Journal of Econometrics 198 (2017) 231–252

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Tests of equal accuracy for nested models with estimated factors[☆]

Sílvia Gonçalves[a], Michael W. McCracken[b], Benoit Perron[c,*]

[a] University of Western Ontario, Economics Department, SSC 4058, 1151 Richmond Street N. London, Ontario, N6A 5C2, Canada

[b] Federal Reserve Bank of St. Louis, P.O. Box 442. St. Louis, MO 63166, USA

[c] Département de sciences économiques, CIREQ and CIRANO, Université de Montréal, C.P.6128, succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada

ARTICLE INFO

Article history:
Received 14 September 2015
Received in revised form
31 August 2016
Accepted 22 January 2017
Available online 21 February 2017

JEL classification:
C12
C32
C38
C52

Keywords:
Factor model
Out-of-sample forecasts
Recursive estimation

ABSTRACT

In this paper we develop asymptotics for tests of equal predictive ability between nested models when factor-augmented regressions are used to forecast. We provide conditions under which the estimation of the factors does not affect the asymptotic distributions developed in Clark and McCracken (2001) and McCracken (2007). This enables researchers to use the existing tabulated critical values when conducting inference despite the presence of estimated predictors. As an intermediate result, we derive the asymptotic properties of the principal components estimator over recursive windows. We provide simulation evidence on the finite sample effects of factor estimation and apply the tests to the case of forecasting excess returns to the S&P 500 Composite Index.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Factors are now a common, parsimonious approach to incorporating many potential predictors into a predictive regression model. Examples that estimate factors using a large collection of macroeconomic variables include works by Stock and Watson (2002), who use factors to predict U.S. inflation and industrial production; Ludvigson and Ng (2009), who use factors to predict bond risk premia; Hofmann (2009), who uses factors to predict euro-area inflation; Shintani (2005), who uses factors to predict Japanese inflation; and Ciccarelli and Mojon (2010), who use factors to predict global inflation. Examples that estimate factors using a large collection of financial variables include works by Stock and Watson (2003), who use factors to predict a variety of macroeconomic variables; Kelly and Pruitt (2013), who use factors to predict stock returns; and Andreou et al. (2013), who use factors to predict U.S. real GDP growth.

In many of these examples, the predictive content of the factors is evaluated using pseudo out-of-sample methods. Specifically, the usefulness of the factors is evaluated by comparing the mean squared forecast error (MSE) of a model that includes factors with one that does not include factors. Typically, the comparison is between a baseline autoregressive model and a comparable model that has been augmented with estimated factors. As such, the models being compared are nested under the null hypothesis of equal accuracy and the theoretical results in West (1996), based on non-nested comparisons, may not apply. McCracken (2007) and Clark and McCracken (2001, 2005) show that, for forecasts from nested models, the distributions of tests for equal forecast accuracy and encompassing are usually not asymptotically normal or chi-square. For one-step-ahead forecasts from models with conditionally homoskedastic errors, tables are provided in Clark and McCracken (2001) and McCracken (2007). For longer horizons or when conditional heteroskedasticity is present, Clark and McCracken (2012)

☆ We thank participants at the European meeting of the Econometric society 2016 (Geneva), the Society for Financial Econometrics 2016 (Hong Kong), the annual meeting of the International Association for Applied Econometrics 2016 (Milan), the Midwest Econometrics Group 2016 (Urbana–Champaign), the "Frontiers of Theoretical Econometrics" conference in celebration of Don Andrews' 60th birthday in Konstanz (August 2015), and the conference in honor of René Garcia (Montreal, August 2015) as well as seminar participants at the Federal Reserve Bank of St. Louis for helpful comments. The views expressed here are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

* Corresponding author.
E-mail addresses: sgoncal9@uwo.ca (S. Gonçalves), michael.w.mccracken@stls.frb.org (M.W. McCracken), benoit.perron@umontreal.ca (B. Perron).

http://dx.doi.org/10.1016/j.jeconom.2017.01.004
0304-4076/© 2017 Elsevier B.V. All rights reserved.

and Hansen and Timmermann (2015) delineate simulation-based methods for estimating asymptotically valid critical values.

The existing literature on out-of-sample methods for nested models is largely silent on situations where generated regressors – and, in particular, factors – are used for forecasting.[1] In most cases, the asymptotic distribution of the test of predictive ability is derived assuming that the predictors are observed and do not themselves need to be estimated prior to forecasting. A related exception is Cheng and Hansen (2015) who show that Mallows-based model averaging remains asymptotically valid when factors are used as predictors. Even so, in the context of in-sample methods, generated regressors have been shown to sometimes affect the asymptotic distribution of a statistic, typically through the standard errors. Examples include works by Pagan (1984, 1986), Randles (1982), and Murphy and Topel (1985). Of particular interest here, Gonçalves and Perron (2014) prove that in some instances an asymptotic bias is also introduced when factors are estimated rather than known.

Building on McCracken's (2007) and Clark and McCracken's (2001, 2005) results, we examine the asymptotic and finite sample properties of tests of equal forecast accuracy and encompassing applied to predictions from nested models when generated factors are used as predictors. Specifically, we establish conditions under which the asymptotic distributions of the *MSE-F* and *ENC-F* tests of predictive ability continue to hold even when the factors are estimated recursively prior to forecasting. Throughout we focus on one-step ahead forecasts made by linear regression models estimated by ordinary least squares (OLS). Our results hold under the null hypothesis of equal predictive ability as well as under a sequence of local alternatives.

Because our asymptotics depend on the relative size of the number of cross sectional units (N), the total number of observations (T), and the number of predictions (P), we provide a collection of Monte Carlo simulations designed to examine the size and power of the tests as we vary these parameters. In most of our simulations the presence of estimated factors leads to only minor size distortions, yielding rejection frequencies akin to those found in the existing literature when the predictors are observed rather than estimated. Nevertheless, we find that the estimation of factors reduces power relative to the case where factors are observed.

The rest of the paper is structured as follows. Section 2 introduces notation and describes the setup. Section 3 provides assumptions and a discussion of them. Section 4 contains our main theoretical results, including some asymptotic results on the estimation of factors over recursive samples which may be of independent interest. Section 5 presents Monte Carlo results on the finite sample size and power of the tests with an emphasis on the degree of distortions induced by factor estimation error. Section 6 applies the tests to forecasting excess stock returns using a large cross-section of macroeconomic aggregates and financial data. Finally, Section 7 concludes.

Before we proceed, we define some notation. For any matrix A, we let $\|A\| = tr^{1/2}(A'A)$ denote the Frobenius norm and $\|A\|_1 = \lambda_{\max}^{1/2}(A'A)$ the spectral norm. For $R = T - P$, we let $\sup_t$ denote $\sup_{R \leq t \leq T}$.

2. Setup and statistics of interest

2.1. Setup

In this paper we compare the predictive ability of a benchmark linear regression model with an alternative linear model that also includes factors. These additional predictors f_t are latent and denote a vector of r common factors underlying a panel factor model

$$x_{it} = \lambda_i' f_t + e_{it}, t = 1, \dots, T; \quad i = 1, \dots, N, \tag{1}$$

or in matrix form

$$X = F\Lambda' + e,$$

where X is an observed data matrix of size $T \times N$, $F = (f_1, \dots, f_T)'$ is a $T \times r$ matrix of factors, $\Lambda = (\lambda_1, \dots, \lambda_N)'$ is a $N \times r$ matrix, and e is $T \times N$.

We base our comparison of the two models on their ability to forecast y one step-ahead. To do so, we divide the total sample of $T = R+P$ observations into in-sample and out-of-sample portions. The in-sample observations span 1 to R, whereas the out-of-sample observations span $R + 1$ through T for a total of P one-step-ahead predictions.

For each forecast origin $t = R, \dots, T - 1$, we first generate forecasts of y_{t+1} recursively using the benchmark model by regressing y_{s+1} on the $k_1 \times 1$ observed predictor w_s for $s = 1, \dots, t - 1$ and hence

$$\hat{y}_{1,t+1} = w_t' \hat{\beta}_t,$$

where

$$\hat{\beta}_t = \left(\sum_{s=1}^{t-1} w_s w_s' \right)^{-1} \sum_{s=1}^{t-1} w_s y_{s+1}.$$

For the alternative model, forecasts of y_{t+1} are constructed in two steps. First, we estimate the latent factors recursively at each forecast origin $t = R, \dots, T - 1$ by the method of principal components. More specifically, at each time t, we compute an $N \times r$ matrix of factor loadings defined as $\tilde{\Lambda}_t = \left(\tilde{\lambda}_{1,t}, \dots, \tilde{\lambda}_{N,t} \right)'$ and the corresponding factors collected in the $t \times r$ matrix $\tilde{F}_t = \left(\tilde{f}_{1,t}, \dots, \tilde{f}_{t,t} \right)'$. Since estimation takes place recursively over $t = R, \dots, T - 1$, we index $\tilde{\Lambda}_t$ and $\tilde{F}_t$ by t. Thus, $\tilde{f}_{s,t}$ denotes the sth observation on the $r \times 1$ vector of factors estimated using data on x_{is} up to time t.

The principal components estimators are defined as

$$\left(\tilde{\Lambda}_t, \tilde{F}_t \right) = \arg \min_{\{\lambda_i, f_s\}} \frac{1}{Nt} \sum_{i=1}^{N} \sum_{s=1}^{t} \left(x_{is} - \lambda_i' f_s \right)^2,$$

subject to the normalization condition that $\frac{\tilde{F}_t' \tilde{F}_t}{t} = I_r$, an identity matrix of dimension r. Under this normalization, we can show that $\tilde{F}_t$ is equal to $\sqrt{t}$ times the $t \times 1$ eigenvector of the $t \times t$ matrix $X_t X_t' / (tN)$ and $\tilde{\Lambda}_t = \frac{X_t' \tilde{F}_t}{t}$, where we let

$$\underset{t \times N}{X_t} = \begin{pmatrix} x_{1,1} & \dots & x_{1,N} \\ \vdots & \ddots & \vdots \\ x_{t,1} & \dots & x_{t,N} \end{pmatrix}, \text{ for } t = R, \dots, T - 1.$$

Thus, $\frac{X_t X_t'}{tN} \tilde{F}_t = \tilde{F}_t \tilde{V}_t$, where $\tilde{V}_t = diag\left(\tilde{v}_{1,t}, \dots, \tilde{v}_{r,t} \right)$, with $\tilde{v}_{1,t} \geq \tilde{v}_{2,t} \geq \dots \geq \tilde{v}_{r,t} > 0$ and $\tilde{v}_{j,t}$ is the jth largest eigenvalue of $\frac{X_t X_t'}{tN}$. As is well known, the principal components estimator is not consistent for the true latent factors even under the normalization imposed above. As Bai (2003) and Stock and Watson (2002) showed, for given t, $\tilde{F}_t$ is mean square consistent for a rotated version of F_t, where the rotation matrix is defined as follows:

$$H_t = \tilde{V}_t^{-1} \left(\frac{\tilde{F}_t' F_t}{t} \right) \left(\frac{\Lambda' \Lambda}{N} \right). \tag{2}$$

[1] Li and Patton (2015) establish the first-order validity of the asymptotics in Clark and McCracken (2001, 2005) when the dependent variable, not the regressors, is generated. Fosten (2015) considers the non-nested case with estimated factors.

Given the estimated factors $\left\{\tilde{f}_{s,t} : s = 1, \ldots, t\right\}$, we then regress[2] y_{s+1} onto w_s and $\tilde{f}_{s,t}$ to get forecasts of y_{t+1} of the form

$$\hat{y}_{2,t+1} = w_t'\hat{\theta}_t + \tilde{f}_{t,t}'\hat{\alpha}_t \equiv \tilde{z}_{t,t}'\hat{\delta}_t,$$

where

$$\hat{\delta}_t = \left(\sum_{s=1}^{t-1} \tilde{z}_{s,t}\tilde{z}_{s,t}'\right)^{-1} \sum_{s=1}^{t-1} \tilde{z}_{s,t} y_{s+1},$$

with $\tilde{z}_{s,t} = \left(w_s', \tilde{f}_{s,t}'\right)'$ for $s = 1, \ldots, t-1$ and $t = R, \ldots, T-1$.

2.2. Test statistics

The statistics of interest are given by

$$ENC\text{-}F_{\tilde{f}} = \frac{\sum_{t=R}^{T-1} \hat{u}_{1,t+1}\left(\hat{u}_{1,t+1} - \hat{u}_{2,t+1}\right)}{\hat{\sigma}^2} \tag{3}$$

and

$$MSE\text{-}F_{\tilde{f}} = \frac{\sum_{t=R}^{T-1} \left(\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2\right)}{\hat{\sigma}^2}, \tag{4}$$

with

$$\hat{\sigma}^2 \equiv P^{-1} \sum_{t=R}^{T-1} \hat{u}_{2,t+1}^2, \tag{5}$$

where the forecast errors of the benchmark and alternative models are respectively denoted as:

$$\hat{u}_{1,t+1} = y_{t+1} - \hat{y}_{1,t+1} = y_{t+1} - w_t'\hat{\beta}_t$$

and

$$\hat{u}_{2,t+1} = y_{t+1} - \hat{y}_{2,t+1} = y_{t+1} - \tilde{z}_{t,t}'\hat{\delta}_t.$$

Let $u_{1,t+1}$ and $u_{2,t+1}$ denote these forecast errors when evaluated at the probability limits of $\hat{\beta}_t$ and $\hat{\delta}_t$ respectively. Finally, we will denote by $\ddot{u}_{1,t+1}$ and $\ddot{u}_{2,t+1}$ the analog out-of-sample forecast errors based on the model where factors are observed, i.e.

$$\ddot{u}_{1,t+1} = y_{t+1} - w_t'\ddot{\beta}_t = \hat{u}_{1,t+1},$$

given that[3] $\ddot{\beta}_t = \hat{\beta}_t$, and

$$\ddot{u}_{2,t+1} = y_{t+1} - z_t'\ddot{\delta}_t,$$

where $\ddot{\delta}_t$ is the recursive OLS estimator obtained from regressing y_{s+1} on $z_s = \left(w_s', f_s'\right)'$, for $s = 1, \ldots, t-1$, i.e.

$$\ddot{\delta}_t = \left(\sum_{s=1}^{t-1} z_s z_s'\right)^{-1} \sum_{s=1}^{t-1} z_s y_{s+1}.$$

Note that z_t differs from $\tilde{z}_{s,t}$ because it depends on the observed factors f_s and not the sample estimates $\tilde{f}_{s,t}$. This creates a discrepancy between $\ddot{u}_{2,t+1}$ and $\hat{u}_{2,t+1}$. Our main goal is to prove the asymptotic equivalence between the statistics based on $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$ and those based on $\ddot{u}_{1,t+1}$ and $\ddot{u}_{2,t+1}$, which we call $ENC\text{-}F_f$ and $MSE\text{-}F_f$.

Clark and McCracken (2001) derive the asymptotic distribution of the $ENC\text{-}F$ statistic under the null hypothesis H_0 : $E(u_{1,t+1}(u_{1,t+1} - u_{2,t+1})) = 0$ when all predictors are observed—that is, when both w_t and f_t are observed. Similarly, McCracken (2007) derives the asymptotic distribution of the $MSE\text{-}F$ statistic under the null hypothesis H_0 : $E(u_{1,t+1}^2 - u_{2,t+1}^2) = 0$ when all predictors are observed. In the same environment, Inoue and Kilian (2007) derive the asymptotic distribution of both statistics under a sequence of local alternatives. To prove asymptotic equivalence with estimated factors, we follow Inoue and Kilian (2007) and maintain that the DGP linking the predictors to the target variable y takes the form

$$y_{t+1} = w_t'\theta + f_t'(T^{-1/2}\alpha) + u_{t+1} \equiv z_t'\delta + u_{t+1}, \quad t = 1, \ldots, T, \tag{6}$$

where we let[4] $\delta \equiv \delta_T = \left(\theta' \quad T^{-1/2}\alpha'\right)'$. Given (6), our results have implications under both the null, for which $\alpha = 0$, and under small deviations from the null for which $\alpha \neq 0$. This result justifies using the usual critical values when testing for equal predictability with estimated factors in the larger, nesting model. It also suggests that factor estimation error need not lead to any substantial differences in the power of the test – at least when deviations from the null are small.

The first step in our proof is simply to establish an algebraic relationship between the out-of-sample test statistics given in (3), (4), and (5) and their analogs based on $\ddot{u}_{1,t+1}$ and $\ddot{u}_{2,t+1}$.

We can write

$$ENC\text{-}F_{\tilde{f}} = ENC\text{-}F_f + ENC\text{-}F_f\left(\frac{\ddot{\sigma}^2}{\hat{\sigma}^2} - 1\right) + \frac{1}{\hat{\sigma}^2}\sum_{t=R}^{T-1} \ddot{u}_{1,t+1}\left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1}\right)$$

and

$$MSE\text{-}F_{\tilde{f}} = MSE\text{-}F_f + MSE\text{-}F_f\left(\frac{\ddot{\sigma}^2}{\hat{\sigma}^2} - 1\right) - \frac{2}{\hat{\sigma}^2}\sum_{t=R}^{T-1}\left(\ddot{u}_{2,t+1}\left(\hat{u}_{2,t+1} - \ddot{u}_{2,t+1}\right)\right) - \frac{1}{\hat{\sigma}^2}\sum_{t=R}^{T-1}\left(\hat{u}_{2,t+1} - \ddot{u}_{2,t+1}\right)^2.$$

This decomposition allows us to identify a set of sufficient conditions for the asymptotic equivalence of the two sets of statistics (based on $\tilde{f}$ and on f, respectively).

Lemma 2.1. *Suppose that*

(a) $\sum_{t=R}^{T-1}\left(\hat{u}_{2,t+1} - \ddot{u}_{2,t+1}\right)^2 = o_P(1)$.
(b) $\sum_{t=R}^{T-1}\ddot{u}_{2,t+1}\left(\hat{u}_{2,t+1} - \ddot{u}_{2,t+1}\right) = o_P(1)$.
(c) $\sum_{t=R}^{T-1}\ddot{u}_{1,t+1}\left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1}\right) = o_P(1)$.
(d) $\hat{\sigma}^2 - \ddot{\sigma}^2 = o_P(1)$.

Then, it follows that $ENC\text{-}F_{\tilde{f}} = ENC\text{-}F_f + o_P(1)$ *and* $MSE\text{-}F_{\tilde{f}} = MSE\text{-}F_f + o_P(1)$.

[2] Our results remain valid when some of the additional predictors in the alternative model are observed as these will not introduce new terms in the decomposition of the discrepancy between the statistics with known and estimated factors presented below. We thank a referee for raising this possibility.

[3] By assumption, the benchmark model depends only on the observed regressors w_t and therefore there is no discrepancy between $\hat{u}_{1,t+1}$ and $\ddot{u}_{1,t+1}$. This would not be true if estimated factors were also present in the benchmark model. In this case, extra terms would appear in Lemma 2.1. For simplicity, we do not consider this extension here.

[4] For simplicity, we omit the index T in δ_T. Similarly, we ignore the triangle array notation in Eq. (6) by omitting the index T in $y_{t+1,T}$ and $u_{t+1,T}$.

The proof of Lemma 2.1 is trivial and therefore omitted. Our next goal is to provide a set of assumptions such that conditions (a)–(d) of Lemma 2.1 are verified. An important step in that direction is to provide bounds on the mean squared factors estimation uncertainty over recursive samples. But first we need to introduce our assumptions.

