}'_{s+h}\right)\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{e}'_{s+h}\right)^{-1}\bar{e}_{t+h}$ and $\zeta_{i,t+h} = -\lambda'_i(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\left(\sum_{s+h=1}^{T}\bar{u}_{s+h}\bar{e}'_{s+h}\right)\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{e}'_{s+h}\right)^{-1}\bar{e}_{t+h}$.

Next, observe that

$$\begin{aligned}&n^{-1/2}\sum_{i=1}^{n}[(\hat{\lambda}'_{i,1}\bar{e}_{t+h})^2-(\lambda'_{i,1}f_{t+h})^2]\\ &= n^{-1/2}\sum_{i=1}^{n}(\xi_{i,t+h,1}+\varepsilon_{i,t+h,1}+\zeta_{i,t+h,1})^2 + 2n^{-1/2}\sum_{i=1}^{n}(\xi_{i,t+h,1}+\varepsilon_{i,t+h,1}+\zeta_{i,t+h,1})\,\lambda'_{i,1}f_{t+h}.\end{aligned}$$

and

$$n^{-1/2}\sum_{i=1}^{n}\xi_{i,t+h,1}\lambda_{i,1}'f_{t+h}=n^{-1/2}\bar{u}_{t+h}'\bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\left(\sum_{i=1}^{n}\lambda_{i,1}\lambda_{i,1}'\right)f_{t+h}.$$

Therefore, we have

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n}[(\hat{\lambda}_{i,1}'\bar{e}_{t+h})^2-(\lambda_{i,1}'f_{t+h})^2]\\
&=n^{-1/2}\bar{u}_{t+h}'\bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\left(\sum_{i=1}^{n}\lambda_{i,1}\lambda_{i,1}'\right)f_{t+h}\\
&\qquad+n^{-1/2}\sum_{i=1}^{n}\left(\xi_{i,t+h,1}+\varepsilon_{i,t+h,1}+\zeta_{i,t+h,1}\right)^2+2n^{-1/2}\sum_{i=1}^{n}\left(\varepsilon_{i,t+h,1}+\zeta_{i,t+h,1}\right)\lambda_{i,1}'f_{t+h}.
\end{aligned}$$

The rest of the proof proceeds in four steps, bounding different components in the above display.

Step 1: show that $n^{-1/2}\sum_{i=1}^{n}\left(\varepsilon_{i,t+h,1}+\zeta_{i,t+h,1}\right)\lambda_{i,1}'f_{t+h}=o_P(1)$.

We observe that

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n}\zeta_{i,t+h,1}\lambda_{i,1}'f_{t+h}\\
&\quad=-n^{-1/2}\bar{e}_{t+h}'\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{e}_{s+h}'\right)^{-1}\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{u}_{s+h}'\right)\bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\left(\sum_{i=1}^{n}\lambda_{i,1}\lambda_{i,1}'\right)f_{t+h} \qquad (33)
\end{aligned}$$

and

$$n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i,t+h,1}\lambda_{i,1}'f_{t+h}=n^{-1/2}f_{t+h}'\left(\sum_{s+h=1}^{T}\left(\sum_{i=1}^{n}\lambda_{i,1}u_{i,s+h,1}\right)\bar{e}_{s+h}'\right)\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{e}_{s+h}'\right)^{-1}\bar{e}_{t+h}. \qquad (34)$$

Recall that $\bar{e}_{s+h}=\bar{\lambda}'f_{s+h}+\bar{u}_{s+h}$. Since $\bar{\lambda}\bar{\lambda}'$ has eigenvalues bounded away from zero and infinity and f_{s+h} has non-trivial variance, it follows that $\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{e}_{s+h}'\right)^{-1}=O_P(T^{-1})$. Moreover,

$$\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{u}_{s+h}'=\bar{\lambda}'\sum_{s+h=1}^{T}f_{s+h}\bar{u}_{s+h}'+\sum_{s+h=1}^{T}\bar{u}_{s+h}\bar{u}_{s+h}'. \qquad (35)$$

Notice that $\sum_{s+h=1}^{T}f_{s+h}\bar{u}_{s+h}'=n^{-1}\sum_{i=1}^{n}(\sum_{s+h=1}^{T}f_{s+h}u_{i,s+h}')$ and $T^{-1/2}\sum_{s+h=1}^{T}f_{s+h}u_{i,s+h}'$

has mean zero with bounded variance and is independent across i conditional on $\{f_{s+h}\}_{s+h=1}^T$. Therefore, $\sum_{s+h=1}^T f_{s+h}\bar{u}'_{s+h} = O_P(\sqrt{T/n})$. Since $E\sum_{s+h=1}^T \bar{u}_{s+h}\bar{u}'_{s+h} = O(T/n)$, we have $\sum_{s+h=1}^T \bar{u}_{s+h}\bar{u}'_{s+h} = O(T/n)$. By (35), we have $\sum_{s+h=1}^T \bar{e}_{s+h}\bar{u}'_{s+h} = O_P(\sqrt{T/n}+T/n)$. Therefore, (33) implies

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^n \zeta_{i,t+h,1}\lambda'_{i,1}f_{t+h} \\
&= -n^{-1/2}\bar{e}'_{t+h}\left(\sum_{s+h=1}^T \bar{e}_{s+h}\bar{e}'_{s+h}\right)^{-1}\left(\sum_{s+h=1}^T \bar{e}_{s+h}\bar{u}'_{s+h}\right)\bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\left(\sum_{i=1}^n \lambda_{i,1}\lambda'_{i,1}\right)f_{t+h} \\
&= n^{-1/2}O_P(1)\cdot O_P(T^{-1})\cdot O_P(\sqrt{T/n}+T/n)\cdot O_P(1)\cdot O_P(n)\cdot O_P(1) \\
&= O_P(n^{-1/2}+T^{-1/2}) = o_P(1). \qquad (36)
\end{aligned}
$$

We observe that

$$
\begin{aligned}
& \sum_{s+h=1}^T\left(\sum_{i=1}^n \lambda_{i,1}u_{i,s+h,1}\right)\bar{e}'_{s+h} \\
&= \sum_{s+h=1}^T\sum_{i=1}^n \lambda_{i,1}u_{i,s+h,1}\left(\bar{u}'_{s+h}+f'_{s+h}\bar{\lambda}\right) \\
&= \sum_{s+h=1}^T\sum_{i=1}^n \lambda_{i,1}u_{i,s+h,1}\bar{u}'_{s+h} + \sum_{s+h=1}^T\sum_{i=1}^n \lambda_{i,1}u_{i,s+h,1}f'_{s+h}\bar{\lambda} \\
&\overset{\text{(i)}}{=} \sum_{s+h=1}^T\sum_{i=1}^n \lambda_{i,1}u_{i,s+h,1}\bar{u}'_{s+h} + O_P(\sqrt{nT}) \\
&= O_P\left(\sqrt{\sum_{s+h=1}^T\left(\sum_{i=1}^n \lambda_{i,1}u_{i,s+h,1}\right)^2}\times\sqrt{\sum_{s+h=1}^T \|\bar{u}_{s+h}\|^2}\right) + O_P(\sqrt{nT}) \\
&\overset{\text{(ii)}}{=} O_P\left(\sqrt{Tn}\times\sqrt{T/n}\right) + O_P(\sqrt{nT}) = O_P\left(\sqrt{nT}(1+\sqrt{T/n})\right),
\end{aligned}
$$