3. Assumptions

In this section, we state the assumptions that will be used to establish the asymptotic equivalence of the test statistics with estimated factors (ENC-$F_{\tilde f}$ and MSE-$F_{\tilde f}$) and the analogous statistics with known factors (ENC-F_f and MSE-F_f). Throughout, we let M be a generic finite constant.

Assumption 1. (a) For each s, $E\|f_s\|^{16} \leq M$, $\Sigma_f \equiv E\left(f_s f_s'\right) = I_r$ and $\Lambda'\Lambda$ is a diagonal matrix with distinct entries arranged in decreasing order.

(b) $\sup_t \left\|\frac{1}{T}\sum_{s=1}^{t}\left(f_s f_s' - I_r\right)\right\| = O_P\left(1/\sqrt{T}\right)$.

(c) $\left\|\frac{\Lambda'\Lambda}{N} - \Sigma_\Lambda\right\| = O\left(1/\sqrt{N}\right)$ such that $\lambda_{\min}\left(\Sigma_\Lambda\right)$ is bounded away from zero and bounded away from infinity – that is, $0 < a < \lambda_{\min}\left(\Sigma_\Lambda\right) < b < \infty$ for some constants a and b.

The first part of Assumption 1(a) requires f_s to have finite moments of order 16. Although some of our results could be derived under less stringent moment conditions (for instance, the results in Section 4.1 require only finite second moments), we will rely on these many moments in later parts of our proofs and therefore we impose this condition from the beginning. The remaining part of Assumption 1(a) is a normalization condition often used in factor analysis. It implies that both the factors and the factor loadings are orthogonal. See, in particular, Stock and Watson (2002) for a similar assumption and more recently, Fan et al. (2013), Bai and Li (2012), and Bai and Ng (2013). Part (b) of Assumption 1 implies that $T^{-1}\sum_{s=1}^{T} f_s f_s'$ converges to $\Sigma_f = I_r$ at $\sqrt{T}$-rate. This condition follows under standard mixing conditions on f_s by a maximal inequality for mixing processes. Part (c) of Assumption 1 requires that a nonnegligible fraction of the factor loadings is nonvanishing, implying that factors are pervasive. We specify the rate at which $\frac{\Lambda'\Lambda}{N}$ converges to Σ_Λ, a diagonal matrix given part (a). For simplicity, we treat the factors loadings as fixed, but if the factor loadings were assumed random, this would be the rate implied by the central limit theorem, see Han and Inoue (2015) for a similar assumption.

Assumption 2. (a) $E\left(e_{it}\right) = 0$, $E\left|e_{it}\right|^8 \leq M$.

(b) There exists a constant $m > 0$ such that $\lambda_{\min}\left(\Sigma_e\right) > m$, $\|\Sigma_e\|_1 = \lambda_{\max}\left(\Sigma_e\right) < M$, where $\Sigma_e = E\left(e_t e_t'\right)$.

(c) $\gamma_{t,s} \equiv E\left(e_t' e_s/N\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(e_{it}e_{is}\right)$ is such that $\gamma_{t,s} = 0$ whenever $|t-s| > \tau$ for some $\tau < \infty$, and $\left|\gamma_{t,s}\right| \leq M$ for all (t,s).

(d) For every (t,s), $E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(e_{it}e_{is} - E\left(e_{it}e_{is}\right)\right)\right)^{16} \leq M$.

(e) For every t, $E\left(\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right\|^{16}\right) \leq M$.

(f) $\frac{1}{N}\sum_{i,j=1}^{N}\frac{1}{T}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\left|cov\left(e_{is_1}e_{js_1}, e_{is_2}e_{js_2}\right)\right| \leq M$.

Assumption 2 is rather standard in this literature, allowing for weak dependence of e_{it} across i and over t. Part (b) of Assumption 2 corresponds to the approximate factor model of Chamberlain and Rothschild (1983), imposing an upper bound on the maximum eigenvalue of the covariance matrix of $e_t = \left(e_{1t}, \ldots, e_{Nt}\right)'$. It is satisfied if we assume that e_{it} is independent across i, given the moment conditions on e_{it}. However, it does not require cross sectional independence and is implied by the condition that $\max_{1\leq j\leq N}\sum_{i=1}^{N}\left|E\left(e_{it}e_{jt}\right)\right| \leq M$, an assumption used by Bai (2003).

Part (c) of Assumption 2 allows for serial correlation on e_{it} but restricts it to be of the moving average type. This is a stronger assumption than usual. The main reason we impose this assumption is the need for tight uniform bounds on the time average of the product of the factors estimation error $\tilde f_{s,t} - H_t f_s$ and u_{s+1}, the error term in the DGP for y. In particular, we need that

$$\sup_t\left\|t^{-1}\sum_{s=1}^{t-1}\left(\tilde f_{s,t} - H_t f_s\right)u_{s+1}\right\| = O_P\left(\frac{1}{\sqrt{R}\delta_{N,R}}\right),$$

where $\delta_{N,R}^2 = \min(N,R)$. This in turn is satisfied if

$$\sup_t t^{-1}\sum_{l=1}^{t}\left(\sum_{s=1}^{t-1}\gamma_{l,s}u_{s+1}\right)^2 = O_P(1),$$

which follows under the assumption that $\gamma_{t,s} = 0$ whenever $|t-s| > \tau$ for some $\tau < \infty$ (but not necessarily if we impose only the more usual weak dependence assumption that $T^{-1}\sum_{l,s=1}^{T}\gamma_{l,s}^2 \leq M$).

Part (d) of Assumption 2 strengthens Bai's (2003) Assumption C.5 by requiring 16 instead of 4 finite moments. Part (e) is also a strengthened version of Assumption 3(d) by Gonçalves and Perron (2014), where 16 instead of 2 finite moments are required. The relatively large number of moments we require is explained by the fact that we use mean square convergence arguments to show the asymptotic equivalence of our statistics. Since these are a function of squared forecast errors that depend on the product of estimated factors and regression estimates, the 16 moments assumptions are the result of an application of Cauchy–Schwarz inequality. Finally, Bai (2009) relies on an assumption similar to part (f) (see, in particular, his Assumption C.4) in the context of the interactive fixed effects model.

Assumptions 3 and 4 are new to the out-of-sample forecasting context.

Assumption 3. (a) $\frac{1}{N}\sum_{i=1}^{N}\sup_t\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{t} f_s e_{is}\right\|^2 = O_P(1)$, where $E\left(f_s e_{is}\right) = 0$.

(b) $\frac{1}{T}\sum_{l=1}^{T}\sup_t\left|\frac{1}{\sqrt{NT}}\sum_{s=1}^{t-1}\sum_{i=1}^{N}\left(e_{il}e_{is} - E\left(e_{il}e_{is}\right)\right)u_{s+1}\right|^2 = O_P(1)$, where $E\left(\left(e_{il}e_{is} - E\left(e_{il}e_{is}\right)\right)u_{s+1}\right) = 0$.

(c) $\sup_t\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{t-1}\sum_{i=1}^{N}\lambda_i e_{is}u_{s+1}\right\|^2 = O_P(1)$, where $E\left(\lambda_i e_{is}u_{s+1}\right) = 0$.

(d) $\frac{1}{T}\sum_{l=1}^{T}\sup_t\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{t-1}\sum_{i=1}^{N}\left(e_{il}e_{is} - E\left(e_{il}e_{is}\right)\right)z_s\right\|^2 = O_P(1)$, where $E\left(\left(e_{il}e_{is} - E\left(e_{il}e_{is}\right)\right)z_s\right) = 0$.

(e) $\sup_t\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{t-1}\sum_{i=1}^{N}\lambda_i e_{is}z_s'\right\|^2 = O_P(1)$, where $E\left(\lambda_i e_{is}z_s'\right) = 0$.

Assumption 3 requires that several partial sums be bounded in probability, uniformly in $t = R, \ldots, T$. Since the summands have mean zero, this assumption is not very restrictive and is implied by an application of the functional central limit theorem. For instance, we can verify that Assumption 3(b) and (c) hold under Assumption 2(d) and (e) if we assume that $\{u_t\}$ and $\{e_t\}$ are independent of each other and u_{t+1} is a martingale difference sequence, as we assume in Assumption 5.

Our next assumption requires that the partial sums in parts (b) and (c) of Assumption 3 have 8 finite moments.

Assumption 4. (a) For each (l,t), $E\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{t-1}\sum_{i=1}^{N}(e_{il}e_{is} - E(e_{il}e_{is}))u_{s+1}\right\|^8 \leq M$.

(b) For each t, $E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{t-1}\sum_{i=1}^{N}\lambda_i e_{is}u_{s+1}\right\|^8 \leq M$.

Assumption 4 is a strengthening of Assumption 4 of Gonçalves and Perron (2014), according to which $E\left\| \frac{1}{\sqrt{TN}} \sum_{s=1}^{T} \sum_{i=1}^{N} (e_{il}e_{is} - E(e_{il}e_{is}))u_{s+1} \right\|^2 \leq M$ and $E\left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{T} \sum_{i=1}^{N} \lambda_i e_{is} u_{s+1} \right\|^2 \leq M$. A sufficient condition for Assumption 4 is that u_{t+1} is i.i.d., given Assumption 2(d) and (e), respectively. Our assumption on u_{t+1} is nevertheless weaker, as we assume only that the regression innovation is a martingale difference sequence. More formally, we impose the following assumption on u_{t+1}.

Assumption 5. $E\left(u_{t+1}|\mathcal{F}^t\right) = 0$, where $\mathcal{F}^t = \sigma\left(y_t, y_{t-1}, \ldots; X_t, X_{t-1}, \ldots; z_t, z_{t-1}, \ldots\right)$.

Assumption 5 is analogous to the martingale difference assumption in Bai and Ng (2006) when the forecasting horizon is 1.

Assumption 6. (a) $E\|z_t\|^{16} \leq M$, $E\left(u_{t+1}^8\right) \leq M$ for all $t = 1, \ldots, T$.

(b) Let $V(t) = t^{-1} \sum_{s=1}^{t-1} z_s u_{s+1}$, where $E(z_s u_{s+1}) = 0$. Then $\sup_t \|V(t)\| = O_P\left(1/\sqrt{T}\right)$ and $E\left\|\sqrt{T}V(t)\right\|^8 \leq M$ for all t.

(c) Let $B_0(t) \equiv \left(t^{-1} \sum_{s=1}^{t-1} w_s w_s'\right)^{-1}$, $B(t) \equiv \left(t^{-1} \sum_{s=1}^{t-1} z_s z_s'\right)^{-1}$, $B_0 = \left(E\left(w_s w_s'\right)\right)^{-1}$, and $B = \left(E\left(z_s z_s'\right)\right)^{-1}$. Then, $\sup_t \|B_0(t)\| = O_P(1)$, $\sup_t \|B(t)\| = O_P(1)$, $\sup_t \|B_0(t) - B_0\| = O_P\left(1/\sqrt{T}\right)$, and $\sup_t \|B(t) - B\| = O_P\left(1/\sqrt{T}\right)$.

(d) $P, R \to \infty$ such that $P/R = O(1)$.

Assumption 6 imposes conditions on the regressors and errors of the predictability regression model based on the latent factors. In particular, Assumption 6(a) extends the moment conditions on the factors assumed in Assumption 1 to the observed regressors w_t. In addition, while it requires the innovations u_{t+1} to have eight finite moments it permits conditional heteroskedasticity. Parts (b) and (c) of Assumption 6 imply that $\sup_t \left\|\ddot{\beta}_t - \theta\right\|$ and $\sup_t \left\|\ddot{\delta}_t - \delta\right\|$ are both $O_P\left(1/\sqrt{T}\right)$, which is the usual rate of convergence. These assumptions are implied by more primitive assumptions that ensure the application of mixingale-type inequalities. Part (d) is implied by the usual rate condition on P and R according to which $\lim_{P,R\to\infty} \frac{P}{R} = \pi$, with $0 \leq \pi < \infty$.

4. Theory

4.1. Properties of factors over recursive samples

As mentioned earlier, factor models are subject to a fundamental identification problem. In our setup, this means that factors $\tilde{f}_{s,t}$ estimated at each forecasting origin t are consistent for a rotation of the true factors given by $H_t f_s$, where the rotation matrix H_t is time-varying. The asymptotic properties of the rotation matrix H_t and its inverse (as well as the inverse of the eigenvalue matrix $\tilde{V}_t^{-1}$) are a crucial step to proving the asymptotic negligibility of the factors estimation error in the out-of-sample forecasting context. Lemmas A.5 and A.6 in Appendix A contain these results. These results are crucial to bounding terms involving the estimated factors in the expansions of our test statistics.

The following theorem establishes the rates of convergence of the estimated factors over recursive subsamples.

Theorem 4.1. *Under Assumptions 1–6,*

(a) $\frac{1}{P} \sum_{t=R}^{T-1} \left\|\tilde{f}_{t,t} - H_t f_t\right\|^2 = O_P\left(1/\delta_{NR}^2\right)$, *where* $\delta_{N,R}^2 = \min(N, R)$.

(b) $\sup_t \left\{\frac{1}{t} \sum_{s=1}^{t} \left\|\tilde{f}_{s,t} - H_t f_s\right\|^2\right\} = O_P\left(1/\delta_{NR}^2\right)$.

Theorem 4.1 extends Lemma A.1 of Bai (2003) to the out-of-sample context. Just as Bai's (2003) result is crucial to proving the asymptotic theory of the method of principal components over one sample when both N and T are large, Theorem 4.1 is crucial to proving the asymptotic theory of this estimation method over recursive samples. Part (a) of Theorem 4.1 shows that the time average over the out-of-sample period of the squared deviations between the estimated factor $\tilde{f}_{t,t}$ and the rotated factor $H_t f_t$ converges to zero at rate $O_P\left(1/\delta_{N,R}^2\right)$. Part (b) shows that the same rate of convergence applies uniformly over $t = R, \ldots, T$ to the time average of the squared deviations between $\tilde{f}_{s,t}$ and $H_t f_s$ for each recursive sample estimation period $s = 1, \ldots, t$.

4.2. Asymptotic equivalence

This section's main contribution is to show that Lemma 2.1 holds under the sequence of local alternatives implied by (6). In particular, we verify conditions (a)–(d), which involve the forecast errors $\ddot{u}_{1,t+1}$, $\ddot{u}_{2,t+1}$, $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$, and their respective differences.

Given the definitions of $\ddot{u}_{2,t+1}$ and $\hat{u}_{2,t+1}$, we have that

$$\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} = \tilde{z}_{t,t}'\hat{\delta}_t - z_t'\ddot{\delta}_t.$$

Let $\Phi_t = diag\left(I_{k_1}, H_t\right)$. Adding and subtracting appropriately yields

$$\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} = \left(\tilde{z}_{t,t} - \Phi_t z_t\right)' \Phi_t'^{-1}\left(\ddot{\delta}_t - \delta\right) + \tilde{z}_{t,t}'\left(\hat{\delta}_t - \Phi_t'^{-1}\ddot{\delta}_t\right) + \left(\tilde{z}_{t,t} - \Phi_t z_t\right)' \Phi_t'^{-1}\delta. \quad (7)$$

The first term in (7) measures the discrepancy between the "ideal" predictors at time t, given by $z_t = \left(w_t', f_t'\right)'$, and their estimated version given by $\tilde{z}_{t,t} = \left(w_t', \tilde{f}_{t,t}'\right)'$, interacted with the error in estimating δ using $\ddot{\delta}_t$. The second term in (7) measures the impact of the factor estimation error on the OLS estimators. The third term in (7) exists only under the alternative and hence captures the impact of factor estimation error under the alternative. Given our assumption that there are no generated regressors in the benchmark model, we can write this term as $\left(\tilde{f}_{t,t} - H_t f_t\right)' H_t'^{-1}\alpha/\sqrt{T}$, given the definition of δ implicit in (6).

Our first goal is to show that $\sum_{t=R}^{T-1}\left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1}\right)^2 = o_P(1)$. By Assumption 6, $\sup_t \left\|\ddot{\delta}_t - \delta\right\| = O_P\left(1/\sqrt{T}\right)$, whereas

$$\frac{1}{T}\sum_{t=R}^{T-1}\left\|\tilde{z}_{t,t} - \Phi_t z_t\right\|^2 = \frac{1}{T}\sum_{t=R}^{T-1}\left\|\tilde{f}_{t,t} - H_t f_t\right\|^2 = O_P\left(1/\delta_{N,R}^2\right)$$

by Theorem 4.1. Thus, by an application of the Cauchy–Schwarz inequality, we can show that the sum of the squared contributions from the first term in (7) is of order $O_P\left(1/\delta_{N,R}^2\right)$. This is $o_P(1)$ given that $N, R \to \infty$. Similarly, we can bound the sum of the squared contributions from the third term by

$$\sum_{t=R}^{T-1}\left\|\tilde{f}_{t,t} - H_t f_t\right\|^2 \left\|H_t'^{-1}\right\|^2 \left\|T^{-1/2}\alpha\right\|^2 \leq \sup_t \left\|H_t^{-1}\right\|^2 \|\alpha\|^2 \frac{1}{T}\sum_{t=R}^{T-1}\left\|\tilde{f}_{t,t} - H_t f_t\right\|^2 = O_P\left(1/\delta_{N,R}^2\right)$$

given Theorem 4.1. Thus, this term's contribution is also $o_P(1)$ given that $N, R \to \infty$.

Finally, the sum of squared contributions from the second term is bounded by

$$T \sup_t \left\| \hat{\delta}_t - \Phi_t'^{-1} \ddot{\delta}_t \right\|^2 \frac{1}{T} \sum_{t=R}^{T-1} \left\| \tilde{z}_{t,t} \right\|^2,$$

where $T^{-1} \sum_{t=R}^{T-1} \left\| \tilde{z}_{t,t} \right\|^2 = O_P(1)$ under our assumptions. Thus, condition (a) in Lemma 2.1 follows if $\sup_t \left\| \hat{\delta}_t - \Phi_t'^{-1} \ddot{\delta}_t \right\|^2 = o_P(1/T)$, as shown in the next lemma.

Lemma 4.1. *Suppose Assumptions 1–6 hold. Then,*

$$\sup_t \left\| \hat{\delta}_t - \Phi_t'^{-1} \ddot{\delta}_t \right\| = O_P\left(\frac{1}{\sqrt{R}\delta_{N,T}} \right).$$

Since Lemma 4.1 implies that $\sup_t \left\| \hat{\delta}_t - \Phi_t'^{-1} \ddot{\delta}_t \right\|^2 = O_P \left(\frac{1}{R\delta_{N,T}^2} \right) = o_P\left(\frac{1}{T}\right)$, this result and the above arguments imply that

$$\sum_{t=R}^{T-1} \left(\hat{u}_{2,t+1} - \ddot{u}_{2,t+1} \right)^2 = O_P\left(\frac{1}{\delta_{N,T}^2} \right) = o_P(1),$$

which is condition (a) of Lemma 2.1.

Next, we focus on condition (b) of Lemma 2.1. Given the definition of $\ddot{u}_{2,t+1}$, we can write

$$\begin{aligned} \ddot{u}_{2,t+1} &= y_{t+1} - z_t' \ddot{\delta}_t = y_{t+1} - z_t'\delta - z_t' \left(\ddot{\delta}_t - \delta \right) \\ &= u_{t+1} - z_t' \left(\ddot{\delta}_t - \delta \right), \end{aligned}$$

where $u_{t+1} = y_{t+1} - z_t'\delta$ given (6). It follows that

$$\begin{aligned} \sum_{t=R}^{T-1} \ddot{u}_{2,t+1} \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right) &= \sum_{t=R}^{T-1} u_{t+1} \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right) \\ &\quad - \sum_{t=R}^{T-1} z_t' \left(\ddot{\delta}_t - \delta \right) \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right). \end{aligned}$$

The second term is bounded by

$$\begin{aligned} &\sum_{t=R}^{T-1} \left| z_t' \left(\ddot{\delta}_t - \delta \right) \right| \left| \ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right| \\ &\leq \left(\sum_{t=R}^{T-1} \left| z_t' \left(\ddot{\delta}_t - \delta \right) \right|^2 \right)^{1/2} \left(\sum_{t=R}^{T-1} \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right)^2 \right)^{1/2}, \end{aligned}$$

given the Cauchy–Schwarz inequality. Since the square of the first term is bounded by

$$\sum_{t=R}^{T-1} \left| z_t' \left(\ddot{\delta}_t - \delta \right) \right|^2 \leq \sup_t \left\| \sqrt{T} \left(\ddot{\delta}_t - \delta \right) \right\|^2 \frac{1}{T} \sum_{t=R}^{T-1} \| z_t \|^2,$$

which is $O_P(1)$ under Assumption 6, condition (b) is implied by condition (a) and the next lemma.

Lemma 4.2. *Suppose Assumptions 1–6 holds. If $\sqrt{T}/N \to 0$, then $\sum_{t=R}^{T-1} u_{t+1} \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right) = o_P(1)$.*

Condition (c) of Lemma 2.1 can be verified as follows. Adding and subtracting appropriately,

$$\begin{aligned} \ddot{u}_{1,t+1} &= y_{t+1} - w_t' \ddot{\beta}_t = y_{t+1} - z_t'\delta + z_t'\delta - w_t'\ddot{\beta}_t \\ &= u_{t+1} + w_t' \left(\theta - \ddot{\beta}_t \right) + f_t' \left(T^{-1/2} \alpha \right) \end{aligned}$$

where $\delta = \left(\theta', \left(T^{-1/2} \alpha' \right) \right)'$. It follows that

$$\begin{aligned} &\sum_{t=R}^{T-1} \ddot{u}_{1,t+1} \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right) \\ &= \sum_{t=R}^{T-1} u_{t+1} \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right) \\ &\quad + \sum_{t=R}^{T-1} w_t' \left(\theta - \ddot{\beta}_t \right) \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right) \\ &\quad + \sum_{t=R}^{T-1} f_t' \left(T^{-1/2} \alpha \right) \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right). \end{aligned} \tag{8}$$

The first term in (8) is $o_P(1)$ from Lemma 4.2. The second term can be bounded by

$$\begin{aligned} &\sum_{t=R}^{T-1} \| w_t \| \left\| \theta - \ddot{\beta}_t \right\| \left| \ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right| \\ &\leq \sup_t \left\| \theta - \ddot{\beta}_t \right\| \sum_{t=R}^{T-1} \| w_t \| \left| \ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right| \\ &\leq \underbrace{\sup_t \left(\left\| \sqrt{T} \left(\theta - \ddot{\beta}_t \right) \right\| \right)}_{=O_P(1)} \underbrace{\left(\frac{1}{T} \sum_{t=R}^{T-1} \| w_t \|^2 \right)^{1/2}}_{=O_P(1)} \\ &\quad \times \underbrace{\left(\sum_{t=R}^{T-1} \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right)^2 \right)^{1/2}}_{=o_P(1) \text{ by part (a)}}, \end{aligned}$$

where we can show that $\sup_t \left(\left\| \sqrt{T} \left(\theta - \ddot{\beta}_t \right) \right\| \right) = O_P(1)$ under our assumptions.

Finally, the third term in (8) can be bounded by

$$\| \alpha \| \left(\frac{1}{T} \sum_{t=R}^{T-1} \| f_t \|^2 \right)^{1/2} \left(\sum_{t=R}^{T-1} \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right)^2 \right)^{1/2},$$

which is $o_P(1)$ by part (a).

Similarly, condition (d) follows from conditions (a) and (b) since

$$\begin{aligned} \hat{\sigma}^2 - \ddot{\sigma}^2 &= P^{-1} \sum_{t=R}^{T-1} \left(\hat{u}_{2,t+1}^2 - \ddot{u}_{2,t+1}^2 \right) = P^{-1} \sum_{t=R}^{T-1} \left(\hat{u}_{2,t+1} - \ddot{u}_{2,t+1} \right)^2 \\ &\quad + 2P^{-1} \sum_{t=R}^{T-1} \ddot{u}_{2,t+1} \left(\hat{u}_{2,t+1} - \ddot{u}_{2,t+1} \right) = O_P\left(\frac{1}{P} \right), \end{aligned}$$

and $P \to \infty$. The next theorem contains our main theoretical result.

Theorem 4.2. *Under Assumptions 1–6, if $\sqrt{T}/N \to 0$ then ENC-$F_{\tilde{f}}$ = ENC-$F_f + o_P(1)$ and MSE-$F_{\tilde{f}}$ = MSE-$F_f + o_P(1)$.*

This result justifies using the tabulated asymptotic critical values given in Clark and McCracken (2001) for the *ENC-F* statistic and in McCracken (2007) for the *MSE-F* statistic. The rate restriction $\sqrt{T}/N \to 0$ is the same as the one used by Bai and Ng (2006) to show that factor estimation uncertainty disappears asymptotically when conducting inference in factor-augmented regression models. In our context, this rate restriction is used to show that the difference between the out-of-sample test statistics based on the estimated factors and those based on the latent factors is asymptotically negligible. It is sufficient for proving our result, but we have been unable to prove that it is necessary. It is certainly possible that it is stronger than needed.

Note that the asymptotic critical values of Clark and McCracken (2001) and McCracken (2007) are computed assuming conditional homoskedasticity, whereas our asymptotic equivalence result holds under conditional heteroskedasticity. Without this assumption, a different set of critical values should be used. Clark and McCracken (2012) have proposed bootstrap critical values to approximate the asymptotic distribution under conditional heteroskedasticity. Likewise, for the $MSE - F$ statistic, Hansen and Timmermann (2015) develop a simulation-based method for estimating asymptotically valid critical values under conditional heteroskedasticity. It is likely that both of these methods are valid in this context, but formal proofs are left for future research.

5. Simulations

In this section, we analyze the finite sample behavior of tests of equal predictive ability with estimated factors. For this purpose, we use a data-generating process similar to the one in Clark and McCracken (2001):

$$\begin{pmatrix} y_s \\ f_s \end{pmatrix} = \begin{pmatrix} 0.5 & b \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} y_{s-1} \\ f_{s-1} \end{pmatrix} + \begin{pmatrix} u_{y,s} \\ u_{f,s} \end{pmatrix}$$

for $s = 1, \ldots, T$, where the vector of errors is i.i.d. $N(0, I_2)$ and $y_0 = f_0 = 0$. The value of b will be set to 0 for the size experiment and 0.3 for the power experiment.[5]

The panel of variables used to extract the factors is obtained as

$$x_{is} = \lambda_i f_s + \sigma e_{is},$$

where $\lambda_i \sim U(0, 1)$, $e_{is} \sim N(0, \sigma_i^2)$, $\sigma_i^2 \sim U(0.5, 1.5)$, and $\sigma = 1$.[6]

We consider as sample sizes $N \in \{10, 20, 50, 100, 200\}$ and $T \in \{25, 50, 100, 200\}$. We report results for three sample splits: $\pi = \frac{P}{R} \in \{0.2, 1, 2\}$. The number of replications is set at 100,000.

We test the predictive ability of factors using recursive windows. For each $t = R, \ldots, T - 1$, we estimate the parameters of the benchmark model:

$$y_{s+1} = \beta_0 + \beta_1 y_s + u_{1,s+1},$$

using data from $s = 1, \ldots, t - 1$, while the factor-augmented model is given by

$$y_{s+1} = \theta_0 + \theta_1 y_s + \alpha f_s + u_{2,s+1},\ s = 1, \ldots, t - 1,$$

where we estimate f_s with $\tilde{f}_{s,t}$ recursively using data up to time t, x_{is}, $i = 1, \ldots, N$, and $s = 1, \ldots, t$. We consider the $ENC\text{-}F_{\tilde{f}}$ and $MSE\text{-}F_{\tilde{f}}$ statistics. All tests are conducted at the 5% significance level.

The results are reported in Table 1 and in Figs. 1 and 2. We report right-tail rejection probabilities using the asymptotic critical values for each statistic for two cases: when the factor is known and when the factor is estimated. The asymptotic critical values are from Clark and McCracken (2001) for the $ENC\text{-}F$ statistic and McCracken (2007) for the $MSE\text{-}F$ statistic. The results for $\pi = 0.2$ are reported first, followed by $\pi = 1$ and $\pi = 2$. In each case, we report results in the Table for the $ENC\text{-}F$ and $MSE\text{-}F$, first with the factor assumed known and then with the factor estimated. In the figures, the known factor case is represented with dotted lines, and the estimated factor case with solid lines. The $ENC - F$ statistic is in blue, and the $MSE - F$ statistic is in red with dots. We also draw a line at the nominal size of 5%.

The size results are reported in the top panel of Table 1 and in Fig. 1. While there should be no effect of changes in N for the observed factor case, we do see a little variation in the Table, but it is sufficiently small to be impossible to see in the figure. The main point however, is that, as predicted by theory, the statistics behave very similarly under the null hypothesis regardless of whether the factor is estimated or known. We do see some discrepancies for smaller values of N and T. However, for $N \geq 50$ or $T \geq 100$, the size properties are almost identical. Overall, all tests have good size properties, though there is a tendency to overreject for smaller values of T especially for $T = 25$. The $MSE\text{-}F$ statistic tends to be less subject to size distortions than the $ENC\text{-}F$ test statistic.

The bottom panel of Table 1 and Fig. 2 reports power results for the two tests considered. This panel and figure contain a very important result: despite our asymptotic equivalence result under local alternatives, estimating the factor reduces power. For example, for $N = 50$, $T = 50$, and $\pi = 0.2$, power of the $MSE\text{-}F$ statistic is 41.7% if the factor is known, but 40.3% if the factor is estimated. This is a general result for all values of N, T, and π. However, this loss in power is reduced when the sample size, especially the cross-sectional dimension, increases. Again, for $N \geq 50$, the loss of power is marginal.

Table 1 and Fig. 1 also contain other results found in the literature. For example, we see a power advantage for the $ENC\text{-}F$ statistic over the $MSE\text{-}F$ statistic (see Clark and McCracken, 2001) and that power is increasing in P which can be achieved by increasing T for fixed π or increasing π for fixed T.

6. Empirical results

In this section, we provide empirical evidence on the predictive content of macroeconomic factors for the equity premium associated with the S&P 500 Composite Index. Others who have looked into factor-based prediction of the equity premium include Bai (2010), Cakmakli and van Dick (2016), and Batje and Menkhoff (2012). Like us, they use monthly frequency data to investigate whether factor-based model forecasts are more accurate than just using the historical mean. Unlike us, formal tests of statistical significance are not necessarily conducted (though to be fair, prior to the current paper a theoretical justification for doing so did not exist). In addition, in each of these three papers an in-sample information criterion is used to select the "best" model prior to forecasting—an approach we do not follow. Instead, we consider a range of possible models each defined by a given permutation of the current estimated factors.

In our exercise, we construct factors $\tilde{f}_{s,t}$ using a large monthly frequency macroeconomic database of 134 predictors documented in McCracken and Ng (2016).[7] This dataset is designed to be representative of those most frequently used in the literature on factor-based forecasting and appears to be similar to those used in the papers cited above. If we let P_{t+1} denote the value of the stock index at the close of business on the last business day of month $t + 1$, we construct our equity premium as $y_{t+1} = \ln(P_{t+1}/P_t) - r_{t+1}$, where r_{t+1} denotes the 1-month Treasury bill rate.[8] After transforming the data to induce stationarity and truncating the dataset to avoid missing values, our dataset spans 1960:03 through 2014:12 for a total of $T = 658$ observations.

[5] We also conducted an experiment with the autoregressive coefficient on y_s set to 0 in both the benchmark and alternative models. This specification is analogous to our empirical specification in the next section, but because the results were similar, we decided not to report them. They are available from the authors upon request.

[6] We have also experimented with errors drawn from a Student distribution with 3 degrees of freedom to check sensitivity to the strong moment conditions imposed in our assumptions. We also amplified estimation error in the factor estimates by considering $\sigma = 5$. Results are qualitatively similar and are available upon request from the authors.

[7] We extract the factors using the April 2015 vintage of FRED-MD. Four of the series are dropped to avoid the presence of missing values. By doing so we can estimate the factors using precisely the same formulas given in the text.

[8] The 1-month Treasury bill rates are obtained from Kenneth French's website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

Table 1
Size and power for MSE-F and ENC-F (%).

	T	N	$\pi=0.2$				$\pi=1$				$\pi=2$			
			$ENC-F_f$	$ENC-F_{\tilde{f}}$	$MSE-F_f$	$MSE-F_{\tilde{f}}$	$ENC-F_f$	$ENC-F_{\tilde{f}}$	$MSE-F_f$	$MSE-F_{\tilde{f}}$	$ENC-F_f$	$ENC-F_{\tilde{f}}$	$MSE-F_f$	$MSE-F_{\tilde{f}}$
$b=0$ (size)	25	10	11.8	11.4	10.1	9.9	9.6	9.2	6.7	6.8	10.0	9.4	5.8	6.1
		20	11.9	11.6	10.2	10.0	9.6	9.3	6.8	6.7	9.9	9.7	5.8	5.9
		50	11.9	11.8	10.2	10.2	9.7	9.6	6.8	6.8	10.1	10.0	5.7	5.8
		100	11.7	11.8	10.1	10.0	9.6	9.6	6.7	6.8	10.0	9.9	5.7	5.7
		200	12.1	12.1	10.3	10.4	9.7	9.6	6.8	6.8	10.0	10.0	5.8	5.8
	50	10	8.4	8.3	7.8	7.8	7.0	6.8	6.0	5.9	7.2	6.9	5.2	5.4
		20	8.5	8.6	8.0	8.1	7.1	7.1	5.9	6.0	7.3	7.2	5.3	5.4
		50	8.3	8.4	7.8	7.9	7.1	7.0	6.0	5.9	7.2	7.0	5.2	5.2
		100	8.4	8.4	7.9	7.8	7.0	6.9	5.9	5.9	7.2	7.2	5.2	5.2
		200	8.3	8.4	7.7	7.8	6.9	6.9	5.9	5.9	7.1	7.1	5.1	5.1
	100	10	6.6	6.6	6.7	6.6	6.0	5.9	5.5	5.5	5.9	5.8	5.2	5.2
		20	6.7	6.7	6.7	6.7	6.0	6.0	5.5	5.5	5.9	5.9	5.1	5.1
		50	6.8	6.7	6.8	6.8	6.0	5.9	5.5	5.4	5.9	5.9	5.1	5.1
		100	6.7	6.7	6.7	6.7	6.0	6.0	5.6	5.6	5.9	5.9	5.2	5.2
		200	6.8	6.8	6.7	6.8	6.0	6.1	5.6	5.6	6.0	6.0	5.2	5.2
	200	10	5.9	5.9	6.1	6.1	5.4	5.4	5.3	5.3	5.2	5.1	5.0	4.9
		20	6.1	6.0	6.2	6.2	5.4	5.5	5.3	5.3	5.2	5.3	5.0	5.1
		50	6.0	6.0	6.1	6.2	5.4	5.5	5.3	5.3	5.3	5.3	5.0	5.0
		100	5.9	6.0	6.2	6.1	5.5	5.5	5.2	5.3	5.3	5.3	5.0	5.0
		200	6.0	6.0	6.2	6.3	5.5	5.5	5.3	5.2	5.4	5.4	5.1	5.1
$b=0.3$ (power)	25	10	35.7	28.7	29.0	23.7	38.0	28.8	28.7	22.0	37.3	27.9	27.1	20.8
		20	35.6	32.1	29.2	26.3	37.9	32.9	28.5	25.1	37.1	32.0	27.0	23.5
		50	35.8	34.4	28.8	27.7	38.0	36.0	28.7	27.4	37.3	35.3	27.2	26.0
		100	35.9	35.1	29.1	28.6	37.7	36.8	28.4	27.7	37.1	36.2	27.0	26.3
		200	35.7	35.3	28.9	28.7	37.9	37.5	28.6	28.3	37.3	36.7	27.2	26.7
	50	10	51.4	41.3	41.5	33.8	60.7	47.4	48.8	38.3	61.9	47.9	51.0	39.7
		20	51.3	46.3	41.3	37.5	60.7	54.2	48.9	43.9	61.7	54.9	51.0	45.6
		50	51.5	49.5	41.4	40.0	60.7	58.2	49.0	47.0	61.9	59.1	51.2	48.9
		100	51.4	50.3	41.4	40.6	60.8	59.5	49.0	48.0	61.8	60.5	51.2	50.1
		200	51.6	51.1	41.6	41.2	60.7	59.9	48.7	48.2	61.7	61.1	51.0	50.4
	100	10	73.6	62.9	58.0	49.9	88.0	76.5	74.4	63.5	89.1	77.3	79.7	68.0
		20	73.6	68.6	58.1	54.0	87.6	82.6	74.1	69.2	88.8	83.8	79.2	74.1
		50	73.3	71.5	57.6	56.2	88.0	86.1	74.1	72.2	89.1	87.3	79.5	77.5
		100	73.4	72.4	57.8	56.9	87.9	87.0	74.2	73.3	88.9	88.1	79.4	78.4
		200	73.7	73.3	58.1	57.8	87.9	87.4	74.2	73.7	89.0	88.5	79.3	78.9
	200	10	92.2	85.3	74.8	67.6	99.2	96.1	92.2	85.7	99.5	96.8	96.0	90.6
		20	92.2	89.2	74.6	71.3	99.1	98.2	92.2	89.4	99.5	98.7	95.9	93.8
		50	92.2	91.1	74.7	73.4	99.2	98.9	92.1	91.1	99.4	99.2	95.9	95.3
		100	92.2	91.6	74.7	74.0	99.1	99.0	92.0	91.6	99.5	99.3	95.8	95.5
		200	92.2	91.9	74.5	74.2	99.2	99.1	92.1	91.8	99.4	99.4	95.8	95.6

We construct our forecasting exercises with two goals in mind. First, we want to identify the factors with the greatest predictive content for the equity premium. Toward this goal we consider each of the first 8 macroeconomic factors $(\tilde{f}_{1t}, \ldots, \tilde{f}_{8t})$ as potential predictors.[9] As shown later, the second factor, which loads heavily on interest rate variables, is clearly the strongest predictor.