where (i) and (ii) follow by Assumption 3. Hence, by (34), we have

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^n \varepsilon_{i,t+h,1}\lambda'_{i,1}f_{t+h} \\
&= n^{-1/2}f'_{t+h}\left(\sum_{s+h=1}^T\left(\sum_{i=1}^n \lambda_{i,1}u_{i,s+h,1}\right)\bar{e}'_{s+h}\right)\left(\sum_{s+h=1}^T \bar{e}_{s+h}\bar{e}'_{s+h}\right)^{-1}\bar{e}_{t+h}
\end{aligned}
$$

$$= n^{-1/2} \cdot O_P(1) \cdot O_P\left(\sqrt{nT}(1+\sqrt{T/n})\right) \cdot O_P(T^{-1}) \cdot O_P(1)$$

$$= O_P(T^{-1/2} + n^{-1/2}) = o_P(1). \tag{37}$$

By (36) and (37), we have proved the claim in Step 1.

Step 2: show that $n^{-1/2}\sum_{i=1}^n \xi_{i,t+h,1}^2 = o_P(1)$.

Clearly $E\bar{u}_{t+h}\bar{u}_{t+h}' = O(n^{-1})$ and thus $\bar{u}_{t+h}\bar{u}_{t+h}' = O_P(n^{-1})$. It follows that

$$\begin{aligned}
\sum_{i=1}^n \xi_{i,t+h}\xi_{i,t+h}' &= \sum_{i=1}^n \lambda_i'(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\bar{u}_{t+h}\bar{u}_{t+h}'\bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\lambda_i \\
&= \sum_{i=1}^n \text{trace } \left(\bar{u}_{t+h}\bar{u}_{t+h}'\bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\lambda_i\lambda_i'(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\right) \\
&= \text{trace } \left(\bar{u}_{t+h}\bar{u}_{t+h}'\bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\left[\sum_{i=1}^n \lambda_i\lambda_i'\right](\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\right) \\
&= \text{trace } \left(O_P(n^{-1}) \cdot O_P(1) \cdot O_P(n) \cdot O_P(1)\right) = O_P(1).
\end{aligned}$$

Therefore, $n^{-1/2}\sum_{i=1}^n \xi_{i,t+h,1}^2 = O_P(n^{-1/2}) = o_P(1)$.

Step 3: show that $n^{-1/2}\sum_{i=1}^n \varepsilon_{i,t+h,1}^2 = o_P(1)$.

Let $q_{n,1} = \left(\sum_{s+h=1}^T \bar{e}_{s+h}\bar{e}_{s+h}'\right)^{-1}\bar{e}_{t+h}$. Then

$$\varepsilon_{i,t+h} = \left(\sum_{s+h=1}^T u_{i,s+h}\bar{e}_{s+h}'\right) q_{n,1} = \sum_{s+h=1}^T u_{i,s+h}\bar{u}_{s+h}' q_{n,1} + \sum_{s+h=1}^T u_{i,s+h} f_{s+h}'\bar{\lambda} q_{n,1}.$$

Therefore,

$$n^{-1/2}\sum_{i=1}^n \|\varepsilon_{i,t+h}\|^2 \leq 2n^{-1/2}\sum_{i=1}^n \left\|\sum_{s+h=1}^T u_{i,s+h}\bar{u}_{s+h}' q_{n,1}\right\|^2 + 2n^{-1/2}\sum_{i=1}^n \left\|\sum_{s+h=1}^T u_{i,s+h} f_{s+h}'\bar{\lambda} q_{n,1}\right\|^2.$$

From the previous argument, $\|q_{n,1}\| = O_P(T^{-1})$. It follows that

$$n^{-1/2}\sum_{i=1}^n \|\varepsilon_{i,t+h}\|^2 = O_P\left(T^{-2}n^{-1/2}\left[\sum_{i=1}^n \left\|\sum_{s+h=1}^T u_{i,s+h}\bar{u}_{s+h}'\right\|^2 + \sum_{i=1}^n \left\|\sum_{s+h=1}^T u_{i,s+h} f_{s+h}'\right\|^2\right]\right). \tag{38}$$

We observe that

$$
\begin{aligned}
&\text{trace } \left[\sum_{i=1}^{n}\left(\sum_{\tau+h=1}^{T}\bar{u}_{\tau+h}u'_{i,\tau+h}\right)\left(\sum_{s+h=1}^{T}u_{i,s+h}\bar{u}'_{s+h}\right)\right]\\
&=\text{trace } \left[\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\bar{u}_{\tau+h}\left(\sum_{i=1}^{n}u'_{i,\tau+h}u_{i,s+h}\right)\bar{u}'_{s+h}\right]\\
&=\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\left(\sum_{i=1}^{n}u'_{i,\tau+h}u_{i,s+h}\right)\bar{u}'_{s+h}\bar{u}_{\tau+h}\\
&\leq\sqrt{\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\left(\sum_{i=1}^{n}u'_{i,\tau+h}u_{i,s+h}\right)^{2}}\times\sqrt{\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}(\bar{u}'_{s+h}\bar{u}_{\tau+h})^{2}}.
\end{aligned}
$$

Notice that

$$
\begin{aligned}
&E\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\left(\sum_{i=1}^{n}u'_{i,\tau+h}u_{i,s+h}\right)^{2}\\
&=\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}Eu'_{i_1,\tau+h}u_{i_1,s+h}u'_{i_2,\tau+h}u_{i_2,s+h}\\
&=\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\sum_{i=1}^{n}E(u'_{i,\tau+h}u_{i,s+h})^{2}+\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\sum_{i_1\neq i_2}Eu'_{i_1,\tau+h}u_{i_1,s+h}u'_{i_2,\tau+h}u_{i_2,s+h}\\
&\overset{\text{(i)}}{=}\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\sum_{i=1}^{n}E(u'_{i,\tau+h}u_{i,s+h})^{2}+\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\sum_{i_1\neq i_2}Eu'_{i_1,\tau+h}u_{i_1,s+h}Eu'_{i_2,\tau+h}u_{i_2,s+h}\\
&\leq\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\sum_{i=1}^{n}E(u'_{i,\tau+h}u_{i,s+h})^{2}+\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\left(\sum_{i=1}^{n}Eu'_{i,\tau+h}u_{i,s+h}\right)^{2}\\
&=O(nT^{2})+O\left(n^{2}\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\|\gamma_n(s,\tau)\|^{2}\right)\\
&\overset{\text{(ii)}}{=}O(nT^{2})+O\left(n^{2}\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\|\gamma_n(s,\tau)\|\right)\overset{\text{(iii)}}{=}O(nT^{2})+O(n^{2}T),
\end{aligned}
$$

where (i) follows by the independence of $u_{i,s}$ across i, (ii) follows by $\max_s \max_\tau \|\gamma_n(s,\tau)\| \leq M$ and (iii) follows by Assumption 4. On the other hand, we have

$$
\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}(\bar{u}'_{s+h}\bar{u}_{\tau+h})^{2}\leq\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\|\bar{u}_{s+h}\|^{2}\cdot\|\bar{u}_{\tau+h}\|^{2}=\left(\sum_{s+h=1}^{T}\|\bar{u}_{s+h}\|^{2}\right)^{2}=O_P(T^{2}n^{-1}).
$$