Second, following Ghysels et al. (2014), we investigate the importance of the "timing" of the estimated factors. In our reading of the factor-based forecasting literature (not limited to the papers listed earlier), it is not always clear whether the time t factors are dated using calendar or data-release time. Note that in Eq. (6), the dependent variable is associated with time $t+1$, while the factor is associated with time t. Since our dependent variable is observable immediately at the close of business on the last business day of month $t+1$, the distinction between calendar and data-release time is irrelevant. This is in sharp contrast to the data used to construct the factors. The dataset used consists of macroeconomic series that are typically released with a 1-month delay. As such, the factor estimates based on data associated with (say) April 2015 cannot literally be constructed until May 2015 after the data is released. To evaluate the importance of this distinction, we conduct our forecasting exercise once estimating the factors $\tilde{f}_{s,t}$ such that s and t represent calendar time (i.e., the data are associated with months s and t) and once such that s and t represent data-release time (i.e. the data is associated with months $s-1$ and $t-1$).

For all model comparisons, our benchmark model uses the historical mean as the predictor and hence Model 1 takes the form

$$y_{s+1} = \beta + u_{1,s+1},\ s = 1, \ldots, t-1.$$

The competing models are similar but add estimated factors and hence take the form

$$y_{s+1} = \theta + \tilde{f}_{s,t}'\alpha + u_{2,s+1},\ s = 1, \ldots, t-1.$$

We consider all $2^8 - 1 = 255$ permutations of models that include a single lag of at least 1 of the first 8 factors. For each we construct one-step-ahead forecasts and test the null hypotheses of equal predictive ability at the 5% level using the *MSE-F* and *ENC-F* test statistics based on critical values taken from Clark and McCracken (2001) and McCracken (2007). Note that the appropriate critical values vary with the number of factors used in the competing model. All models are estimated recursively by

9 As discussed in McCracken and Ng (2016), Factor 1 can be interpreted as a real activity/employment factor. Factor 2 is dominated by interest rate variables. Factor 3 is concentrated on price variables. Factors 4 and 5 are a mix of housing and interest rate variables. Like Factor 1, Factor 6 concentrates on real activity/employment variables. Factor 7 is associated with stock market variables, while Factor 8 loads heavily on exchange rates.

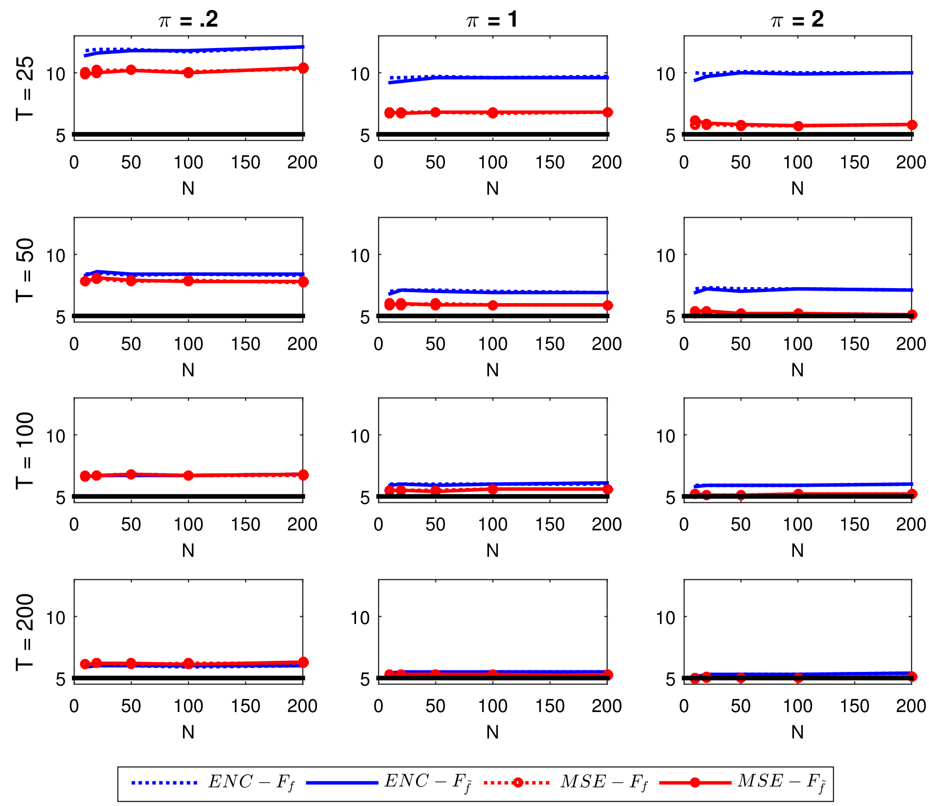

Fig. 1. Size of equal-predictability tests (%).

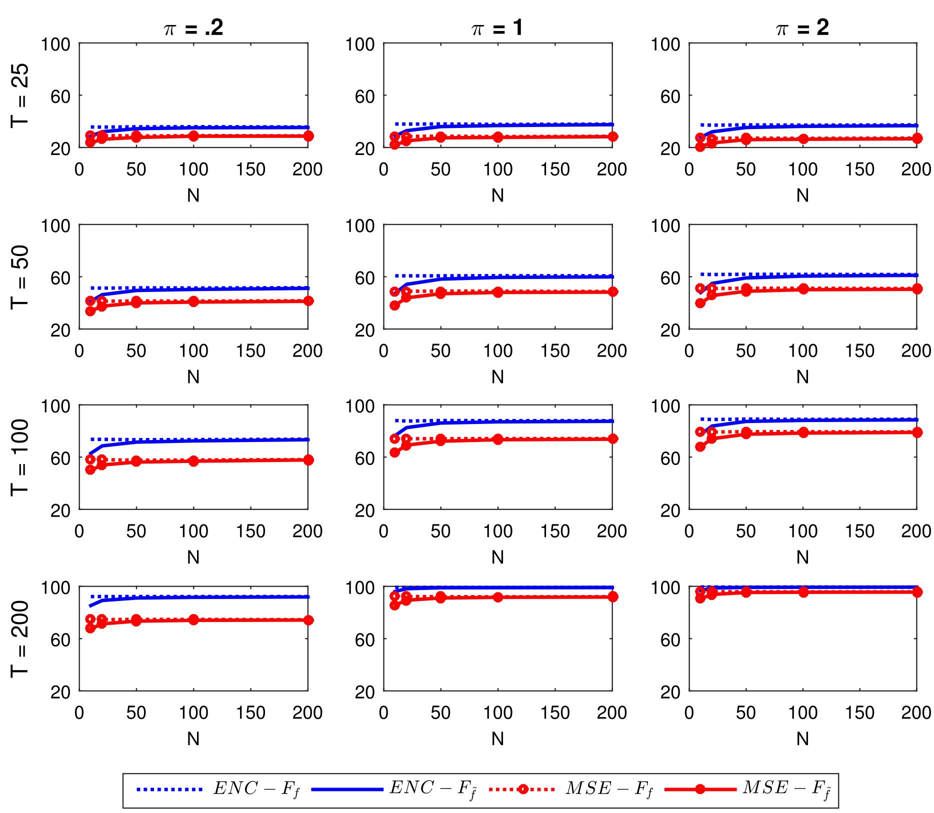

Fig. 2. Power of equal-predictability tests (%).

OLS. We consider two out-of-sample periods: 1985:01–2014:12 and 2005:01–2014:12. These imply sample splits (R, P) equal to (298, 360) and (438, 120), respectively.

Rather than report all of the potential pairwise comparisons, for each out-of-sample period and for each definition of time t (calendar vs. data-release time), we report the results associated

Table 2
Factor-based forecasting of the equity premium.

Calendar time							
1985:01–2014:12				2005:01–2014:12			
Model	Ratio	MSE-F	ENC-F	Model	Ratio	MSE-F	ENC-F
2	0.98	5.52*	7.14*	2, 8	0.95	5.73*	5.49*
2, 6	0.99	4.82*	6.21*	2, 6, 8	0.96	5.51*	4.87*
2, 8	0.99	4.71*	8.05*	2, 5, 8	0.96	5.46*	5.41*
2, 6, 8	0.99	4.22*	7.21*	2, 5, 6, 8	0.96	5.26*	4.79*
2, 4	0.99	3.56*	5.99*	2, 4, 8	0.96	4.66*	3.47*
2, 3	0.99	3.42*	7.43*	2, 4, 5, 8	0.96	4.45*	3.39*
2, 4, 8	0.99	3.28*	6.89*	2	0.97	4.33*	4.21*
2, 7	0.99	3.15*	6.88*	2, 5	0.97	3.99*	4.11*
2, 3, 6	0.99	3.12*	6.58*	2, 6	0.97	3.86*	3.50*
2, 3, 8	0.99	2.97*	8.32*	2, 4, 6, 8	0.97	3.56*	2.86*
Data-release time							
1985:01–2014:12				2005:01–2014:12			
Model	Ratio	MSE-F	ENC-F	Model	Ratio	MSE-F	ENC-F
2, 7	0.99	2.47*	3.08*	2, 8	0.98	2.51*	2.17*
2, 6, 7	1.00	1.51	2.69	2, 4, 8	0.98	2.36*	2.48*
2	1.00	1.47	2.25*	2, 7, 8	0.98	2.21*	1.79*
2, 7, 8	1.00	0.94	2.77	2, 4, 7, 8	0.98	2.15*	2.09*
2, 6	1.00	0.07	1.87	2	0.98	2.03*	1.77*
2, 6, 7, 8	1.00	0.07	2.41	2, 4	0.98	1.83*	2.05*
7	1.00	−0.09	0.80	2, 7	0.99	1.68*	1.38*
2, 3, 7	1.00	−0.10	2.50	2, 4, 7	0.99	1.58	1.66*
2, 4, 7	1.00	−0.13	2.28	2, 3, 8	0.99	1.26	1.90*
2, 8	1.00	−0.15	1.95	2, 3, 4, 8	0.99	1.17	2.21*

Note: The table reports empirical evidence regarding the use of factor-based models for predicting the equity premium at the one-month horizon. Only the top 10 models, by MSE, are reported. The top panel uses factor estimates dated by calendar time while the lower panel uses factor estimates dated by data-release time. Model denotes which factors are included in a given model. Ratio denotes MSE(factor)/MSE(benchmark).

* Denote significant at the 5% level.

with the 10 best-performing factor-based models. The results are reported in Table 2. The upper panel is associated with calendar time, while the lower is associated with data-release time. For each of the four sets of results we list (i) the factors in the model, (ii) the ratio of the factor-model MSE to the benchmark model MSE, and (iii) the *MSE-F* and *ENC-F* statistics. For each pairwise comparison between the benchmark and competing model, an asterisk is used to denote significance at the 5% level. This indicator of significance should be treated with care however since it does not account for the multiple testing problem created by our presentation of 255 pairwise comparisons.

The upper panel of Table 2 suggests the presence of predictability for a variety of factor-based models. Despite the fact that the gains in forecast accuracy are nominally small, both the *MSE-F* and *ENC-F* tests indicate significant improvements relative to the benchmark. The predictive content is largely concentrated in the second factor – a feature also documented in Bai (2010) and Batje and Menkhoff (2012). Over the shorter, more recent, out-of-sample period the gains in forecast accuracy are larger than before but still remain modest. Interestingly, it appears that the eighth factor has become increasingly important for forecasting the equity premium and, in fact, the best 6 models all include both the second and eighth factor.

All of that said, in the lower panel of Table 2 we find less predictive content, which suggests that the timing used to estimate factors can play an important role in determining their usefulness for forecasting. Over the longer out-of-sample period, exactly 6 of the 255 factor-based models have a nominally smaller MSE than the benchmark (note that the ratio is reported to only two decimals) and in only two instances is the improvement statistically significant based on either the $MSE-F$ or $ENC-F$. Over the shorter out-of-sample period, the evidence of predictability is a bit stronger. As in the upper panel, the second and eighth factors seem to contain predictive content, providing an improvement in forecast accuracy of about 2% that is significant at the 5% level regardless of which test statistic we consider. Unlike the upper panel, the seventh factor shows up in roughly half of the models and is one of the better predictors over the longer out-of-sample period. The seventh factor, which loads heavily on stock market variables, shows up in only one model under calendar based timing.

Overall, it seems the second factor, which loads heavily on interest rates, is the strongest and most consistent predictor of the equity premium. This is consistent with extensive empirical results in the literature on the ability of interest rates to forecast stock returns, see for example Hjalmarsson (2010) for evidence for developed countries. Even so, the eighth factor, which loads heavily on exchange rate variables, seems to be of increasing importance. Of course, one could certainly argue that the gains in forecast accuracy are small but, as is common in the financial forecasting literature, even small statistical improvements in forecast accuracy can provide large economic gains. On a final note, it is interesting to note that the first factor, which loads heavily on real activity and employment variables, never appears in Table 2. This is of independent interest since the first factor is the one most often used in the literature on factor-based forecasting.

7. Conclusion

Factor-based forecasting is increasingly common at central banks and in academic research. The accuracy of these forecasts is sometimes compared with other benchmark models using tests of equal accuracy and encompassing for nested models developed in Clark and McCracken (2001) and McCracken (2007). In this paper, we establish conditions under which these tests remain asymptotically valid when the factors are estimated recursively across forecast origins. While some of these conditions are straightforward generalizations of existing theoretical work on factors, others are new and of independent interest for applications that estimate factors recursively over time.

Simulation evidence supports the theoretical results showing that the size of the tests is typically near its nominal level despite finite sample estimation error in the factors. Perhaps not surprisingly, we find that this finite sample estimation error reduces power compared with a hypothetical case in which the factors can be observed directly. The paper concludes with an empirical application in which a large macroeconomic dataset is used to construct factors and forecast the equity premium. While the predictive content is not high, we find significant evidence that the second factor has consistently exhibited predictive content for the equity premium and that of late, the eighth factor has become increasingly more relevant.

Appendix

This appendix contains two sections. First, we provide several lemmas useful to prove the results in Section 4, followed by their proofs. Then we prove the results in Section 4.

A.1. Auxiliary lemmas

Our first set of results are instrumental in proving some of the auxiliary lemmas that follow. As in Bai (2003), we use the following notation:

$$\gamma_{l,s} = E\left(\frac{e_l' e_s}{N}\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(e_{il}e_{is}\right);$$

$$\zeta_{l,s} = \frac{1}{N}\sum_{i=1}^{N}\left(e_{il}e_{is} - E\left(e_{il}e_{is}\right)\right);$$

$$\eta_{l,s} = \frac{f_l' \Lambda' e_s}{N} = \frac{1}{N}\sum_{i=1}^{N} f_l'\lambda_i e_{is}, \text{ and } \xi_{l,s} = \frac{f_s'\Lambda' e_l}{N} = \frac{1}{N}\sum_{i=1}^{N} f_s'\lambda_i e_{il}.$$

Moreover, for $t = R, \ldots, T$, we let

$$\mathcal{U}_{t,T} \equiv u_{t+1}\left(\frac{T}{t}\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{t-1} z_s u_{s+1}\right),$$

and note that $\mathcal{U}_{t,T}$ is a martingale difference sequence array with respect to $\mathcal{F}^t$ given Assumption 5.

Lemma A.1. *Under Assumptions 1–6,*

(a) $\sup_t \left(\frac{1}{t}\sum_{l=1}^{t}\left(\sum_{s=1}^{t-1}\gamma_{l,s}u_{s+1}\right)^2\right) = O_P(1)$, where $\gamma_{l,s} \equiv \frac{1}{N}\sum_{i=1}^{N} E\left(e_{il}e_{is}\right)$.

(b) $\sup_t \left(\frac{1}{t}\sum_{l=1}^{t}\left(\sum_{s=1}^{t-1}\zeta_{l,s}u_{s+1}\right)^2\right) = O_P\left(\frac{T}{N}\right)$, where $\zeta_{l,s} \equiv \frac{1}{N}\sum_{i=1}^{N}\left(e_{il}e_{is} - E\left(e_{il}e_{is}\right)\right)$.

(c) $\sup_t \left(\frac{1}{t}\sum_{l=1}^{t}\left(\sum_{s=1}^{t-1}\eta_{l,s}u_{s+1}\right)^2\right) = O_P\left(\frac{T}{N}\right)$, where $\eta_{l,s} \equiv \frac{f_l'\Lambda' e_s}{N} = f_l'\frac{1}{N}\sum_{i=1}^{N}\lambda_i e_{is}$.

(d) $\sup_t \left(\frac{1}{t}\sum_{l=1}^{t}\left(\sum_{s=1}^{t-1}\xi_{l,s}u_{s+1}\right)^2\right) = O_P\left(\frac{T}{N}\right)$, where $\xi_{l,s} \equiv \eta_{s,l}$.

(e) $\sup_t \left(\frac{1}{t}\sum_{l=1}^{t}\left\|\sum_{s=1}^{t-1}\gamma_{l,s}z_s'\right\|^2\right) = O_P(1)$.

(f) $\sup_t \left(\frac{1}{t}\sum_{l=1}^{t}\left\|\sum_{s=1}^{t-1}\zeta_{l,s}z_s'\right\|^2\right) = O_P\left(\frac{T}{N}\right)$.

(g) $\sup_t \left(\frac{1}{t}\sum_{l=1}^{t}\left\|\sum_{s=1}^{t-1}\eta_{l,s}z_s'\right\|^2\right) = O_P\left(\frac{T}{N}\right)$.

Lemma A.2. *Under Assumptions 1–6,*

(a) $\sup_t E\left\|\frac{1}{t}\sum_{l=1}^{t} f_l\left(\sum_{s=1}^{t-1}\gamma_{l,s}u_{s+1}\right)\right\|^4 = O(1)$.

(b) $\sup_t E\left\|\frac{1}{t}\sum_{l=1}^{t} f_l\left(\sum_{s=1}^{t-1}\zeta_{l,s}u_{s+1}\right)\right\|^4 = O\left(\left(\frac{T}{N}\right)^2\right)$.

(c) $\sup_t E\left\|\frac{1}{t}\sum_{l=1}^{t} f_l\left(\sum_{s=1}^{t-1}\eta_{l,s}u_{s+1}\right)\right\|^4 = O\left(\left(\frac{T}{N}\right)^2\right)$.

(d) $\sup_t E\left\|\frac{1}{t}\sum_{l=1}^{t} f_l\left(\sum_{s=1}^{t-1}\xi_{l,s}u_{s+1}\right)\right\|^4 = O\left(\left(\frac{T}{N}\right)^2\right)$.

Lemma A.3. *Under Assumptions 1–6,*

(a) $\frac{1}{T}\sum_{t=R}^{T}\left\|\mathcal{U}_{t,T}\right\|^2 = O_P(1)$.

(b) $\sum_{t=R}^{T}\left(\frac{1}{t}\sum_{s=1}^{t} f_s\gamma_{s,t}\right)' A\,\mathcal{U}_{t,T} = O_P(1)$, where A is any $r \times (r+k_1)$ matrix of constants.

(c) $\sum_{t=R}^{T}\left(\frac{1}{t}\sum_{s=1}^{t} f_s\zeta_{s,t}\right)' A\,\mathcal{U}_{t,T} = O_P\left(\sqrt{\frac{T}{N}}\right)$.

(d) $\sum_{t=R}^{T}\left(\frac{1}{t}\sum_{s=1}^{t} f_s\eta_{s,t}\right)' A\,\mathcal{U}_{t,T} = O_P\left(\sqrt{\frac{T}{N}}\right)$.

(e) $\sum_{t=R}^{T}\left(\frac{1}{t}\sum_{s=1}^{t} f_s\xi_{s,t}\right)' A\,\mathcal{U}_{t,T} = O_P\left(\sqrt{\frac{T}{N}}\right)$.

Our next result provides bounds on the norms of $A_{1s,t}, A_{2s,t}, A_{3s,t}$, and $A_{4s,t}$, which are defined by the following equality given in Bai (2003). Specifically, for each $t = R, \ldots, T$ and $s = 1, \ldots, t$, we can write

$$\tilde{f}_{s,t} - H_t f_s = \tilde{V}_t^{-1}\left(\underbrace{\frac{1}{t}\sum_{l=1}^{t}\tilde{f}_{l,t}\gamma_{l,s}}_{\equiv A_{1s,t}} + \underbrace{\frac{1}{t}\sum_{l=1}^{t}\tilde{f}_{l,t}\zeta_{l,s}}_{\equiv A_{2s,t}} + \underbrace{\frac{1}{t}\sum_{l=1}^{t}\tilde{f}_{l,t}\eta_{l,s}}_{\equiv A_{3s,t}} + \underbrace{\frac{1}{t}\sum_{l=1}^{t}\tilde{f}_{l,t}\xi_{l,s}}_{\equiv A_{4s,t}}\right). \quad (9)$$

Lemma A.4. *Under Assumptions 1–6,*

(a) $\frac{1}{P}\sum_{t=R}^{T-1}\left\|A_{1t,t}\right\|^2 = O_P\left(\frac{1}{T}\right)$, *whereas for* $j = 2, 3, 4$, $\frac{1}{P}\sum_{t=R}^{T-1}\left\|A_{jt,t}\right\|^2 = O_P\left(\frac{1}{N}\right)$.

(b) $\sup_t \frac{1}{t}\sum_{s=1}^{t}\left\|A_{1s,t}\right\|^2 = O_P\left(\frac{1}{R}\right)$, *whereas for* $j = 2, 3, 4$, $\sup_t \frac{1}{t}\sum_{s=1}^{t}\left\|A_{js,t}\right\|^2 = O_P\left(\frac{1}{N}\right)$.

Lemma A.4 and part (a) of the following result (which provides uniform bounds on $\tilde{V}_t^{-1}$ over $t = R, \ldots, T$) are used to prove Theorem 4.1.

Lemma A.5. *Under Assumptions 1–6, we have that (a)* $\sup_t \left\|\tilde{V}_t^{-1}\right\| = O_P(1)$, *and (b)* $\sup_t \|H_t\|^q = O_P(1)$, *for any* $q > 0$.

Our next result derives the convergence rates of H_t, H_t^{-1}, and $\tilde{V}_t^{-1}$. Given the identification condition Assumption 1(a), we show that H_t converges to $H_{0t} = diag\,(\pm 1)$ at rate $O\left(1/\delta_{N,R}\right)$ uniformly over $t = R, \ldots, T$. A similar result holds for H_t^{-1} and for $\tilde{V}_t^{-1}$, which converge to H_{0t}^{-1} and to $V_0^{-1} \equiv \Sigma_\Lambda^{-1}$ at rate $O\left(1/\delta_{N,R}\right)$ uniformly over $t = R, \ldots, T$, respectively.

Lemma A.6. *Under Assumptions 1–6,*

(a) $\sup_t \|H_t - H_{0t}\| = \sup_t \left\|H_t^{-1} - H_{0t}^{-1}\right\| = O_P\left(1/\delta_{N,R}\right)$, *where* $H_{0t} = diag\,(\pm 1)$ *and* $\delta_{N,R} = \min\left(\sqrt{N}, \sqrt{R}\right)$.

(b) $\sup_t \left\|\tilde{V}_t^{-1} - V_0^{-1}\right\| = O_P\left(1/\delta_{N,R}\right)$, *where* $V_0 = \Sigma_\Lambda > 0$.

Next, we provide a result that is auxiliary in proving Lemma 4.1.

Lemma A.7. *Under Assumptions 1–6,*

(a) $\sup_t \left\| \frac{1}{t} \sum_{s=1}^{t-1} A_{1s,t} u_{s+1} \right\| = O_P\left(\frac{1}{R}\right)$ *whereas for* $j = 2, 3, 4$, $\sup_t \left\| \frac{1}{t} \sum_{s=1}^{t-1} A_{js,t} u_{s+1} \right\| = O_P\left(\frac{1}{\sqrt{RN}}\right)$.

(b) $\sup_t \left\| \frac{1}{t} \sum_{s=1}^{t-1} A_{1s,t} z_s' \right\| = O_P\left(\frac{1}{R}\right)$, *for* $j = 2, 3$, $\sup_t \left\| \frac{1}{t} \sum_{s=1}^{t-1} A_{js,t} z_s' \right\| = O_P\left(\frac{1}{\sqrt{RN}}\right)$, *and* $\sup_t \left\| \frac{1}{t} \sum_{s=1}^{t-1} A_{4s,t} z_s' \right\| = O_P\left(\frac{1}{\delta_{N,R}^2}\right)$.

To prove Lemma 4.1, we need to introduce some additional notation. Let

$$\hat{B}(t) = \left(t^{-1} \sum_{s=1}^{t-1} \tilde{z}_{s,t} \tilde{z}_{s,t}' \right)^{-1}, \text{ and } \hat{V}(t) = t^{-1} \sum_{s=1}^{t-1} \tilde{z}_{s,t} u_{s+1},$$

and note that $y_{s+1} = z_s'\delta + u_{s+1}$ given (6). Moreover, adding and subtracting appropriately it is also true that

$$y_{s+1} = \tilde{z}_{s,t}' \Phi_t'^{-1} \delta - \left(\tilde{z}_{s,t} - \Phi_t z_s \right)' \Phi_t'^{-1} \delta + u_{s+1}.$$

Therefore,

$$\hat{\delta}_t = \Phi_t'^{-1}\delta + \hat{B}(t)\hat{V}(t) - \hat{B}(t) t^{-1} \sum_{s=1}^{t-1} \tilde{z}_{s,t} \left(\tilde{z}_{s,t} - \Phi_t z_s \right)' \Phi_t'^{-1} \delta$$

$$\text{and } \ddot{\delta}_t = \delta + B(t) V(t),$$

which implies that

$$\hat{\delta}_t - \Phi_t'^{-1}\ddot{\delta}_t = \hat{B}(t)\hat{V}(t) - \Phi_t'^{-1} B(t) V(t) - \hat{B}(t) t^{-1} \sum_{s=1}^{t-1} \tilde{z}_{s,t} \left(\tilde{z}_{s,t} - \Phi_t z_s \right)' \Phi_t'^{-1} \delta.$$

Thus, we can write

$$\begin{aligned} \hat{\delta}_t - \Phi_t'^{-1}\ddot{\delta}_t &= \Phi_t'^{-1} B(t) \Phi_t^{-1} \left(\hat{V}(t) - \Phi_t V(t) \right) \\ &\quad + \left(\hat{B}(t) - \Phi_t'^{-1} B(t) \Phi_t^{-1} \right) \hat{V}(t) \\ &\quad - \hat{B}(t) t^{-1} \sum_{s=1}^{t-1} \tilde{z}_{s,t} \left(\tilde{z}_{s,t} - \Phi_t z_s \right)' \Phi_t'^{-1} \delta, \end{aligned} \tag{10}$$

where the first term captures the factors estimation error in the score process and the second term captures the factor estimation error in the (inverse) Hessian. The third term is equal to

$$-\hat{B}(t) t^{-1} \sum_{s=1}^{t-1} \tilde{z}_{s,t} \left(\tilde{f}_{s,t} - H_t f_s \right)' H_t'^{-1} \frac{\alpha}{\sqrt{T}},$$

given our assumption that there are no generated regressors in the benchmark model. This term disappears under the null when $\alpha = 0$ but is present under local alternatives.

The following result analyzes each term of $\hat{\delta}_t - \Phi_t'^{-1}\ddot{\delta}_t$.

Lemma A.8. *Suppose Assumptions 1–6 hold. Then,*

(a) $\sup_t \left\| \hat{V}(t) - \Phi_t V(t) \right\| = O_P\left(\frac{1}{\sqrt{R}\delta_{N,T}} \right)$.

(b) $\sup_t \left\| \hat{B}(t) - \Phi_t'^{-1} B(t) \Phi_t^{-1} \right\| = O_P\left(\frac{1}{\delta_{N,R}^2} \right)$.

(c) $\sup_t \left\| \hat{V}(t) \right\| = O_P\left(\frac{1}{\sqrt{T}} \right)$.

(d) $\sup_t \left\| t^{-1} \sum_{s=1}^{t-1} \tilde{z}_{s,t} \left(\tilde{z}_{s,t} - \Phi_t z_s \right)' \Phi_t'^{-1} \delta \right\| = O_P\left(\frac{1}{\sqrt{T}\delta_{N,T}} \right)$.

Our next result is useful to prove Lemma 4.2. Given (7), we can write

$$\begin{aligned} \sum_{t=R}^{T-1} u_{t+1} \left(\ddot{u}_{2,t+1} - \hat{u}_{2,t+1} \right) &= \sum_{t=R}^{T-1} u_{t+1} \left(\tilde{z}_{t,t} - \Phi_t z_t \right)' \Phi_t'^{-1} \left(\ddot{\delta}_t - \delta \right) \\ &\quad + \sum_{t=R}^{T-1} u_{t+1} \tilde{z}_{t,t}' \left(\hat{\delta}_t - \Phi_t'^{-1} \ddot{\delta}_t \right) \\ &\quad + \sum_{t=R}^{T-1} u_{t+1} \left(\tilde{z}_{t,t} - \Phi_t z_t \right)' \Phi_t'^{-1} \delta. \end{aligned} \tag{11}$$

We analyze each piece separately. The following lemma contains the results.

Lemma A.9. *Suppose Assumptions 1- 6 hold. If* $\sqrt{T}/N \to 0$, *then*

(a) $\sum_{t=R}^{T-1} u_{t+1} \left(\tilde{z}_{t,t} - \Phi_t z_t \right)' \Phi_t'^{-1} \left(\ddot{\delta}_t - \delta \right) = o_P(1)$.

(b) $\sum_{t=R}^{T-1} u_{t+1} \tilde{z}_{t,t}' \left(\hat{\delta}_t - \Phi_t'^{-1} \ddot{\delta}_t \right) = o_P(1)$.

(c) $\sum_{t=R}^{T-1} u_{t+1} \left(\tilde{z}_{t,t} - \Phi_t z_t \right)' \Phi_t'^{-1} \delta = o_P(1)$.

A.2. Proofs of auxiliary lemmas

Proof of Lemma A.1. We start with (a). Under Assumption 2(c), we can write

$$\sum_{s=1}^{t-1} \gamma_{l,s} u_{s+1} = \sum_{s=l-\tau}^{l+\tau} \gamma_{l,s} u_{s+1},$$

where we let $\gamma_{l,s} \equiv 0$ if $\min(l,s) \leq 0$ or $\max(l,s) > t$. By Cauchy–Schwarz,

$$\begin{aligned} \left(\sum_{s=1}^{t-1} \gamma_{l,s} u_{s+1} \right)^2 &= \left(\sum_{s=l-\tau}^{l+\tau} \gamma_{l,s} u_{s+1} \right)^2 \leq \sum_{s=l-\tau}^{l+\tau} \gamma_{l,s}^2 \sum_{s=l-\tau}^{l+\tau} u_{s+1}^2 \\ &\leq (2\tau + 1) \max_{l,s} \left| \gamma_{l,s}^2 \right| \sum_{s=l-\tau}^{l+\tau} u_{s+1}^2 \leq M \sum_{s=l-\tau}^{l+\tau} u_{s+1}^2, \end{aligned}$$

for some constant M, given Assumption 2(c), implying that

$$\begin{aligned} \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left(\sum_{s=1}^{t-1} \gamma_{l,s} u_{s+1} \right)^2 \right) &\leq M \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \sum_{s=l-\tau}^{l+\tau} u_{s+1}^2 \right) \\ &= M \frac{1}{R} \sum_{l=1}^{T-1} \sum_{s=l-\tau}^{l+\tau} u_{s+1}^2 = O_P(1) \end{aligned}$$

provided $E\left(u_{s+1}^2\right) \leq M$, which follows under Assumption 6(a).

Next, consider part (b). We have that

$$\begin{aligned} &\sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left(\sum_{s=1}^{t-1} \zeta_{l,s} u_{s+1} \right)^2 \right) \\ &= \frac{1}{N} \sup_t \left(\frac{T}{t} \sum_{l=1}^{t} \left(\frac{1}{\sqrt{NT}} \sum_{s=1}^{t-1} \sum_{i=1}^{N} \left(e_{il} e_{is} - E\left(e_{il} e_{is}\right) \right) u_{s+1} \right)^2 \right) \\ &= \frac{T}{N} \frac{T}{R} \left(\frac{1}{T} \sum_{l=1}^{T} \sup_t \left(\frac{1}{\sqrt{NT}} \sum_{s=1}^{t-1} \sum_{i=1}^{N} \left(e_{il} e_{is} - E\left(e_{il} e_{is}\right) \right) u_{s+1} \right)^2 \right) \\ &= O_P\left(\frac{T}{N} \right) \end{aligned}$$

under Assumption 3(b).

For (c),

$$\sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left(\sum_{s=1}^{t-1} \eta_{l,s} u_{s+1} \right)^2 \right)$$
$$= \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left(f_l' \sum_{s=1}^{t-1} \frac{1}{N} \sum_{i=1}^{N} \lambda_i e_{is} u_{s+1} \right)^2 \right)$$
$$\leq \frac{T}{N} \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \|f_l\|^2 \left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{t-1} \sum_{i=1}^{N} \lambda_i e_{is} u_{s+1} \right\|^2 \right)$$
$$\leq \frac{T}{N} \underbrace{\sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \|f_l\|^2 \right)}_{=O_P(1)} \underbrace{\sup_t \left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{t-1} \sum_{i=1}^{N} \lambda_i e_{is} u_{s+1} \right\|^2}_{=O_P(1)}$$
$$= O_P\left(\frac{T}{N}\right),$$

given in particular Assumption 3(c).

Finally, for (d),

$$\sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left(\sum_{s=1}^{t-1} \xi_{l,s} u_{s+1} \right)^2 \right)$$
$$= \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left(\sum_{s=1}^{t-1} \frac{1}{N} \sum_{i=1}^{N} f_s' \lambda_i e_{il} u_{s+1} \right)^2 \right)$$
$$= \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i' e_{il} \sum_{s=1}^{t-1} f_s u_{s+1} \right)^2 \right)$$
$$\leq \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left\| \frac{1}{N} \sum_{i=1}^{N} \lambda_i' e_{il} \right\|^2 \left\| \sum_{s=1}^{t-1} f_s u_{s+1} \right\|^2 \right)$$
$$= \underbrace{\frac{1}{R} \sum_{l=1}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i' e_{il} \right\|^2}_{=O_P(1)} \left(\frac{T}{N}\right) \underbrace{\sup_t \left\| \frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} f_s u_{s+1} \right\|^2}_{=O_P(1)}$$
$$= O_P\left(\frac{T}{N}\right),$$

given Assumptions 2(e) and 6(b).

Part (e) follows exactly as the proof of (a) given Assumption 2(c) and the fact that $E\|z_s\|^2 \leq M$ by Assumption 6(a).

Next, consider part (f). We have that

$$\sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left\| \sum_{s=1}^{t-1} \zeta_{l,s} z_s' \right\|^2 \right)$$
$$= \frac{T}{N} \frac{T}{R} \left(\frac{1}{T} \sum_{l=1}^{T} \sup_t \left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{t-1} \sum_{i=1}^{N} (e_{il} e_{is} - E(e_{il} e_{is})) z_s' \right\|^2 \right)$$
$$= O_P\left(\frac{T}{N}\right)$$

under the analog of Assumption 3(b) with u_{s+1} replaced with z_s' (cf. Assumption 3(d)).

For (g),

$$\sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left\| \sum_{s=1}^{t-1} \eta_{l,s} z_s' \right\|^2 \right)$$
$$= \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left\| f_l' \sum_{s=1}^{t-1} \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i e_{is} \right) z_s' \right\|^2 \right)$$
$$\leq \frac{T}{N} \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \|f_l\|^2 \left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{t-1} \sum_{i=1}^{N} \lambda_i e_{is} z_s' \right\|^2 \right)$$
$$\leq \frac{T}{N} \underbrace{\sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \|f_l\|^2 \right)}_{=O_P(1)} \underbrace{\sup_t \left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{t-1} \sum_{i=1}^{N} \lambda_i e_{is} z_s' \right\|^2}_{=O_P(1)}$$
$$= O_P\left(\frac{T}{N}\right),$$

given in particular Assumption 3(e) (the analog of 3(c)). ■

Proof of Lemma A.2. Starting with (a), by the Cramer–Rao and Cauchy–Schwarz inequalities,

$$E\left\| \frac{1}{t} \sum_{l=1}^{t} f_l \left(\sum_{s=1}^{t-1} \gamma_{l,s} u_{s+1} \right) \right\|^4 \leq \frac{1}{t} \sum_{l=1}^{t} E\left(\|f_l\|^4 \left\| \sum_{s=1}^{t-1} \gamma_{l,s} u_{s+1} \right\|^4 \right)$$
$$\leq \frac{1}{t} \sum_{l=1}^{t} \left(E\|f_l\|^8 \right)^{1/2} \left(E\left\| \sum_{s=1}^{t-1} \gamma_{l,s} u_{s+1} \right\|^8 \right)^{1/2}$$
$$\leq \left(\frac{1}{t} \sum_{l=1}^{t} E\|f_l\|^8 \right)^{1/2} \left(\frac{1}{t} \sum_{l=1}^{t} E\left\| \sum_{s=1}^{t-1} \gamma_{l,s} u_{s+1} \right\|^8 \right)^{1/2}.$$

The first term in parenthesis is bounded given the moment conditions on f_l. For the second term, by the moving average-type assumption on $\gamma_{l,s}$, we have that for given l,

$$E\left\| \sum_{s=1}^{t-1} \gamma_{l,s} u_{s+1} \right\|^8 = E\left\| \sum_{s=l-\tau}^{l+\tau} \gamma_{l,s} u_{s+1} \right\|^8$$
$$\leq (2\tau+1)^7 \sum_{s=l-\tau}^{l+\tau} \gamma_{l,s}^8 E\left(u_{s+1}^8\right) \leq M,$$

given that $|\gamma_{l,s}| \leq M$ and $E\left(u_{s+1}^8\right) \leq M$ by Assumptions 2(c) and 6(a).

For part (b), following exactly the same steps as for (a), we have that

$$\frac{1}{t} \sum_{l=1}^{t} E\left\| \sum_{s=1}^{t-1} \zeta_{l,s} u_{s+1} \right\|^8$$
$$= \frac{1}{t} \sum_{l=1}^{t} E\left\| \sum_{s=1}^{t-1} \frac{1}{N} \sum_{i=1}^{N} (e_{il} e_{is} - E(e_{il} e_{is})) u_{s+1} \right\|^8$$
$$\leq \left(\frac{T}{N}\right)^4 \frac{1}{R} \sum_{l=1}^{T} \sup_{R \leq t \leq T} E\left\| \frac{1}{\sqrt{TN}} \sum_{s=1}^{t-1} \sum_{i=1}^{N} (e_{il} e_{is} - E(e_{il} e_{is})) u_{s+1} \right\|^8,$$

which is $O\left(\left(\frac{T}{N}\right)^4\right)$ given Assumption 4(a).