The above three displays imply that

$$\begin{aligned}
&\text{trace }\left[\sum_{i=1}^{n}\left(\sum_{\tau+h=1}^{T}\bar{u}_{\tau+h}u'_{i,\tau+h}\right)\left(\sum_{s+h=1}^{T}u_{i,s+h}\bar{u}'_{s+h}\right)\right]\\
&\leq\sqrt{\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}\left(\sum_{i=1}^{n}u'_{i,\tau+h}u_{i,s+h}\right)^2}\times\sqrt{\sum_{s+h=1}^{T}\sum_{\tau+h=1}^{T}(\bar{u}'_{s+h}\bar{u}_{\tau+h})^2}\\
&=\sqrt{O_P(nT^2)+O_P(n^2T)}\times\sqrt{O_P(T^2n^{-1})}=O_P\left(T^{3/2}(n^{1/2}+T^{1/2})\right).
\end{aligned}$$

Since $\sum_{i=1}^{n}\left(\sum_{\tau+h=1}^{T}\bar{e}_{\tau+h}u'_{i,\tau+h}\right)\left(\sum_{s+h=1}^{T}u_{i,s+h}\bar{e}'_{s+h}\right)$ is positive semi-definite, we have

$$\sum_{i=1}^{n}\left(\sum_{\tau+h=1}^{T}\bar{u}_{\tau+h}u'_{i,\tau+h}\right)\left(\sum_{s+h=1}^{T}u_{i,s+h}\bar{u}'_{s+h}\right)=O_P\left(T^{3/2}(n^{1/2}+T^{1/2})\right).$$

By a CLT, we have

$$\sum_{i=1}^{n}\left\|\sum_{s+h=1}^{T}u_{i,s+h}f'_{s+h}\right\|^2=O_P(nT).$$

The above two displays and (38) imply

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n}\|\varepsilon_{i,t+h}\|^2\\
&=O_P\left(T^{-2}n^{-1/2}\left[\sum_{i=1}^{n}\left\|\sum_{s+h=1}^{T}u_{i,s+h}\bar{u}'_{s+h}\right\|^2+\sum_{i=1}^{n}\left\|\sum_{s+h=1}^{T}u_{i,s+h}f'_{s+h}\right\|^2\right]\right)\\
&=O_P\left(T^{-2}n^{-1/2}\left[O_P\left(T^{3/2}(n^{1/2}+T^{1/2})\right)+O_P(nT)\right]\right)\\
&=O_P\left(n^{-1/2}+T^{-1/2}+n^{1/2}T^{-1}\right)\overset{\text{(i)}}{=}o_P(1),
\end{aligned}$$

where (i) follows by $n/T^2=o(1)$.

Step 4: show that $n^{-1/2}\sum_{i=1}^{n}\zeta_{i,t+h,1}^2=o_P(1)$.

Let $q_{n,2}=(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda}\left(\sum_{s+h=1}^{T}\bar{u}_{s+h}\bar{e}'_{s+h}\right)\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{e}'_{s+h}\right)^{-1}\bar{e}_{t+h}$. Then $\zeta_{i,t+h}=-\lambda'_i q_{n,2}$.

It follows that

$$n^{-1/2}\sum_{i=1}^{n}\|\zeta_{i,t+h}\|^2 = n^{-1/2}\sum_{i=1}^{n} q_{n,2}'\lambda_i\lambda_i' q_{n,2} \le O_P\left(n^{1/2}\|q_{n,2}\|^2\right). \tag{39}$$

By the previous argument, $\left(\sum_{s+h=1}^{T}\bar{e}_{s+h}\bar{e}_{s+h}'\right)^{-1}\bar{e}_{t+h} = O_P(T^{-1})$. Notice that

$$\sum_{s+h=1}^{T}\bar{u}_{s+h}\bar{e}_{s+h}' = \sum_{s+h=1}^{T}\bar{u}_{s+h}\bar{u}_{s+h}' + \sum_{s+h=1}^{T}\bar{u}_{s+h}f_{s+h}'\bar{\lambda}.$$

It is simple to show that $\sum_{s+h=1}^{T}\bar{u}_{s+h}\bar{u}_{s+h}' = O_P(T/n)$ and $\sum_{s+h=1}^{T}\bar{u}_{s+h}f_{s+h}'\bar{\lambda} = O_P(\sqrt{T/n})$. Therefore, $\|q_{n,2}\| = O_P\left(T/n + \sqrt{T/n}\right)\cdot O_P(T^{-1})$. By (39), we have

$$n^{-1/2}\sum_{i=1}^{n}\|\zeta_{i,t+h}\|^2 = O_P\left(n^{1/2}\|q_{n,2}\|^2\right) = O_P\left(n^{-3/2} + n^{-1/2}T^{-1}\right) = o_P(1).$$

This completes the proof. □

A.7 Theorem 5

Proof. By (12) and Lemma 1, we have that

$$\begin{aligned}
&\sqrt{n}\left[\overline{\Delta\hat{u}}_{t+h}^2 - n^{-1}\sum_{i=1}^{n}E(u_{i,t+h,1}^2 - u_{i,t+h,2}^2 \mid \mathcal{F})\right] \\
&= n^{-1/2}\sum_{i=1}^{n}\Big[(u_{i,t+h,1}^2 - u_{i,t+h,2}^2) - E(u_{i,t+h,1}^2 - u_{i,t+h,2}^2 \mid \mathcal{F}) \\
&\quad + 2(\lambda_{i,1}'f_{t+h}u_{i,t+h,1} - \lambda_{i,2}'f_{t+h}u_{i,t+h,2}) + u_{i,t+h}'D_{t+h}\Big] + o_P(1),
\end{aligned} \tag{40}$$

where $D_{t+h} = \bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\left(n^{-1}\sum_{i=1}^{n}[\lambda_{i,1}\lambda_{i,1}' - \lambda_{i,2}\lambda_{i,2}']\right)f_{t+h}$. Since $\hat{\lambda}_i' - \lambda_i'(\bar{\lambda}\bar{\lambda}')^{-1}\bar{\lambda} = o_P(1)$, $\bar{e}_{t+h} = \bar{\lambda}'f_{t+h} + o_P(1)$ and $(\bar{\lambda}\bar{\lambda}')^{-1}$ exists asymptotically, we have $\hat{D}_{t+h} = D_{t+h} + o_P(1)$. Since $\{u_{i,t+h,m}\}_{i=1}^{n}$ is independent across i, the result then follows by the classical CLT and a self-normalized CLT; see e.g., Theorem 4.1 of Chen et al. (2016); Peña et al. (2008). □