For part (c),

$$\sup_t E\left\| \frac{1}{t} \sum_{l=1}^{t} f_l \left(\sum_{s=1}^{t-1} \eta_{l,s} u_{s+1} \right) \right\|^4$$
$$= \sup_t E\left\| \frac{1}{t} \sum_{l=1}^{t} f_l f_l' \frac{1}{N} \sum_{s=1}^{t-1} \sum_{i=1}^{N} \lambda_i e_{is} u_{s+1} \right\|^4$$

$$\leq \sup_t \frac{1}{t}\sum_{l=1}^{t} E\left(\|f_l\|^8 \left\|\frac{1}{N}\sum_{s=1}^{t-1}\sum_{i=1}^{N}\lambda_i e_{is}u_{s+1}\right\|^4\right)$$

$$\leq \sup_l E\left(\|f_l\|^{16}\right)^{1/2}\left(\frac{T^4}{N^4}\sup_t E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{t-1}\sum_{i=1}^{N}\lambda_i e_{is}u_{s+1}\right\|^8\right)^{1/2}$$

$$= O\left(\left(\frac{T}{N}\right)^2\right),$$

given Assumptions 1(a) and 4(b).

Finally, for part (d), we have that

$$\sup_t E\left\|\frac{1}{t}\sum_{l=1}^{t} f_l\left(\sum_{s=1}^{t-1}\xi_{l,s}u_{s+1}\right)\right\|^4$$

$$= \sup_t E\left\|\frac{1}{t}\sum_{l=1}^{t} f_l\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i' e_{il}\sum_{s=1}^{t-1} f_s u_{s+1}\right)\right\|^4$$

$$\leq \frac{T^2}{N^2}\sup_t \frac{1}{t}\sum_{l=1}^{t} E\left(\|f_l\|^4\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{il}\right\|^4\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{t-1} f_s u_{s+1}\right\|^4\right)$$

$$\leq \frac{T^2}{N^2}\sup_t \frac{1}{t}\sum_{l=1}^{t}\left(E\left(\|f_l\|^8\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{il}\right\|^8\right)\right)^{1/2}$$

$$\times\left(E\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{t-1} f_s u_{s+1}\right\|^8\right)^{1/2}$$

$$\leq \frac{T^2}{N^2}\left(\sup_l E\|f_l\|^{16}\right)^{1/4}\left(\sup_l E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{il}\right\|^{16}\right)^{1/4}$$

$$\times\left(\sup_t E\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{t-1} f_s u_{s+1}\right\|^8\right)^{1/2} = O\left(\left(\frac{T}{N}\right)^2\right),$$

given Assumptions 1(a), 2(e) and 6(b). ■

Proof of Lemma A.3. For part (a), note that by definition of $\mathcal{U}_{t,T}$,

$$\frac{1}{T}\sum_{t=R}^{T}\left\|\mathcal{U}_{t,T}\right\|^2 \leq \frac{1}{T}\sum_{t=R}^{T}\|u_{t+1}\|^2 (T/R)^2 \sup_t\left\|T^{-1/2}\sum_{s=1}^{t-1} z_s u_{s+1}\right\|^2 = O_P(1)$$

given Assumptions 6(a) and (b).

For part (b), it suffices to show that

$$E\left\|\sum_{t=R}^{T}\left(\frac{1}{t}\sum_{l=1}^{t} f_l\gamma_{l,t}\right)' A\mathcal{U}_{t,T}\right\| = O(1).$$

Using the triangle and Cauchy–Schwarz inequalities,

$$E\left\|\sum_{t=R}^{T}\left(\frac{1}{t}\sum_{l=1}^{t} f_l\gamma_{l,t}\right)' A\mathcal{U}_{t,T}\right\|$$

$$\leq \|A\|\, E\left(\sum_{t=R}^{T}\left\|\mathcal{U}_{t,T}\right\|\frac{1}{t}\sum_{l=1}^{t}\|f_l\|\left|\gamma_{l,t}\right|\right)$$

$$\leq \|A\|\frac{1}{R}\sum_{t=R}^{T}\sum_{l=1}^{T} E\left\|\mathcal{U}_{t,T}\right\|\|f_l\|\left|\gamma_{l,t}\right|$$

$$\leq \|A\|\frac{1}{R}\sum_{t=R}^{T}\sum_{l=1}^{T}\left(E\left\|\mathcal{U}_{t,T}\right\|^2\right)^{1/2}\left(E\|f_l\|^2\right)^{1/2}\left|\gamma_{l,t}\right|$$

$$\leq \|A\|\left(\sup_t E\left\|\mathcal{U}_{t,T}\right\|^2\right)^{1/2}\left(\sup_l E\|f_l\|^2\right)^{1/2}\frac{1}{R}\sum_{t=R}^{T}\sum_{l=1}^{T}\left|\gamma_{l,t}\right|$$

$$= O(1),$$

provided (i) $\sup_t E\left\|\mathcal{U}_{t,T}\right\|^2 = O(1)$, (ii) $\sup_l E\|f_l\|^2 = O(1)$, and (iii) $\frac{1}{R}\sum_{t=R}^{T}\sum_{l=1}^{T}\left|\gamma_{l,t}\right| = O(1)$. (ii) follows by Assumption 1(a) and (iii) follows by Assumption 2(c). Next, we show that (i) follows under our assumptions. By definition of $\mathcal{U}_{t,T}$,

$$\sup_t E\left\|\mathcal{U}_{t,T}\right\|^2 \leq \sup_t E\left\|u_{t+1}\left(\frac{T}{t}\right)\left(T^{-1/2}\sum_{s=1}^{t-1} z_s u_{s+1}\right)\right\|^2$$

$$\leq (T/R)^2\sup_t E\left(u_{t+1}^2\left\|T^{-1/2}\sum_{s=1}^{t-1} z_s u_{s+1}\right\|^2\right)$$

$$\leq (T/R)^2\left(\sup_t E\left(u_{t+1}^4\right)\right)^{1/2}\left(\sup_t E\left\|T^{-1/2}\sum_{s=1}^{t-1} z_s u_{s+1}\right\|^4\right)^{1/2},$$

implying that $\sup_{R\leq t\leq T} E\left\|\mathcal{U}_{t,T}\right\|^2 = O(1)$ since $\sup_t E\left(u_{t+1}^4\right) \leq M$ and $\sup_t E\left\|T^{-1/2}\sum_{s=1}^{t-1} z_s u_{s+1}\right\|^4 \leq M$ under Assumptions 6(a) and (b).

For part (c), we can write

$$\sum_{t=R}^{T}\left(\frac{1}{t}\sum_{l=1}^{t} f_l\zeta_{l,t}\right)' A\mathcal{U}_{t,T} = \sum_{t=R}^{T} u_{t+1}\mathcal{W}_{t,T},$$

where

$$\mathcal{W}_{t,T} = \left(\frac{T}{t}\right)\left(\frac{1}{t}\frac{1}{N}\sum_{l=1}^{t}\sum_{i=1}^{N} f_l\left(e_{il}e_{it} - E\left(e_{il}e_{it}\right)\right)\right)'$$

$$\times A\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{t-1} z_s u_{s+1}\right)$$

is measurable with respect to $\mathcal{F}^t$. Thus, $E\left(u_{t+1}\mathcal{W}_{t,T}|\mathcal{F}^t\right) = 0$ given the martingale difference Assumption 5 and it follows that

$$Var\left(\sum_{t=R}^{T} u_{t+1}\mathcal{W}_{t,T}\right) = \sum_{t=R}^{T} E\left(u_{t+1}^2\mathcal{W}_{t,T}^2\right)$$

$$\leq \underbrace{\left(\sum_{t=R}^{T} E\left(u_{t+1}^4\right)\right)^{1/2}}_{=O(\sqrt{T})}\underbrace{\left(\sum_{t=R}^{T} E\left(\mathcal{W}_{t,T}^4\right)\right)^{1/2}}_{=O\left(\left(\frac{T}{N^2}\right)^{1/2}\right)} = O\left(\frac{T}{N}\right),$$

which suffices to prove the result. Note that

$$\sup_t E\left\|\mathcal{W}_{t,T}\right\|^4$$

$$= \sup_t\left(\frac{T}{t}\right)^4 E\left\|\left(\frac{1}{t}\frac{1}{N}\sum_{l=1}^{t}\sum_{i=1}^{N} f_l\left(e_{il}e_{it} - E\left(e_{il}e_{it}\right)\right)\right)'\right.$$

$$\left.\times A\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{t-1} z_s u_{s+1}\right)\right\|^4$$

$$\leq \left(\frac{T}{R}\right)^4\left(\sup_t E\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{t-1} z_s u_{s+1}\right\|^8\right)^{1/2}$$

$$\times\sup_t\left(E\left\|\frac{1}{t}\frac{1}{N}\sum_{l=1}^{t}\sum_{i=1}^{N} f_l\left(e_{il}e_{it} - E\left(e_{il}e_{it}\right)\right)\right\|^8\right)^{1/2}\|A\|^4$$

$$= O\left(1/N^2\right),$$

where we have used Assumption 6(b) to bound the first term in the square root parentheses. The second term can be bounded by

$$\sup_t \left(\frac{1}{t} \sum_{l=1}^{t} E \|f_l\|^8 \left\| \frac{1}{N} \sum_{i=1}^{N} (e_{il}e_{it} - E(e_{il}e_{it})) \right\|^8 \right)^{1/2}$$
$$\leq \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left(E \|f_l\|^{16}\right)^{1/2} \times \left(E \left\| \frac{1}{N} \sum_{i=1}^{N} (e_{il}e_{it} - E(e_{il}e_{it})) \right\|^{16} \right)^{1/2} \right)^{1/2}$$
$$\leq \sup_t \left(\frac{1}{t} \sum_{l=1}^{t} E \|f_l\|^{16} \right)^{1/4} \times \left(\frac{1}{t} \sum_{l=1}^{t} E \left\| \frac{1}{N} \sum_{i=1}^{N} (e_{il}e_{it} - E(e_{il}e_{it})) \right\|^{16} \right)^{1/4}$$
$$\leq \left(\sup_l E \|f_l\|^{16} \right)^{1/4} \times \left(\frac{1}{N^8} \sup_{l,t} E \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (e_{il}e_{it} - E(e_{il}e_{it})) \right\|^{16} \right)^{1/4}$$
$$= O\left(\frac{1}{N^2}\right),$$

given Assumptions 1(a) and 2(d).

Next, consider part (d). By replacing $\eta_{l,t}$ and $\mathcal{U}_{t,T}$, we can write

$$\sum_{t=R}^{T} \left(\frac{1}{t} \sum_{l=1}^{t} f_l \eta_{l,t} \right)' A \mathcal{U}_{t,T} = \sum_{t=R}^{T} \left(\frac{1}{t} \sum_{l=1}^{t} f_l f_l' \frac{1}{N} \sum_{i=1}^{N} \lambda_i e_{it} \right)' A u_{t+1} \left(\frac{T}{t}\right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} z_s u_{s+1} \right).$$

Adding and subtracting appropriately, we can decompose the term above as the sum of two terms,

$$\chi_1 \equiv \sum_{t=R}^{T} \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i e_{it} \right)' \left(\frac{1}{t} \sum_{l=1}^{t} f_l f_l' - I_r \right)' \times A u_{t+1} \left(\frac{T}{t}\right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} z_s u_{s+1} \right)$$

and

$$\chi_2 = \sum_{t=R}^{T} \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i e_{it} \right)' A u_{t+1} \left(\frac{T}{t}\right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} z_s u_{s+1} \right).$$

We can bound χ_1 by

$$\|\chi_1\| \leq \left(\frac{T}{R}\right) \|A\| \sup_t \left\| \frac{1}{t} \sum_{l=1}^{t} f_l f_l' - I_r \right\| \sup_t \left\| \frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} z_s u_{s+1} \right\| \times \sum_{t=R}^{T} \|u_{t+1}\| \left\| \frac{1}{N} \sum_{i=1}^{N} \lambda_i e_{it} \right\|$$
$$\leq O_P\left(\frac{1}{\sqrt{T}}\right) T \left(\frac{1}{T} \sum_{t=R}^{T} \|u_{t+1}\|^2 \right)^{1/2} \times \left(\frac{1}{N} \frac{1}{T} \sum_{t=R}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} \right\|^2 \right)^{1/2}$$
$$= O_P\left(\sqrt{\frac{T}{N}} \right),$$

given Assumptions 1(a) and (b), 2(e), and 6(a). For χ_2, we can write

$$\chi_2 = \sum_{t=R}^{T} u_{t+1} \chi_{t,T},$$

where

$$\chi_{t,T} = \left(\frac{T}{t}\right) \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i e_{it} \right)' A \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} z_s u_{s+1} \right)$$

is measurable with respect to $\mathcal{F}^t$. Since $E\left(u_{t+1}\chi_{t,T}|\mathcal{F}^t\right) = 0$ given the martingale difference Assumption 5, it follows that

$$Var\left(\sum_{t=R}^{T} u_{t+1}\chi_{t,T} \right) = \sum_{t=R}^{T} E\left(u_{t+1}^2 \chi_{t,T}^2\right) \leq \underbrace{\left(\sum_{t=R}^{T} E\left(u_{t+1}^4\right) \right)^{1/2}}_{=O(\sqrt{T})} \underbrace{\left(\sum_{t=R}^{T} E\left(\chi_{t,T}^4\right) \right)^{1/2}}_{=O\left(\left(\frac{T}{N^2}\right)^{1/2}\right)} = O\left(\frac{T}{N}\right),$$

where

$$\sup_t E \left\|\chi_{t,T}\right\|^4 = \sup_t \left(\frac{T}{t}\right)^4 E \left\| \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i e_{it} \right)' A \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} z_s u_{s+1} \right) \right\|^4$$
$$\leq \left(\frac{T}{R}\right)^4 \|A\| \left(\sup_t E \left\| \frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} z_s u_{s+1} \right\|^8 \right)^{1/2} \times \sup_t \left(\frac{1}{N^4} E \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} \right\|^8 \right)^{1/2}$$
$$= O\left(1/N^2\right),$$

given our assumptions. This proves that $\|\chi_2\| = O_P\left(\sqrt{\frac{T}{N}}\right)$ and concludes the proof of part (d).

Finally, to prove part (e), by replacing $\xi_{l,t}$ and $\mathcal{U}_{t,T}$, we can write

$$\sum_{t=R}^{T} \left(\frac{1}{t} \sum_{l=1}^{t} f_l \xi_{l,t} \right)' A \mathcal{U}_{t,T} = \sum_{t=R}^{T} f_t' u_{t+1} \left(\frac{T}{t}\right) \times \left(\frac{1}{t} \sum_{l=1}^{t} \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i e_{il} \right) f_l' \right) A \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} z_s u_{s+1} \right)$$
$$= \sum_{t=R}^{T} u_{t+1} f_t' \varsigma_{t,T},$$

where

$$\varsigma_{t,T} \equiv \left(\frac{T}{t}\right) \left(\frac{1}{t} \sum_{l=1}^{t} \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i e_{il} \right) f_l' \right) A \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} z_s u_{s+1} \right)$$

is measurable with respect to $\mathcal{F}^t$. Since $E\left(u_{t+1}f_t'\varsigma_{t,T}|\mathcal{F}^t\right)=0$, we can use the above argument to show that $Var\left(\sum_{t=R}^{T}v_{t+1}\varsigma_{t,T}\right)=O\left(\frac{T}{N}\right)$, given our assumptions. ■

Proof of Lemma A.4. Part (a). For $j=1,\ldots,4$, let

$$I_j\equiv\frac{1}{P}\sum_{t=R}^{T-1}\left\|A_{jt,t}\right\|^2\quad\text{with }A_{jt,t}=\frac{1}{t}\sum_{l=1}^{t}\tilde{f}_{l,t}\phi_{j,l,t},$$

where $\phi_{1,l,t}=\gamma_{l,t}$, $\phi_{2,l,t}=\zeta_{l,t}$, $\phi_{3,l,t}=\eta_{l,t}$ and $\phi_{4,l,t}=\xi_{l,t}$. We consider each term separately. Starting with the first term,

$$I_1=\frac{1}{P}\sum_{t=R}^{T-1}\left\|\frac{1}{t}\sum_{l=1}^{t}\tilde{f}_{l,t}\gamma_{l,t}\right\|^2\leq\frac{1}{P}\sum_{t=R}^{T-1}\left(\frac{1}{t}\sum_{l=1}^{t}\left\|\tilde{f}_{l,t}\right\|^2\right)\left(\frac{1}{t}\sum_{l=1}^{t}\gamma_{l,t}^2\right)$$
$$\leq r\frac{1}{PR}\sum_{t=1}^{T}\sum_{l=1}^{T}\gamma_{l,t}^2=O\left(1/T\right)$$

under Assumption 2(c) and given that $T/R=O(1)$. Next,

$$I_2=\frac{1}{P}\sum_{t=R}^{T-1}\left\|\frac{1}{t}\sum_{l=1}^{t}\tilde{f}_{l,t}\zeta_{l,t}\right\|^2\leq\frac{1}{P}\sum_{t=R}^{T-1}\left(\frac{1}{t}\sum_{l=1}^{t}\left\|\tilde{f}_{l,t}\right\|^2\right)\left(\frac{1}{t}\sum_{l=1}^{t}\zeta_{l,t}^2\right)$$
$$\leq r\frac{1}{PR}\sum_{t=1}^{T}\sum_{l=1}^{T}\zeta_{l,t}^2=O_P\left(1/N\right),$$

given Assumption 2(d). I_3 and I_4 follow similarly using the fact that

$$\frac{1}{PR}\sum_{t=1}^{T}\sum_{l=1}^{T}\eta_{l,t}^2=\frac{1}{PR}\sum_{t=1}^{T}\sum_{l=1}^{T}\xi_{l,t}^2=O_P\left(1/N\right).$$

Part (b). The proof follows the same argument as part (a) and is therefore omitted. ■

Proof of Lemma A.5. We start with (a). Recall that $\tilde{V}_t$ denotes the eigenvalue matrix of X_tX_t'/tN for $t=R,\ldots,T$, with its eigenvalues arranged in a non-increasing order: $\tilde{v}_{1,t}\geq\tilde{v}_{2,t}\geq\cdots\geq\tilde{v}_{r,t}$. It follows that $\tilde{V}_t^{-1}=diag\left(\tilde{v}_{j,t}^{-1}\right)$, where $\tilde{v}_{1,t}^{-1}\leq\tilde{v}_{2,t}^{-1}\leq\cdots\leq\tilde{v}_{r,t}^{-1}$. Hence, $\left\|\tilde{V}_t^{-1}\right\|^2=tr\left(\tilde{V}_t^{-2}\right)=\sum_{j=1}^{r}\tilde{v}_{j,t}^{-2}\leq r\tilde{v}_{r,t}^{-2}$, where $\tilde{v}_{r,t}$ is the rth largest eigenvalue of X_tX_t'/tN. Since

$$\sup_t\left\|\tilde{V}_t^{-1}\right\|\leq r^{1/2}\sup_t\left|\tilde{v}_{r,t}^{-1}\right|=r^{1/2}\frac{1}{\inf_t\left|\tilde{v}_{r,t}\right|},$$

it suffices to show that $\inf_t\left|\tilde{v}_{r,t}\right|\geq C>0$ for some constant C, with probability approaching 1. We use an argument similar to that used by Fan et al. (2013, cf. Lemma C.4). First, note that for $t=R,\ldots,T$ and $s=1,\ldots,t$, $x_s=\Lambda f_s+e_s$, where $x_s=(x_{1s},\ldots,x_{Ns})'$ is $N\times 1$ and a similar definition applies to e_s. This implies that the $N\times N$ covariance matrix of x_s is equal to

$$\Sigma\equiv E\left(x_sx_s'\right)=\Lambda E\left(f_sf_s'\right)\Lambda'+E\left(e_se_s'\right)=\Lambda\Lambda'+\Sigma_e,$$

or equivalently, $\frac{1}{N}\Sigma=\frac{1}{N}\Lambda\Lambda'+\frac{1}{N}\Sigma_e$, given that $E\left(f_se_{is}\right)=0$ (by Assumption 3(a)) and the identification condition Assumption 1. Given this assumption, the $r\times r$ matrix $\frac{\Lambda'\Lambda}{N}$ is diagonal with distinct and positive eigenvalues on the main diagonal. Since $\Lambda'\Lambda$ and $\Lambda\Lambda'$ share the same nonzero eigenvalues, it follows that $\Lambda\Lambda'$ has r nonzero eigenvalues, with the remaining eigenvalues equal to 0. By Proposition 1 of Fan et al. (2013) (which relies on Weyl's eigenvalue theorem), and given Assumption 1, we can then show that for $j=1,\ldots,r$, the difference between the jth eigenvalue of Σ/N and the jth eigenvalue of $\Lambda\Lambda'/N$ is bounded by the maximum eigenvalue of Σ_e/N; that is,

$$\left|\lambda_j\left(\Sigma/N\right)-\lambda_j\left(\Lambda'\Lambda/N\right)\right|\leq\frac{1}{N}\left\|\Sigma_e\right\|_1=O\left(\frac{1}{N}\right)=o\left(1\right),$$

given Assumption 2(b). By a second application of Weyl's Theorem (cf. Fan et al., 2013, Lemma 1), for each $j=1,\ldots,r$,

$$\left|\tilde{v}_{j,t}-\lambda_j\left(\Lambda'\Lambda/N\right)\right|\leq\left\|X_t'X_t/tN-N^{-1}\Sigma\right\|_1,$$

given that $X_t'X_t$ and X_tX_t' share the same nonzero eigenvalues. Thus,

$$\tilde{v}_{r,t}\geq\lambda_r\left(\Lambda'\Lambda/N\right)-\left\|X_t'X_t/tN-N^{-1}\Sigma\right\|_1$$
$$\geq\lambda_{\min}\left(\Sigma_\Lambda\right)/2-\left\|X_t'X_t/tN-N^{-1}\Sigma\right\|_1,$$

since $\left|\lambda_j\left(\Sigma/N\right)-\lambda_j\left(\Lambda'\Lambda/N\right)\right|=o\left(1\right)$ and the fact that $\left|\lambda_j\left(\Lambda'\Lambda/N\right)-\lambda_j\left(\Sigma_\Lambda\right)\right|=o\left(1\right)$. It follows that

$$\inf_t\left|\tilde{v}_{r,t}\right|\geq\lambda_{\min}\left(\Sigma_\Lambda\right)/2-\sup_t\left\|X_t'X_t/tN-N^{-1}\Sigma\right\|_1,$$

and hence it suffices to show that $\sup_t\left\|X_t'X_t/tN-N^{-1}\Sigma\right\|_1=o_P\left(1\right)$. But $\frac{X_t'X_t}{tN}-\frac{\Sigma}{N}=D_{1t}+D_{2t}+D_{3t}+D_{4t}$, where

$$D_{1t}=\frac{\Lambda}{N}\left(\frac{1}{t}\sum_{s=1}^{t}f_sf_s'-I_r\right)\Lambda',\quad D_{2t}=\frac{1}{N}\left(\frac{1}{t}\sum_{s=1}^{t}e_se_s'-\Sigma_e\right),$$

$$D_{3t}=\frac{\Lambda}{N}\frac{1}{t}\sum_{s=1}^{t}f_se_s',\text{ and }D_{4t}=D_{3t}',$$

where we recall that $E\left(f_sf_s'\right)=I_r$ by assumption. It follows that

$$\sup_t\left\|X_t'X_t/tN-N^{-1}\Sigma\right\|_1\leq\sup_t\left\|D_{1t}\right\|+\sup_t\left\|D_{2t}\right\|+2\sup_t\left\|D_{3t}\right\|,$$

since $\|A\|_1\leq\|A\|$ for any matrix A (see Horn and Johnson, 1985, p. 314). Starting with D_{1t}, we have that

$$\sup_t\left\|D_{1t}\right\|=\underbrace{\left\|\Lambda/\sqrt{N}\right\|^2}_{=tr\left(\Lambda'\Lambda/N\right)=O(1)}\sup_t\left\|\left(\frac{T}{t}\right)\left(\frac{1}{T}\sum_{s=1}^{t}\left(f_sf_s'-I_r\right)\right)\right\|$$
$$\leq C\cdot\frac{T}{R}\sup_t\left\|\frac{1}{T}\sum_{s=1}^{t}\left(f_sf_s'-I_r\right)\right\|$$

for some constant C. Since $T/R=O\left(1\right)$ by assumption, it follows that $\sup_t\left\|D_{1t}\right\|=o_P\left(1\right)$ given that $\sup_t\left\|\frac{1}{T}\sum_{s=1}^{t}\left(f_sf_s'-I_r\right)\right\|=O_P\left(1/\sqrt{T}\right)=o_P\left(1\right)$ by Assumption 1. Next, consider D_{2t}. We have that for any $\varepsilon>0$,

$$P\left(\sup_t\left\|D_{2t}\right\|>\varepsilon\right)\leq\sum_{t=R}^{T}P\left(\left\|D_{2t}\right\|>\varepsilon\right)\leq\frac{1}{\varepsilon^2}\sum_{t=R}^{T}E\left\|D_{2t}\right\|^2,$$

thus it suffices to prove that $\sum_{t=R}^{T}E\left\|D_{2t}\right\|^2=o\left(1\right)$. By the definition of the Frobenius norm, $\|A\|^2=tr\left(A'A\right)=\sum_{i,j=1}^{N}a_{ij}^2$ for any $N\times N$ matrix. Thus, for given $t=R,\ldots,T$,

$$E\left\|D_{2t}\right\|^2=\frac{1}{N^2}\sum_{i,j=1}^{N}E\left(\frac{1}{t}\sum_{s=1}^{t}\left(e_{is}e_{js}-E\left(e_{is}e_{js}\right)\right)\right)^2$$
$$=\frac{1}{N^2}\sum_{i,j=1}^{N}\frac{1}{t^2}\sum_{s_1=1}^{t}\sum_{s_2=1}^{t}cov\left(e_{is_1}e_{js_1},e_{is_2}e_{js_2}\right)$$
$$\leq\frac{T}{R^2}\frac{1}{N}\left(\frac{1}{N}\sum_{i,j=1}^{N}\frac{1}{T}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\left|cov\left(e_{is_1}e_{js_1},e_{is_2}e_{js_2}\right)\right|\right),$$

which implies that

$$\sum_{t=R}^{T} E\left\|D_{2t}\right\|^2 \leq \left(\frac{T}{R}\right)^2 \frac{1}{N}\left(\frac{1}{N}\sum_{i,j=1}^{N}\frac{1}{T}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\left|cov\left(e_{is_1}e_{js_1}, e_{is_2}e_{js_2}\right)\right|\right) = O\left(1/N\right) = o\left(1\right),$$

given in particular Assumption 2(f). Finally, for $D_{3t} = D'_{4t}$, we have that

$$\|D_{3t}\|^2 \leq \frac{1}{N}\underbrace{\left\|\frac{\Lambda}{\sqrt{N}}\right\|^2}_{\leq C}\left\|\frac{1}{t}\sum_{s=1}^{t}\underbrace{f_s e_s'}_{r\times N}\right\|^2 \leq C\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{t}\sum_{s=1}^{t}f_s e_{is}\right\|^2$$
$$\leq C\frac{1}{R}\left(\frac{T}{R}\right)\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{t}f_s e_{is}\right\|^2 = O_P\left(\frac{1}{R}\right) = o_P\left(1\right),$$

given Assumption 3(a) and the fact that $T/R = O\left(1\right)$.

To prove (b), we prove the result for $q = 2$ and note that this implies the result for any $q > 0$ since

$$\sup_t \|H_t\|^q = \sup_t\left(\|H_t\|^2\right)^{q/2} = \left(\sup_t\|H_t\|^2\right)^{q/2} = O_P\left(1\right)$$

if $\sup_t \|H_t\|^2 = O_P\left(1\right)$. By definition, for $t = R, \dots, T$, $H_t = \tilde{V}_t^{-1}\left(\frac{\tilde{F}_t'F_t}{t}\right)\left(\frac{\Lambda'\Lambda}{N}\right)$, implying that $\|H_t\| \leq \left\|\tilde{V}_t^{-1}\right\|\left\|\frac{\tilde{F}_t'F_t}{t}\right\|\left\|\frac{\Lambda'\Lambda}{N}\right\|$, by the fact that $\|A\cdot B\| \leq \|A\|\cdot\|B\|$. Thus,

$$\sup_t\|H_t\|^2 \leq \sup_t\left\|\tilde{V}_t^{-1}\right\|^2 \sup_t\left\|\frac{\tilde{F}_t'F_t}{t}\right\|^2\left\|\frac{\Lambda'\Lambda}{N}\right\|^2,$$

where the first factor is $O_P\left(1\right)$ given part (a) and $\left\|\frac{\Lambda'\Lambda}{N}\right\| = O\left(1\right)$ under Assumption 1(c). Thus, it suffices to show that $\sup_t\left\|\frac{\tilde{F}_t'F_t}{t}\right\|^2 = O_P\left(1\right)$. By the Cauchy–Schwarz inequality,

$$\left\|\frac{\tilde{F}_t'F_t}{t}\right\|^2 = \left\|t^{-1}\sum_{s=1}^{t}\tilde{f}_{s,t}f_s'\right\|^2 \leq \underbrace{t^{-1}\sum_{s=1}^{t}\left\|\tilde{f}_{s,t}\right\|^2}_{=r} t^{-1}\sum_{s=1}^{t}\|f_s\|^2$$
$$\leq r\cdot t^{-1}\sum_{s=1}^{t}\|f_s\|^2,$$

implying that

$$\sup_t\left\|\frac{\tilde{F}_t'F_t}{t}\right\|^2 \leq r\left(\frac{T}{R}\right)T^{-1}\sum_{s=1}^{T}\|f_s\|^2 = O_P\left(1\right)$$

since $\sup_s E\|f_s\|^2 = O\left(1\right)$ by Assumption 1(a). ■

Proof of Lemma A.6. For part (a), we follow the proof of result (2) in Bai and Ng (2013, cf. Appendix B, p. 27). In particular, by the definition of H_t, we can show that for $t = R, \dots, T$,

$$H_tH_t' = I_r + C_{1t} + C_{2t} + C_{3t},$$

where

$$C_{1t} = -\frac{1}{t}\sum_{s=1}^{t}\tilde{f}_{s,t}\left(\tilde{f}_{s,t} - H_tf_s\right)', \quad C_{2t} = -\frac{1}{t}\sum_{s=1}^{t}\left(\tilde{f}_{s,t} - H_tf_s\right)f_s'H_t'$$

and

$$C_{3t} = -H_t\left(\frac{1}{t}\sum_{s=1}^{t}f_sf_s' - I_r\right)H_t'.$$

It follows that

$$\sup_t\|C_{1t}\| \leq \sup_t\left\{\left(\frac{1}{t}\sum_{s=1}^{t}\left\|\tilde{f}_{s,t}\right\|^2\right)^{1/2}\left(\frac{1}{t}\sum_{s=1}^{t}\left\|\tilde{f}_{s,t} - H_tf_s\right\|^2\right)^{1/2}\right\}$$
$$= r^{1/2}\sup_t\left(\frac{1}{t}\sum_{s=1}^{t}\left\|\tilde{f}_{s,t} - H_tf_s\right\|^2\right)^{1/2} = O_P\left(1/\delta_{N,R}\right),$$

given Theorem 4.1. Similarly,

$$\sup_t\|C_{2t}\| \leq \left(\sup_t\frac{1}{t}\sum_{s=1}^{t}\|f_sH_t\|^2\right)^{1/2} \times \left(\sup_t\frac{1}{t}\sum_{s=1}^{t}\left\|\tilde{f}_{s,t} - H_tf_s\right\|^2\right)^{1/2} = O_P\left(1/\delta_{N,R}\right),$$

by Theorem 4.1 and the fact that the first factor can be bounded by

$$\left(\sup_t\frac{1}{t}\sum_{s=1}^{t}\|f_sH_t\|^2\right)^{1/2} \leq \sup_t\|H_t\|^2\left(\frac{1}{R}\sum_{s=1}^{T}\|f_s\|^2\right)^{1/2} = O_P\left(1\right)$$

given Lemma A.5(b). Finally, for C_{3t} we have that

$$\sup_t\|C_{3t}\| \leq \sup_t\|H_t\|^2\sup_t\left\|\frac{1}{t}\sum_{s=1}^{t}f_sf_s' - I_r\right\| = O_P\left(1/\sqrt{T}\right)$$

given Assumption 1(b) and the fact that $T/R = O\left(1\right)$. This shows that $H_tH_t' = I_r + O_P\left(1/\delta_{N,R}\right)$, where the remainder is uniform over $t = R, \dots, T$. We can now follow the argument in Bai and Ng (2013) to show that $H_t = H_{0t} + O_P\left(1/\delta_{N,R}\right)$ uniformly over $t = R, \dots, T$, where H_{0t} is a diagonal matrix with ±1 on the main diagonal. Note that we are not assuming exactly the same identifying restrictions as Bai and Ng's (2013) condition PC1. The difference is that we assume that $\Sigma_f = I_r$ instead of assuming that $F_t'F_t/t = I_r$ for each t. This explains why we get only the rate $1/\delta_{N,R}$ instead of $1/\delta_{N,R}^2$. For part (b), note that we can show that

$$\tilde{V}_t = \frac{\Lambda'\Lambda}{N} + O_P\left(1/\delta_{N,R}\right)$$

by exactly the same argument as Bai and Ng (2013, p. 27). Given our Assumption 1(b), we can write

$$\tilde{V}_t = \Sigma_\Lambda + O_P\left(1/\delta_{N,R}\right) \tag{12}$$

uniformly over t. Since $\tilde{V}_t^{-1} - \Sigma_\Lambda^{-1} = \tilde{V}_t^{-1}\left(\Sigma_\Lambda - \tilde{V}_t\right)\Sigma_\Lambda^{-1}$, it follows that

$$\sup_t\left\|\tilde{V}_t^{-1} - \Sigma_\Lambda^{-1}\right\| \leq \sup_t\left\|\tilde{V}_t^{-1}\right\|\sup_t\left\|\Sigma_\Lambda - \tilde{V}_t\right\|\left\|\Sigma_\Lambda^{-1}\right\| = O_P\left(1\right)O_P\left(1/\delta_{N,R}\right)O\left(1\right)$$

given Lemma A.5(b) and (12). ■

Proof of Lemma A.7. Part (a). We rely on Lemma A.1. In particular, for $j = 1, \ldots, 4$,

$$\left\| \frac{1}{t} \sum_{s=1}^{t-1} A_{js,t} u_{s+1} \right\| = \left\| \frac{1}{t^2} \sum_{l=1}^{t} \tilde{f}_{l,t} \left(\sum_{s=1}^{t-1} \phi_{j,l,s} u_{s+1} \right) \right\|$$
$$\leq \frac{1}{t} \left(\underbrace{\frac{1}{t} \sum_{l=1}^{t} \left\| \tilde{f}_{l,t} \right\|^2}_{=r} \right)^{1/2} \left(\frac{1}{t} \sum_{l=1}^{t} \left\| \sum_{s=1}^{t-1} \phi_{j,l,s} u_{s+1} \right\|^2 \right)^{1/2}$$
$$= \frac{1}{t} \sqrt{r} \left(\frac{1}{t} \sum_{l=1}^{t} \left\| \sum_{s=1}^{t-1} \phi_{j,l,s} u_{s+1} \right\|^2 \right)^{1/2},$$

where we let $\phi_{j,l,s}$ be one of $\left\{\gamma_{l,s}, \zeta_{l,s}, \eta_{l,s}, \xi_{l,s}\right\}$ depending on $j = 1, 2, 3, 4$. By Lemma A.1(a) through (d), it then follows that

$$\sup_t \left\| \frac{1}{t} \sum_{s=1}^{t-1} A_{js,t} u_{s+1} \right\|$$
$$\leq \frac{1}{R} \sqrt{r} \left[\sup_t \left(\frac{1}{t} \sum_{l=1}^{t} \left\| \sum_{s=1}^{t-1} \phi_{j,l,s} u_{s+1} \right\|^2 \right) \right]^{1/2}$$
$$= O_P \left(\frac{1}{R} \right) \text{ if } j = 1,$$

whereas it is $O_P \left(\frac{1}{R} \frac{\sqrt{T}}{\sqrt{N}} \right) = O\left(\frac{1}{\sqrt{RN}} \right)$ for $j = 2, 3, 4$, given that $T/R = O(1)$.