A.8 Theorem 6

Proof. By Lemma 1, we have that under the null hypothesis,

$$n^{-1/2}\sum_{i=1}^{n}\left[(\hat{\lambda}_{i,1}'\bar{e}_{t+h})^2-(\hat{\lambda}_{i,2}'\bar{e}_{t+h})^2\right]=2n^{-1/2}\bar{u}_{t+h}'\bar{\lambda}'(\bar{\lambda}\bar{\lambda}')^{-1}\sum_{i=1}^{n}\left(\lambda_{i,1}\lambda_{i,1}'-\lambda_{i,2}\lambda_{i,2}'\right)f_{t+h}+o_P(1).$$

Since $\{u_{i,t+h,m}\}_{i=1}^{n}$ is independent across i, the result then follows by the classical CLT and a self-normalized CLT; see e.g., Theorem 4.1 of Chen et al. (2016); Peña et al. (2008). □

A.9 Lemma 2

Proof. Under Assumptions A-F and Theorem 3 in Bai (2003), recall that the following result holds:

$$\hat{\lambda}_i'\hat{f}_{t+h}-\lambda_i'f_{t+h}=n^{-1}\lambda_i'\left(n^{-1}\sum_{j=1}^{n}\lambda_j\lambda_j'\right)^{-1}\sum_{j=1}^{n}\lambda_j u_{jt+h}$$
$$+T^{-1}f_{t+h}'\left(T^{-1}\sum_{s+h=1}^{T}f_{s+h}f_{s+h}'\right)^{-1}\sum_{s+h=1}^{T}f_{s+h}u_{is+h}+O_P(1/\min\{n,T\}).$$

Using this result, we have $\hat{\lambda}_{i,m}'f_{t+h}=\lambda_{i,m}'f_{t+h}+\xi_{i,t+h,m}$ for $m\in\{1,2\}$, where

$$\xi_{i,t+h,m}=n^{-1}\lambda_i'\left(n^{-1}\sum_{j=1}^{n}\lambda_j\lambda_j'\right)^{-1}\sum_{j=1}^{n}\lambda_{j,m}u_{jt+h,m}$$
$$+T^{-1}f_{t+h}'\left(T^{-1}\sum_{s+h=1}^{T}f_{s+h}f_{s+h}'\right)^{-1}\sum_{s+h=1}^{T}f_{s+h}u_{is+h,m}+O_P(1/\min\{n,T\}).$$

It follows that

$$n^{-1}\sum_{i=1}^{n}\left[(\hat{\lambda}_{i,1}'\hat{f}_{t+h})^2-(\hat{\lambda}_{i,2}'\hat{f}_{t+h})^2\right]=n^{-1}\sum_{i=1}^{n}[(\lambda_{i,1}'f_{t+h})^2-(\lambda_{i,2}'f_{t+h})^2]$$
$$+2n^{-1}\sum_{i=1}^{n}[\lambda_{i,1}'f_{t+h}\xi_{i,t+h,1}-\lambda_{i,2}'f_{t+h}\xi_{i,t+h,2}]$$
$$+n^{-1}\sum_{i=1}^{n}[\xi_{i,t+h,1}^2-\xi_{i,t+h,2}^2].$$

The last term is of order $1/\min\{n,T\}$, which is negligible if $\sqrt{n}/T = o(1)$. Under assumptions of weak (cross-sectional and serial) dependence in $u_{i,t+h,m}$ (e.g., Assumptions E and F in Bai (2003)), we can show that

$$n^{-1}\sum_{i=1}^{n}\lambda_{i,m}' f_{t+h}\xi_{i,t+h,m} = n^{-1}\sum_{i=1}^{n}\lambda_{i,m}' f_{t+h}u_{i,t+h,m} + o_P(n^{-1/2}).$$

Using this, it follows that

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\left[(\hat{\lambda}_{i,1}'\hat{f}_{t+h})^2 - (\hat{\lambda}_{i,2}'\hat{f}_{t+h})^2\right] &= n^{-1}\sum_{i=1}^{n}[(\lambda_{i,1}' f_{t+h})^2 - (\lambda_{i,2}' f_{t+h})^2] \\ &\quad + 2n^{-1}\sum_{i=1}^{n}[\lambda_{i,1}' f_{t+h}u_{i,t+h,1} - \lambda_{i,2}' f_{t+h}u_{i,t+h,2}] + o_P(n^{-1/2}). \end{aligned}$$

The stated result follows from this. □

Table 1: Firm coverage by forecaster

Panel A: Total number of firms covered

	Total	Consumer	Manufacturing	High tech	Health	Other
MERRILL	1825	233	382	409	159	642
JPMORGAN	1796	210	341	455	166	624
FBOSTON	1752	211	375	417	121	628
GOLDMAN	1602	193	358	387	123	541
MORGAN	1473	170	305	352	117	529
LAWRENCE	1437	151	297	372	113	504

Panel B: Average number of firms covered

	Total	Consumer	Manufacturing	High tech	Health	Other
MERRILL	356	45	81	76	29	125
JPMORGAN	311	38	79	72	27	96
FBOSTON	277	34	70	60	17	96
GOLDMAN	283	37	72	64	21	88
MORGAN	243	29	55	53	19	88
LAWRENCE	239	27	57	56	16	82

Note: Panel A reports the number of different firms whose quarterly earnings per share is predicted by each brokerage firm for at least one quarter during our sample. Panel B reports the average number of quarterly earnings per share forecasts generated by each brokerage firm both in the aggregate (first column) and across five industries (columns 2-6).

Table 2: Estimated number of common factors in the earnings forecast errors

	GR	ER	ED
FBOSTON	1	1	1
JPMORGAN	0	0	0
MORGAN	2	2	2
GOLDMAN	1	1	1
LAWRENCE	1	1	3
MERRILL	1	1	1

Note: This table presents estimates of the number of common factors in the earnings forecast errors using the methods in Ahn and Horenstein (2013) and Onatski (2010). Columns labeled "GR" and "ER" report the "'Growth Ratio" and "Eigenvalue Ratio" statistics proposed by Ahn and Horenstein (2013), while the column labeled "ED" reports the Onatski (2010) statistic.

Table 3: Correlations across earnings forecast errors

Average correlations in forecast errors						
No. factors	FBOSTON	JPMORGAN	MORGAN	GOLDMAN	LAWRENCE	MERRILL
0	0.09	0.08	0.07	0.08	0.12	0.08
	(40.22)	(43.50)	(25.12)	(37.41)	(39.02)	(57.97)
1	-0.01	0.00	0.03	0.00	-0.01	0.00
	(-2.62)	(0.34)	(9.95)	(1.08)	(-2.24)	(2.82)
2	-0.01	-0.00	0.01	0.01	-0.01	-0.00
	(-2.39)	(-1.87)	(5.07)	(4.41)	(-2.68)	(-0.71)

Note: This table reports estimates of the average pair-wise correlation in earnings forecast errors along with the test statistic for non-zero average correlations proposed by Pesaran (2004) in brackets underneath. Results are presented using raw forecast errors (row labeled "0") as well as residuals from a regression that accounts for one and two common factors in the residuals (rows labeled "1" and "2").

Table 4: Heterogeneity in factor loadings within and across industries

	Aggregate	consumer	manufacturing	high tech	health	other
FBOSTON	0.067	0.066	0.075	0.049	0.043	0.070
JPMORGAN	0.062	0.049	0.058	0.052	0.067	0.070
MORGAN	0.096	0.048	0.101	0.049	0.064	0.123
GOLDMAN	0.074	0.086	0.060	0.072	0.048	0.080
LAWRENCE	0.069	0.075	0.061	0.077	0.084	0.065
MERRILL	0.057	0.048	0.051	0.067	0.063	0.051

Note: This table reports the standard deviation of the estimated factor loadings for the earnings forecast errors across all firms (column 1) as well as for different industries (columns 2-6). For each set of forecast errors, we estimate a model with a single common factor on the normalized forecast errors, demeaned and scaled to have a unit sample variance.

$$\tilde{e}_{i,t} = \lambda_{i,1} f_{1,t} + \varepsilon_{i,t},$$

sucject to the constraint: $\sum_i^N \lambda_{i,1}^2 = 1$. The table reports the standard deviation of the factor loadings $\lambda_{i,1}$ within each group of firms.

Table 5: Contributions of idiosyncratic error variance and squared bias components

Difference in idiosyncratic variance (%)	MORGAN vs. GOLDMAN		MORGAN vs. MERRILL		GOLDMAN vs. MERRILL		LAWRENCE vs. MERRILL	
	Mean Ratio	Variance Ratio	Mean Ratio	Variance Ratio	Mean Ratio	Variance Ratio	Mean Ratio	Variance Ratio
PCA	39.85	86.97	55.23	61.75	64.10	51.30	74.82	98.34
CCE	1.31	31.91	51.57	63.01	94.79	105.98	43.34	76.81
Cluster	96.46	93.58	96.35	92.91	93.40	90.74	85.46	89.45

Difference in squared bias (%)	MORGAN vs. GOLDMAN		MORGAN vs. MERRILL		GOLDMAN vs. MERRILL		LAWRENCE vs. MERRILL	
	Mean Ratio	Variance Ratio	Mean Ratio	Variance Ratio	Mean Ratio	Variance Ratio	Mean Ratio	Variance Ratio
PCA	60.15	7.70	44.77	16.41	35.90	31.67	25.18	6.07
CCE	98.69	64.37	48.43	15.71	5.21	6.94	56.66	7.74
Cluster	3.54	0.35	3.65	0.64	6.60	0.44	14.20	0.42

Note: Columns labeled mean ratio report the sample average of the ratio of the mean contribution to the total loss difference that comes from differences in idiosyncratic variances (top panel) or differences in squared biases (bottom panel) for a given pair of brokerage firms. Columns labeled variance ratio report the ratio of the sample variance of the squared idiosyncratic error differences to the sample variance of the total loss difference (top panel) or the ratio of the variance of the squared bias difference to the variance of the total loss difference (bottom panel), averaged across all quarters in the sample.

Table 6: Coverage probabilities for 95% confidence intervals constructed to test the null of equal conditional squared error loss

Coverage probability							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	91.6	91.6	91.1	90.8	90.9	90.4	89.8
25	93.8	94.0	93.4	93.4	93.3	93.3	92.9
50	94.5	94.3	94.3	94.1	94.0	94.4	94.2
100	94.5	94.7	94.5	94.5	94.6	94.4	94.4
200	94.8	94.9	94.8	94.8	94.7	94.7	94.8
1000	94.9	95.2	95.1	95.0	95.1	94.8	95.2

Note: This table reports the coverage probability for a 95% confidence interval for the test of equal conditional squared error loss, $E(\overline{\Delta L}_{t+h} \mid \mathcal{F}) = 0$ using the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts.

Table 7: Coverage probabilities for 95% confidence intervals constructed to test the null of equal squared biases

Clustering							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	98.0	96.6	93.8	92.7	92.5	90.8	91.8
25	97.3	96.6	94.4	94.1	93.1	93.2	93.4
50	96.3	95.7	95.6	93.9	95.2	94.1	93.7
100	95.9	95.3	94.9	94.9	94.9	94.1	94.3
200	95.9	94.4	93.8	95.4	95.5	95.3	95.0
1000	95.4	94.8	94.8	95.4	95.1	95.5	95.1

CCE							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	97.7	96.2	93.6	92.1	92.5	90.7	91.4
25	97.1	96.1	94.2	93.9	92.9	93.0	93.2
50	96.3	95.4	95.3	93.6	95.0	93.6	93.1
100	95.7	95.3	94.6	94.8	94.4	93.9	93.8
200	95.3	93.9	93.6	94.9	95.2	95.1	94.5
1000	94.7	94.2	94.7	95.1	94.7	95.2	94.8

PCA							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	99.1	96.9	92.2	91.4	91.9	90.4	91.0
25	97.1	92.6	92.4	93.1	92.5	92.9	93.2
50	95.0	91.7	94.2	92.0	94.6	93.3	93.2
100	90.6	92.7	93.3	94.1	93.6	93.4	93.7
200	90.2	92.5	93.6	93.4	94.8	95.2	94.9
1000	91.3	92.6	93.6	94.3	93.8	94.6	94.9

Note: This table reports the coverage probability for a 95% confidence interval for the test of equal squared biases, using the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts. We show coverage probabilities for the clustering, CCE, and PCA methods described in Section 3. The assumed time-series dimension is $T = 80$.

Table 8: Coverage probabilities for 95% confidence intervals constructed to test the null of equal idiosyncratic error variances

Coverage probability (clustering)							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	93.7	94.8	95.4	96.4	95.6	98.0	99.1
25	94.8	94.6	97.4	97.3	97.8	98.5	99.4
50	95.3	96.3	97.0	97.3	98.1	99.0	99.6
100	95.3	96.6	97.3	97.8	98.6	99.2	99.8
200	95.8	96.6	97.7	98.6	98.2	99.0	99.4
1000	95.7	96.4	98.0	97.8	98.3	99.5	99.7

Coverage probability (CCE)							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	95.3	96.5	97.0	98.0	97.1	99.2	99.5
25	96.1	96.4	98.4	98.5	98.5	99.3	99.7
50	96.4	97.3	97.9	98.3	99.0	99.6	99.8
100	96.5	97.1	98.4	98.7	99.1	99.7	100.0
200	96.4	97.6	98.5	99.1	98.9	99.4	99.7
1000	96.4	97.1	98.7	98.6	99.1	99.8	99.9

Coverage probability (PCA)							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	91.3	91.0	93.1	94.1	95.4	98.0	99.1
25	94.1	92.8	96.1	96.8	97.4	98.6	99.3
50	94.6	94.5	96.5	97.0	97.9	99.0	99.6
100	93.3	95.5	96.3	97.5	98.4	99.1	99.7
200	95.0	95.8	96.8	98.4	98.0	99.1	99.4
1000	94.8	95.3	97.6	97.1	98.0	99.4	99.7

Note: This table reports the coverage probability for a 95% confidence interval for the test of equal idiosyncratic variances, using the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts. We show coverage probabilities for the clustering, CCE, and PCA methods described in Section 3. The assumed time-series dimension is $T = 80$.

Figure 1: Cross-sectional test statistics for comparisons of the null of equal squared error loss conducted on pairs of brokerage firms

Positive values of the test statistics indicate that the second forecaster is more accurate than the first forecaster, while negative values suggest the reverse.

Figure 2: Values of the cross-sectional test of equal idiosyncratic variances conducted on individual quarters

Each panel shows the outcome of a cross-sectional test of the null that a pair of forecasters produce the same idiosyncratic error variance in a given quarter. Red color indicates that the idiosyncratic error variance component of the first forecaster is larger than that of the second forecaster. Blue color indicates the reverse. The first and second rows of each panel estimate the factors by PCA and CCE, respectively, while the third row is calculated by assuming identical factor loadings within each cluster. Asterisks represent quarters with test statistics that are statistically significant at the 5% level.

Figure 3: Values of the cross-sectional test of equal squared biases conducted on individual quarters

Each panel shows the outcome of a cross-sectional test of the null that a pair of forecasters produce the same squared bias in a given quarter. Red color indicates that the squared bias component of the first forecaster is larger than that of the second forecaster. Blue color indicates the reverse. The first and second rows of each panel estimate the factors by PCA and CCE, respectively, while the third row is calculated by assuming identical factor loadings within each cluster. Asterisks represent quarters with test statistics that are statistically significant at the 5% level.

Table A1: Coverage probabilities for 95% confidence intervals constructed to test the null of equal squared biases (5-cluster DGP)

	clustering						
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	88.1	86.6	84.2	83.8	81.8	78.5	77.1
25	97.3	96.0	95.0	94.2	94.2	91.6	90.6
50	98.2	97.5	96.1	95.9	95.0	94.3	93.9
100	98.2	97.4	96.8	95.6	95.0	94.5	94.6
200	98.4	96.5	95.8	95.3	94.9	94.6	95.1
1000	96.9	96.2	96.1	95.8	95.5	95.8	95.3

	CCE						
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	98.5	97.5	95.2	94.6	93.6	90.7	90.6
25	98.6	97.1	95.7	95.5	93.3	93.1	91.2
50	98.4	96.6	94.9	94.5	93.4	92.9	93.3
100	96.6	95.6	94.5	93.7	93.5	93.8	93.2
200	95.7	93.3	92.9	92.8	92.7	92.4	93.9
1000	94.3	93.0	93.5	93.2	93.1	93.7	93.4

	PCA						
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	99.7	99.4	98.7	99.5	99.7	99.8	99.9
25	99.7	99.2	99.9	99.9	99.8	100.0	100.0
50	99.2	99.7	100.0	100.0	99.9	100.0	99.9
100	98.8	99.7	99.9	100.0	100.0	100.0	100.0
200	99.7	99.8	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: This table reports the coverage probability for a 95% confidence interval for the test of equal squared biases, using the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts. We show coverage probabilities for the clustering, CCE, and PCA methods described in Section 3. The assumed time-series dimension is $T = 80$. The underlying data generating process assumes factor loadings that follow a cluster structure with 5 clusters and $n/5$ elements in each cluster.

Table A2: Coverage probabilities for 95% confidence intervals constructed to test the null of equal idiosyncratic error variances (5-cluster DGP)

	clustering						
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	98.0	98.0	98.4	98.8	99.0	99.5	99.8
25	96.4	97.1	98.3	98.8	98.9	99.8	99.7
50	96.4	96.5	97.6	98.7	99.1	99.8	99.8
100	96.8	96.6	97.9	98.4	99.0	99.6	99.8
200	96.4	97.1	98.5	98.6	98.9	99.1	99.7
1000	95.7	96.7	97.9	97.9	98.7	99.4	99.5

	CCE						
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	95.8	95.4	96.6	96.6	97.0	98.4	98.9
25	95.4	96.8	97.4	97.5	97.7	98.8	99.3
50	95.8	95.9	96.6	97.6	98.3	99.3	99.4
100	97.0	96.6	97.5	98.2	98.1	99.0	99.6
200	95.8	96.2	97.5	98.5	98.4	98.4	99.2
1000	95.7	96.7	97.9	97.0	98.3	98.9	98.9

	PCA						
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	92.3	91.2	91.3	92.9	93.2	92.3	92.3
25	93.1	94.4	94.9	94.7	94.2	94.4	95.2
50	94.2	93.6	94.2	95.2	95.5	96.0	95.1
100	95.6	94.8	95.4	95.5	95.5	95.6	95.6
200	94.8	95.5	95.8	96.1	95.8	94.8	95.7
1000	95.1	94.8	95.8	94.1	96.4	95.3	95.7

Note: This table reports the coverage probabilities for 95% confidence intervals for the test of equal idiosyncratic variances using the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts. We show coverage probabilities for the clustering, CCE, and PCA methods described in Section 3. The assumed time-series dimension is $T = 80$. The underlying data generating process assumes that factor loadings follow a cluster structure with 5 clusters and $n/5$ elements in each cluster.

Table A3: Coverage probability of 95% confidence intervals: 2 factors with heterogeneous loadings

CCE: squared bias							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	98.2	95.4	93.2	94.0	92.1	92.0	91.8
25	97.7	94.7	93.1	94.2	93.7	94.3	93.8
50	95.4	94.2	92.5	93.5	94.6	94.3	94.7
100	94.7	94.3	93.7	94.8	95.1	94.6	94.2
200	92.9	94.2	94.5	93.4	94.4	95.0	95.0
1000	93.7	94.2	94.8	93.9	94.0	95.4	94.1

PCA: squared bias							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	97.3	95.1	89.9	88.4	86.0	85.7	84.6
25	95.7	91.9	89.4	90.7	90.5	90.1	90.0
50	93.6	88.7	89.0	91.1	93.0	92.1	93.0
100	90.1	89.5	91.8	92.7	93.1	93.7	93.8
200	87.5	89.7	93.0	93.5	93.5	94.7	94.5
1000	87.3	91.5	93.0	94.7	92.9	94.6	94.7

CCE: variance							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	95.5	96.0	97.1	96.3	96.6	97.8	98.9
25	96.2	96.8	97.3	98.0	97.6	98.5	98.9
50	95.6	96.4	97.5	98.0	98.1	99.1	99.3
100	96.5	96.7	97.7	97.9	98.6	98.9	99.5
200	96.3	96.2	97.5	97.8	98.9	99.4	99.4
1000	95.7	96.5	97.7	98.2	98.3	99.3	99.4

PCA: variance							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	92.4	93.0	92.3	93.8	94.3	96.7	98.5
25	94.3	93.7	94.9	95.8	96.0	98.0	99.0
50	93.4	94.5	96.0	96.7	97.5	98.6	99.3
100	94.8	94.9	96.1	97.4	97.9	98.7	99.2
200	94.6	94.7	96.0	97.3	97.4	99.4	99.4
1000	94.1	95.4	96.5	97.2	97.9	98.9	99.2

Note: This table reports the coverage probability for a 95% confidence interval for the test of equal squared biases (top two panels) or equal idiosyncratic variances (panels 3 and 4), using the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts. We show coverage probabilities for the CCE and PCA methods described in Section 3. The assumed time-series dimension is $T = 80$. The underlying data generating process assumes two factors.

Table A4: Coverage probability of 95% confidence intervals: 3 factors with heterogeneous loadings

	CCE method for difference in squared bias						
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	96.8	95.2	91.9	92.7	91.1	89.1	88.9
25	96.5	94.4	92.8	90.3	91.7	88.9	85.9
50	95.2	94.0	90.6	90.9	88.4	84.6	77.6
100	92.8	92.3	88.4	88.1	84.6	75.1	66.1
200	93.0	89.4	84.8	81.1	75.5	63.9	53.8
1000	90.6	84.6	74.9	68.7	64.0	52.2	40.4

	PCA method for difference in squared bias						
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	98.1	96.9	92.7	91.3	88.7	88.3	88.0
25	98.2	95.1	93.2	93.0	93.3	92.4	93.9
50	96.0	94.3	93.1	92.2	92.9	95.0	94.4
100	94.3	91.9	94.2	94.8	94.8	95.1	94.6
200	92.1	93.3	94.0	95.6	94.7	95.0	95.4
1000	91.1	92.7	96.0	94.7	95.7	95.9	96.1

	CCE method for difference in variance						
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	94.5	97.1	96.3	97.5	97.5	98.3	99.0
25	96.3	97.5	97.7	98.0	97.4	98.5	98.8
50	96.5	96.4	98.4	97.5	98.1	98.3	97.6
100	97.2	96.3	97.8	97.8	96.7	94.6	90.5
200	95.4	97.1	97.4	95.7	94.9	86.5	77.6
1000	95.5	94.9	94.1	90.5	86.1	74.0	59.0

	PCA method for difference in variance						
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	92.1	92.8	91.3	93.6	93.8	96.1	97.5
25	93.7	94.4	94.6	95.9	95.9	98.0	98.8
50	94.4	93.9	96.9	96.0	97.0	98.4	98.7
100	95.1	94.2	96.7	97.1	96.7	99.1	98.6
200	93.5	95.2	96.7	97.6	98.3	98.9	98.9
1000	93.8	94.6	97.1	96.6	97.2	98.6	99.0

Note: This table reports the coverage probability for a 95% confidence interval for the test of equal squared biases (top two panels) or equal idiosyncratic variances (panels 3 and 4), using the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts. We show coverage probabilities for the CCE and PCA methods described in Section 3. The assumed time-series dimension is $T = 80$. The underlying data generating process assumes three factors.

Table A5: Coverage probability of 95% confidence intervals: Breaks in the number of factors with heterogeneous loadings

	CCE method							PCA method						
	difference in squared bias (before the break)													
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	97.4	95.6	93.5	91.9	90.9	89.5	89.4	99.6	98.2	94.5	91.1	89.9	85.8	85.4
25	97.2	95.6	92.1	92.2	91.3	88.7	88.3	99.2	96.6	93.0	92.0	92.3	89.8	91.1
50	95.2	91.7	90.6	90.1	90.3	87.1	85.6	98.0	95.0	93.4	93.1	92.5	91.7	92.0
100	94.1	91.1	89.8	89.9	87.8	84.6	76.7	95.1	94.4	93.6	92.5	94.0	94.0	93.3
200	93.5	92.0	88.7	85.6	85.7	77.6	65.7	94.3	93.7	93.8	93.3	94.0	94.8	94.2
1000	90.9	91.0	83.7	80.2	77.2	62.6	47.4	92.5	93.1	93.8	94.2	95.3	95.1	94.5
	diff in squared bias (after the break)													
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	96.8	95.9	93.2	91.9	90.7	90.4	90.2	99.5	97.3	93.0	90.7	88.6	88.8	87.3
25	96.9	93.9	93.0	92.5	91.8	90.6	86.9	98.6	95.6	94.5	93.1	93.6	93.4	92.4
50	94.5	93.8	91.7	90.9	90.6	84.4	79.8	97.4	93.7	93.9	94.0	95.2	93.9	94.1
100	94.2	92.2	89.9	87.0	86.4	76.0	67.4	96.0	95.3	94.6	94.8	95.8	95.4	95.5
200	92.7	90.9	85.2	81.9	76.9	65.0	54.3	94.1	94.6	94.5	94.9	95.7	96.0	96.2
1000	90.7	85.2	73.7	68.2	64.1	47.8	39.8	94.0	95.2	96.0	95.5	95.0	96.2	96.2
	diff in variance (before the break)													
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	95.4	95.6	95.6	96.8	97.2	98.3	99.0	91.8	91.6	92.3	92.9	94.4	97.3	98.4
25	95.5	96.4	97.4	97.5	98.2	98.7	99.4	93.3	94.6	95.4	96.0	97.5	98.5	99.5
50	95.8	95.8	97.4	98.3	98.7	99.2	99.6	94.4	94.1	96.1	97.2	97.7	99.2	99.5
100	96.2	96.9	97.2	98.0	98.4	98.4	96.8	94.9	94.8	96.2	97.1	98.7	99.3	99.2
200	96.3	96.2	96.9	97.6	98.1	95.7	88.5	94.7	95.2	96.9	97.9	98.5	99.0	99.8
1000	96.0	96.7	95.4	95.6	93.9	83.6	69.2	94.6	95.9	96.3	96.9	98.4	99.6	99.2
	diff in variance (after the break)													
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	95.3	96.0	96.7	97.2	97.6	98.9	99.2	91.8	91.7	92.3	93.6	94.4	96.9	98.0
25	96.4	96.2	98.3	98.3	98.7	99.1	98.7	93.1	92.8	96.0	95.5	96.7	98.6	98.2
50	96.1	97.7	98.0	98.0	98.2	97.9	98.2	93.8	94.4	95.7	96.6	97.2	98.5	99.3
100	97.2	97.0	97.7	97.4	96.7	94.6	89.9	94.8	95.2	96.7	97.2	97.4	98.6	99.0
200	95.9	97.2	96.3	96.4	94.2	87.9	74.8	94.4	95.4	96.5	97.4	97.6	99.0	99.0
1000	96.2	96.5	93.5	89.3	84.9	67.9	56.0	94.9	95.6	96.7	97.1	97.4	98.9	99.3

Note: This table reports the coverage probability for a 95% confidence interval for the test of equal squared biases (top two panels) or equal idiosyncratic variances (panels 3 and 4), using the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts. We show coverage probabilities for the CCE and PCA methods described in Section 3. The assumed time-series dimension is $T = 80$. The underlying data generating process assumes that there are initially two factors but that this changes to three factors in the second half of the sample.

Table A6: 95% Coverage probabilities for a 95% confidence interval for testing the null of equal conditionally expected loss under Linex loss

Coverage probability (linex loss)							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	89.5	90.5	89.2	88.9	89.0	88.1	87.8
25	92.6	93.0	92.9	92.5	93.1	92.1	90.5
50	95.4	94.5	94.3	94.2	93.8	93.0	92.3
100	95.1	94.2	94.3	94.1	94.2	94.0	93.6
200	95.0	94.9	94.5	95.0	95.1	94.9	94.1
1000	94.3	95.4	95.3	94.8	94.7	95.0	95.8

Note: This table reports the coverage probability for a 95% confidence interval for the test of equal expected loss, $E(\overline{\Delta L}_{t+h} \mid \mathcal{F}) = 0$, using the linex loss function. We use the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts.

Table A7: Coverage probabilities for a 95% confidence interval for the average difference in squared bias under conditionally heteroskedastic shocks

clustering							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	97.9	96.3	94.2	93.7	93.0	91.0	91.3
25	98.2	96.1	94.9	94.1	93.6	92.6	94.2
50	96.1	95.7	95.0	94.0	94.8	94.5	94.7
100	96.0	96.0	94.5	94.6	95.3	94.3	94.1
200	96.1	95.8	94.3	94.1	95.4	95.1	95.0
1000	95.3	94.8	95.1	94.5	94.5	94.5	95.7

CCE							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	97.8	95.9	94.0	93.3	92.0	90.2	91.2
25	98.3	96.0	94.6	94.1	93.2	92.2	93.9
50	96.1	95.2	94.8	93.5	94.6	94.0	94.3
100	95.7	95.8	94.3	94.4	95.1	94.2	93.8
200	95.8	95.2	94.1	93.8	94.9	94.6	94.8
1000	94.9	94.7	94.8	94.2	94.1	94.1	95.3

PCA							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	97.7	94.8	90.1	89.2	90.2	89.8	90.8
25	96.1	91.4	91.7	91.8	91.9	92.1	94.1
50	92.1	88.4	91.6	93.1	94.1	93.6	94.5
100	89.8	89.3	91.2	93.3	94.0	94.1	93.6
200	87.0	89.9	93.7	92.9	94.1	94.4	94.8
1000	86.7	91.0	93.9	93.4	94.0	93.8	95.2

Note: This table reports coverage probabilities for 95% confidence intervals for the test of equal squared biases using the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts. We show coverage probabilities for the clustering, CCE, and PCA methods described in Section 3. The assumed time-series dimension is $T = 80$. The table replaces the assumption of i.i.d standard normal errors and factors with an assumption of ARCH dynamics.

Table A8: Coverage probabilities for a 95% confidence interval for the average difference in variance under conditionally heteroskedastic shocks

clustering							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	94.3	94.4	95.7	95.8	96.5	97.9	98.3
25	95.1	96.4	96.8	96.4	98.1	99.2	99.5
50	95.4	96.5	97.5	98.0	98.0	98.7	99.6
100	95.6	96.7	97.3	96.9	98.1	99.0	99.7
200	95.5	96.2	96.4	97.4	97.7	98.9	99.6
1000	94.5	96.2	97.1	97.6	98.3	98.9	99.0

CCE							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	95.8	96.2	97.1	97.0	97.9	98.9	99.3
25	95.8	97.0	98.1	97.6	98.9	99.6	99.7
50	96.3	97.5	98.2	98.4	98.5	99.2	99.8
100	96.1	97.1	98.0	98.1	99.1	99.6	99.8
200	96.4	96.9	97.6	98.3	98.6	99.3	99.9
1000	95.1	96.9	98.1	98.4	98.7	99.5	99.6

PCA							
$n \setminus \rho^2$	0.05	0.1	0.2	0.25	0.3	0.4	0.45
10	93.4	92.7	92.1	93.1	95.1	97.4	98.1
25	93.1	93.4	94.9	95.3	97.7	99.3	99.2
50	94.0	94.4	96.9	97.2	97.7	98.6	99.6
100	94.1	95.0	96.4	96.6	97.8	99.1	99.7
200	94.4	94.5	95.7	97.2	97.8	98.9	99.6
1000	93.8	94.6	96.5	97.3	97.9	98.8	98.9

Note: This table reports the coverage probability for 95% confidence intervals for the test of equal idiosyncratic error variances, using the Monte Carlo simulation setup described in Section 5.1 and 2,000 random samples. n refers to the number of cross-sectional units used in the pair-wise comparison of loss differences, while ρ^2 measures the predictive power of the underlying forecasts. We show coverage probabilities for the clustering, CCE, and PCA methods described in Section 3. The assumed time-series dimension is $T = 80$. The table replaces the assumption of i.i.d standard normal errors and factors with an assumption of ARCH dynamics.

Table A9: Expected length of 95% confidence intervals

	squared bias							variance						
	clustering													
$n \setminus \rho_e^2$	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	3.84	3.35	2.93	2.81	2.40	2.27	2.08	4.95	4.23	3.66	3.44	2.94	2.68	2.43
25	2.42	2.15	1.89	1.69	1.55	1.53	1.40	3.25	2.86	2.46	2.18	1.93	1.91	1.68
50	1.72	1.48	1.33	1.29	1.13	1.13	1.05	2.37	2.00	1.77	1.67	1.45	1.41	1.28
100	1.22	1.07	0.96	0.87	0.85	0.76	0.72	1.70	1.47	1.27	1.14	1.10	0.97	0.90
200	0.88	0.76	0.69	0.61	0.58	0.55	0.50	1.22	1.04	0.92	0.80	0.74	0.70	0.63
1000	0.55	0.47	0.42	0.40	0.37	0.35	0.33	0.77	0.65	0.57	0.53	0.48	0.44	0.41
	CCE													
$n \setminus \rho_e^2$	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	4.28	3.74	3.17	2.79	2.31	1.83	1.31	6.00	4.92	4.03	3.43	2.75	2.12	1.46
25	3.20	2.77	2.30	1.86	1.53	1.22	0.82	4.07	3.41	2.78	2.21	1.79	1.42	0.95
50	2.45	1.96	1.65	1.42	1.14	0.90	0.60	3.02	2.38	1.97	1.67	1.33	1.05	0.70
100	1.78	1.46	1.23	0.97	0.81	0.62	0.42	2.17	1.75	1.45	1.16	0.95	0.72	0.49
200	1.32	1.06	0.87	0.69	0.58	0.44	0.29	1.57	1.25	1.03	0.81	0.67	0.52	0.34
1000	0.83	0.66	0.54	0.45	0.36	0.28	0.18	0.99	0.79	0.64	0.53	0.43	0.33	0.22
	PCA													
$n \setminus \rho_e^2$	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	3.50	2.89	2.41	2.03	1.59	1.24	0.81	3.99	3.13	2.52	2.08	1.63	1.25	0.82
25	2.58	2.12	1.73	1.39	1.12	0.88	0.58	2.67	2.18	1.76	1.41	1.12	0.88	0.58
50	1.94	1.53	1.26	1.07	0.84	0.66	0.44	1.98	1.55	1.27	1.07	0.85	0.67	0.44
100	1.41	1.14	0.93	0.75	0.61	0.47	0.32	1.42	1.14	0.93	0.75	0.62	0.47	0.32
200	1.01	0.81	0.67	0.53	0.43	0.33	0.22	1.02	0.81	0.67	0.53	0.43	0.34	0.22
1000	0.65	0.51	0.42	0.35	0.28	0.22	0.14	0.65	0.52	0.42	0.35	0.28	0.22	0.14

Note: This table assumes a three-factor data generating process with random factor loadings. Parameters are set to match the average cross-sectional R^2 value observed in the empirical application.