Part (b). The same arguments follow by relying on Lemma A.1(e), (f), and (g). For the last term involving $A_{4s,t}$, the analog of Lemma A.1(d) when we replace u_{s+1} with z_s' does not hold because we cannot claim that $\sup_t \left\| \frac{1}{\sqrt{T}} \sum_{s=1}^{t-1} f_s z_s' \right\|^2 = O_P(1)$ since $E\left(f_s z_s'\right) = \left(E\left(f_s w_s'\right), E\left(f_s f_s'\right)\right) \neq 0$. Thus, we need a more refined analysis. Replacing $A_{4s,t}$ with its definition yields

$$\sup_t \left\| t^{-1} \sum_{s=1}^{t-1} A_{4s,t} z_s' \right\|$$
$$= \sup_t \left\| t^{-1} \sum_{s=1}^{t-1} t^{-1} \sum_{l=1}^{t} \tilde{f}_{l,t} \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i' e_{il} \right) f_s z_s' \right\|$$
$$= \sup_t \left\| t^{-1} \sum_{l=1}^{t} \tilde{f}_{l,t} \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i' e_{il} \right) t^{-1} \sum_{s=1}^{t-1} f_s z_s' \right\|$$
$$\leq \sup_t \left\| t^{-1} \sum_{s=1}^{t-1} f_s z_s' \right\| \sup_t \left\| t^{-1} \sum_{l=1}^{t} \tilde{f}_{l,t} \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i' e_{il} \right) \right\|,$$

where $\sup_t \left\| t^{-1} \sum_{s=1}^{t-1} f_s z_s' \right\|$ is bounded under Assumption 6. By adding and subtracting appropriately, we can write

$$\sup_t \left\| t^{-1} \sum_{l=1}^{t} \tilde{f}_{l,t} \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i' e_{il} \right) \right\|$$
$$\leq \sup_t \left\| t^{-1} \sum_{l=1}^{t} \left(\tilde{f}_{l,t} - H_t f_l \right) \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i' e_{il} \right) \right\|$$
$$+ \sup_t \|H_t\| \sup_t \left\| t^{-1} \sum_{l=1}^{t} f_l \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i' e_{il} \right) \right\|.$$

The first term can be shown to be $O_P\left(1/\delta_{N,R}\right) \times O_P\left(1/\sqrt{N}\right)$ given Theorem 4.1 and the fact that $\frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i' e_{il} \right\|^2 = O_P(1)$ under our assumptions. Thus, this term is $O_P\left(1/\delta_{N,R}^2\right)$. For the second term, given that $\sup_t \|H_t\| = O_P(1)$, it suffices that

$$\sup_t \left\| t^{-1} \sum_{l=1}^{t} f_l \left(\frac{1}{N} \sum_{i=1}^{N} \lambda_i' e_{il} \right) \right\|$$
$$= \left(\frac{T}{R} \right) \frac{1}{\sqrt{TN}} \sup_t \left\| \frac{1}{\sqrt{TN}} \sum_{l=1}^{t} \sum_{i=1}^{N} f_l \lambda_i' e_{il} \right\| = O_P\left(1/\sqrt{TN}\right),$$

which follows under Assumption 3(e). This concludes the proof of part (b). ■

Proof of Lemma A.8. Part (a). Given the equality (9) and the definitions of $\hat{V}(t)$ and $V(t)$, we have that

$$\hat{V}(t) - H_t V(t)$$
$$= \frac{1}{t} \sum_{s=1}^{t-1} \left(\tilde{z}_{s,t} - \Phi_t z_s \right) u_{s+1} = \begin{pmatrix} 0 \\ \frac{1}{t} \sum_{s=1}^{t-1} \left(\tilde{f}_{s,t} - H_t f_s \right) u_{s+1} \end{pmatrix},$$

which implies that

$$\sup_t \left\| \hat{V}(t) - H_t V(t) \right\| = \sup_t \left\| \frac{1}{t} \sum_{s=1}^{t-1} \left(\tilde{f}_{s,t} - H_t f_s \right) u_{s+1} \right\|$$
$$\leq \sup_t \left\| \tilde{V}_t^{-1} \right\| \sum_{j=1}^{4} \sup_t \left\| \frac{1}{t} \sum_{s=1}^{t-1} A_{js,t} u_{s+1} \right\|$$
$$= O_P\left(\frac{1}{R}\right) + O_P\left(\frac{1}{\sqrt{RN}}\right) \quad = O_P\left(\frac{1}{\sqrt{R}\delta_{N,R}}\right),$$

by Lemmas A.5 and A.7.

Part (b). Given the equality $A^{-1} - B^{-1} = A^{-1}(B - A) B^{-1}$, and the fact that $\sup_t \left\| \hat{B}(t) \right\| = O_P(1)$ and $\sup_t \left\| \Phi_t'^{-1} B(t) \Phi_t^{-1} \right\| \leq \sup_t \left\| \Phi_t^{-1} \right\|^2 \sup_t \|B(t)\| = O_P(1)$, it suffices to consider

$$\hat{B}^{-1}(t) - \Phi_t B^{-1}(t) \Phi_t' = t^{-1} \sum_{s=1}^{t-1} \left(\tilde{z}_{s,t} \tilde{z}_{s,t}' - \Phi_t z_s z_s' \Phi_t' \right)$$
$$= \mathcal{F}_{1t} + \mathcal{F}_{2t} + \mathcal{F}_{2t}',$$

where

$$\mathcal{F}_{1t} = t^{-1} \sum_{s=1}^{t-1} \left(\tilde{z}_{s,t} - \Phi_t z_s \right) \left(\tilde{z}_{s,t} - \Phi_t z_s \right)' \text{ and}$$

$$\mathcal{F}_{2t} = t^{-1} \sum_{s=1}^{t-1} \left(\tilde{z}_{s,t} - \Phi_t z_s \right) z_s' \Phi_t'.$$

By the triangle inequality and Theorem 4.1,

$$\|\mathcal{F}_{1t}\| \leq t^{-1} \sum_{s=1}^{t-1} \left\| \tilde{z}_{s,t} - \Phi_t z_s \right\|^2 = t^{-1} \sum_{s=1}^{t-1} \left\| \tilde{f}_{s,t} - H_t f_s \right\|^2$$
$$= O_P\left(\frac{1}{\delta_{N,T}^2} \right).$$

For $\mathcal{F}_{2t}$, note that

$$\mathcal{F}_{2t} = t^{-1} \sum_{s=1}^{t-1} \left(\tilde{z}_{s,t} - \Phi_t z_s \right) z_s' \Phi_t' = \begin{pmatrix} 0 \\ t^{-1} \sum_{s=1}^{t-1} \left(\tilde{f}_{s,t} - H_t f_s \right) z_s' \Phi_t' \end{pmatrix},$$

where $\tilde{f}_{s,t} - H_t f_s = \tilde{V}_t^{-1}\left(A_{1s,t} + A_{2s,t} + A_{3s,t} + A_{4s,t}\right)$. Thus,

$$\mathcal{F}_{2t} = \tilde{V}_t^{-1}\left(t^{-1}\sum_{s=1}^{t-1}\left(A_{1s,t} + A_{2s,t} + A_{3s,t} + A_{4s,t}\right)z_s'\right)\Phi_t',$$

which given Lemma A.7 is bounded in norm by

$$\sup_t \|\mathcal{F}_{2t}\| \le \sup_t \left\|\tilde{V}_t^{-1}\right\| \sup_t \|\Phi_t\| \sum_{j=1}^{4} \sup_t \left\|\frac{1}{t}\sum_{s=1}^{t-1} A_{js,t} z_s'\right\|$$
$$= O_P\left(\frac{1}{R}\right) + O_P\left(\frac{1}{\sqrt{RN}}\right) + O_P\left(\frac{1}{\sqrt{RN}}\right) + O_P\left(\frac{1}{\delta_{N,R}^2}\right),$$

which is $O_P\left(\delta_{N,R}^{-2}\right)$.

Part (c). We have that

$$\sup_t \left\|\hat{V}(t)\right\| \le \sup_t \left\|\hat{V}(t) - \Phi_t V(t)\right\| + \sup_t \|\Phi_t\| \sup_t \|V(t)\|,$$

where $\sup_t \|V(t)\| = O_P\left(1/\sqrt{T}\right)$ given Assumption 6(b).

Part (d). Since there are no factors in the benchmark model, we can write

$$t^{-1}\sum_{s=1}^{t-1}\tilde{z}_{s,t}\left(\tilde{z}_{s,t} - \Phi_t z_s\right)'\Phi_t'^{-1}\delta$$
$$= t^{-1}\sum_{s=1}^{t-1}\tilde{z}_{s,t}\left(\tilde{f}_{s,t} - H_t f_s\right)' H_t'^{-1}\left(T^{-1/2}\alpha\right),$$

where we have used the local-to-zero parametrization of the DGP given in (6). It follows that

$$\sup_t \left\|t^{-1}\sum_{s=1}^{t-1}\tilde{z}_{s,t}\left(\tilde{f}_{s,t} - H_t f_s\right)' H_t'^{-1}\left(T^{-1/2}\alpha\right)\right\|$$
$$\le \sup_t \left\|H_t^{-1}\right\| T^{-1/2}\|\alpha\| \sup_t \left\|t^{-1}\sum_{s=1}^{t-1}\tilde{z}_{s,t}\left(\tilde{f}_{s,t} - H_t f_s\right)\right\|$$
$$\le \sup_t \left\|H_t^{-1}\right\| T^{-1/2}\|\alpha\| \sup_t \left(t^{-1}\sum_{s=1}^{t-1}\left\|\tilde{z}_{s,t}\right\|^2\right)^{1/2}$$
$$\times \sup_t \left(t^{-1}\sum_{s=1}^{t-1}\left\|\tilde{f}_{s,t} - H_t f_s\right\|^2\right)^{1/2}$$
$$= O_P\left(\left(\sqrt{T}\delta_{N,R}\right)^{-1}\right). \quad \blacksquare$$

Proof of Lemma A.9. Part (a). Given that $\ddot{\delta}_t - \delta = B(t)V(t)$, we can write

$$\sum_{t=R}^{T-1} u_{t+1}\left(\tilde{z}_{t,t} - \Phi_t z_t\right)'\Phi_t'^{-1}\left(\ddot{\delta}_t - \delta\right)$$
$$= \sum_{t=R}^{T-1} u_{t+1}\left(\tilde{z}_{t,t} - \Phi_t z_t\right)'\Phi_t'^{-1}B(t)V(t) = \mathcal{A}_1 + \mathcal{A}_2,$$

where

$$\mathcal{A}_1 = \sum_{t=R}^{T-1} u_{t+1}\left(\tilde{z}_{t,t} - \Phi_t z_t\right)'\Phi_{0t}'^{-1}BV(t), \quad \text{and}$$
$$\mathcal{A}_2 = \sum_{t=R}^{T-1} u_{t+1}\left(\tilde{z}_{t,t} - \Phi_t z_t\right)'\left(\Phi_t'^{-1}B(t) - \Phi_{0t}'^{-1}B\right)V(t).$$

Let $B = \left(E\left(z_s z_s'\right)\right)^{-1}$ and note that by Assumption 6, $\sup_t \|B(t) - B\| = O_P\left(1/\sqrt{T}\right)$. By the triangle and Cauchy–Schwarz inequalities,

$$|\mathcal{A}_2| \le \sum_{t=R}^{T-1}|u_{t+1}|\left\|\tilde{z}_{t,t} - \Phi_t z_t\right\|\left\|\Phi_t'^{-1}B(t) - \Phi_{0t}'^{-1}B\right\|\|V(t)\|$$
$$\le \sup_t \left\|\Phi_t'^{-1}B(t) - \Phi_{0t}'^{-1}B\right\| \sup_t \|V(t)\| T\left(\frac{1}{T}\sum_{t=1}^{T}u_{t+1}^2\right)^{1/2}$$
$$\times \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{z}_{t,t} - \Phi_t z_t\right\|^2\right)^{1/2}$$
$$= O_P\left(\frac{1}{\delta_{N,T}}\right)O_P\left(\frac{1}{\sqrt{T}}\right)O(T)\,O_P\left(\frac{1}{\delta_{N,T}}\right) = O_P\left(\frac{\sqrt{T}}{\delta_{N,T}^2}\right)$$
$$= o_P(1),$$

if $\sqrt{T}/N \to 0$, where we have used Theorem 4.1 to bound $\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{z}_{t,t} - \Phi_t z_t\right\|^2$, and where (1) $\sup_t \left\|\Phi_t'^{-1}B(t) - \Phi_{0t}'^{-1}B\right\| = O_P\left(1/\delta_{N,T}\right)$, (2) $\sup_t \|V(t)\| = O_P\left(1/\sqrt{T}\right)$, (3) $\frac{1}{T}\sum_{t=1}^{T}u_{t+1}^2 = O_P(1)$ hold under our assumptions. To see (1), note that

$$\sup_t \left\|\Phi_t'^{-1}B(t) - \Phi_{0t}'^{-1}B\right\|$$
$$\le \sup_t \left\|\Phi_t^{-1} - \Phi_{0t}^{-1}\right\| \sup_t \|B(t)\| + \sup_t \left\|\Phi_{0t}^{-1}\right\| \sup_t \|B(t) - B\|$$
$$= O_P\left(1/\delta_{N,T}\right) + O_P\left(1/\sqrt{T}\right) = O_P\left(1/\delta_{N,T}\right),$$

given Assumption 6 and the fact that $\sup_t \left\|\Phi_t^{-1} - \Phi_{0t}^{-1}\right\|$ is bounded by $\sup_t \left\|H_t^{-1} - H_{0t}^{-1}\right\|$, which is $O_P\left(1/\delta_{NT}\right)$ by Lemma A.6; (2) and (3) follow easily under Assumption 6. For $\mathcal{A}_1$, note that given that $\tilde{z}_{t,t} - \Phi_t z_t = \left(0', \left(\tilde{f}_{t,t} - H_t f_t\right)'\right)'$ and $\Phi_{0t} = diag\left(I_{k_1}, H_{0t}\right)$, with $H_{0t} = diag\left(\pm 1\right)$, we can write

$$\left(\tilde{z}_{t,t} - \Phi_t z_t\right)'\Phi_{0t}'^{-1}B = \left(\tilde{f}_{t,t} - H_t f_t\right)' H_{0t}C,$$

where C is a $r \times (r + k_1)$ submatrix of B containing its last r rows. It follows that

$$\mathcal{A}_1 = \sum_{t=R}^{T-1} u_{t+1}\left(\tilde{f}_{t,t} - H_t f_t\right)' H_{0t}CV(t)$$
$$= \frac{1}{\sqrt{T}}\sum_{t=R}^{T-1}\left(\tilde{f}_{t,t} - H_t f_t\right)' H_{0t}C\underbrace{\left(\sqrt{T}V(t)\,u_{t+1}\right)}_{\equiv \mathcal{U}_{t,T}},$$

where we note that $\mathcal{U}_{t,T}$ is a martingale difference sequence array with respect to $\mathcal{F}^t$ given Assumption 5. Given the decomposition (9), we can write

$$\tilde{f}_{t,t} - H_t f_t = V_0^{-1}\left(A_{1t,t} + A_{2t,t} + A_{3t,t} + A_{4t,t}\right)$$
$$+ \left(\tilde{V}_t^{-1} - V_0^{-1}\right)\left(A_{1t,t} + A_{2t,t} + A_{3t,t} + A_{4t,t}\right),$$

which implies that we can decompose $\mathcal{A}_1$ into the sum of $\mathcal{A}_{1.1}$ and $\mathcal{A}_{1.2}$ with

$$\mathcal{A}_{1.1} = \frac{1}{\sqrt{T}}\sum_{t=R}^{T-1}\left(A_{1t,t} + A_{2t,t} + A_{3t,t} + A_{4t,t}\right)' V_0'^{-1} H_{0t} C\,\mathcal{U}_{t,T},$$
$$\mathcal{A}_{1.2} = \frac{1}{\sqrt{T}}\sum_{t=R}^{T-1}\left(A_{1t,t} + A_{2t,t} + A_{3t,t} + A_{4t,t}\right)'$$
$$\times \left(\tilde{V}_t^{-1} - V_0^{-1}\right)' H_{0t} C\,\mathcal{U}_{t,T}.$$

We can prove that $\|\mathcal{A}_{1.2}\| = O_P\left(\frac{\sqrt{T}}{\delta_{N,T}^2}\right) = o_P(1)$ if $\sqrt{T}/N \to 0$, given that $\sup_t \left\|\tilde{V}_t^{-1} - V_0^{-1}\right\| = O_P\left(\frac{1}{\delta_{NT}}\right)$ by Lemma A.6; that $\frac{1}{T}\sum_{t=R}^{T-1}\left\|A_{jt,t}\right\|^2 = O_P\left(\frac{1}{\delta_{NT}^2}\right)$ for all $j = 1, \ldots, 4$ by Lemma A.4(a), and given that $\frac{1}{T}\sum_{t=R}^{T}\left\|\mathcal{U}_{t,T}\right\|^2 = O_P(1)$ by Lemma A.3(a). So, we concentrate on $\mathcal{A}_{1.1}$. We can decompose

$$\mathcal{A}_{1.1} = \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 + \mathcal{R}_4 \equiv \sum_{j=1}^{4}\mathcal{R}_j$$

with obvious definitions. Let $\phi_{1,l,t} = \gamma_{l,t}, \phi_{2,l,t} = \zeta_{l,t}, \phi_{3,l,t} = \eta_{l,t}$, and $\phi_{4,l,t} = \xi_{l,t}$. Given the definition of $A_{jt,t}$ for $j = 1, \ldots, 4$, we can write

$$A_{jt,t} = \frac{1}{t}\sum_{l=1}^{t}\tilde{f}_{l,t}\phi_{j,l,t} = H_{0t}\frac{1}{t}\sum_{l=1}^{t}f_l\phi_{j,l,t} + \frac{1}{t}\sum_{l=1}^{t}\left(\tilde{f}_{l,t} - H_{0t}f_l\right)\phi_{j,l,t}, \tag{13}$$

which implies that

$$\mathcal{R}_j = \frac{1}{\sqrt{T}}\sum_{t=R}^{T-1}\left(\frac{1}{t}\sum_{l=1}^{t}f_l\phi_{j,l,t}\right)'V_0^{-1}C\mathcal{U}_{t,T} + \frac{1}{\sqrt{T}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{l=1}^{t}\left(\tilde{f}_{l,t} - H_{0t}f_l\right)'\phi_{j,l,t}V_0'^{-1}H_{0t}C\mathcal{U}_{t,T} \equiv \mathcal{R}_{1j} + \mathcal{R}_{2j},$$

where we have used the fact that V_0 is diagonal under our assumptions and $H_{0t} = diag(\pm 1)$ to write $H_{0t}'V_0'^{-1}H_{0t} = V_0^{-1}$ when defining $\mathcal{R}_{1j}$. By Lemma A.3, we can conclude that $\mathcal{R}_{11} = O_P\left(\frac{1}{\sqrt{T}}\right)$, whereas $\mathcal{R}_{1j} = O_P\left(\frac{1}{\sqrt{N}}\right)$ for $j = 2, 3, 4$. For $\mathcal{R}_{2j}$, note that

$$\begin{aligned}
\left\|\mathcal{R}_{2j}\right\| &\leq \sqrt{T}\left\|V_0^{-1}\right\|\sup_t\|H_{0t}\|\frac{1}{T}\sum_{t=R}^{T}\left\|\mathcal{U}_{t,T}\right\| \\
&\quad\times\left\|\frac{1}{t}\sum_{l=1}^{t}\left(\tilde{f}_{l,t} - H_{0t}f_l\right)'\phi_{j,l,t}\right\| \\
&\leq \sqrt{T}\left\|V_0^{-1}\right\|\sup_t\|H_{0t}\|\left(\frac{1}{T}\sum_{t=R}^{T}\left\|\mathcal{U}_{t,T}\right\|^2\right)^{1/2} \\
&\quad\times\left(\frac{1}{T}\sum_{t=R}^{T}\left\|\frac{1}{t}\sum_{l=1}^{t}\left(\tilde{f}_{l,t} - H_{0t}f_l\right)'\phi_{j,l,t}\right\|^2\right)^{1/2} \\
&\leq \sqrt{T}\left\|V_0^{-1}\right\|\sup_t\|H_{0t}\|\left(\sup_t\frac{1}{t}\sum_{l=1}^{t}\left\|\tilde{f}_{l,t} - H_{0t}f_l\right\|^2\right)^{1/2} \\
&\quad\times\left(\frac{1}{T}\sum_{t=R}^{T}\left\|\mathcal{U}_{t,T}\right\|^2\right)^{1/2}\left(\frac{1}{TR}\sum_{t=1}^{T}\sum_{l=1}^{T}\phi_{j,l,t}^2\right)^{1/2}.
\end{aligned}$$

Note that $\frac{1}{TR}\sum_{t=1}^{T}\sum_{l=1}^{T}\gamma_{l,t}^2 = O(1/T)$ under Assumption 2(c), whereas $\frac{1}{TR}\sum_{t=1}^{T}\sum_{l=1}^{T}\phi_{j,l,t}^2 = O_P(1/N)$ for $j = 2, 3$, and 4 under Assumption 2(d) and (e). Moreover,

$$\begin{aligned}
\sup_t\frac{1}{t}\sum_{l=1}^{t}\left\|\tilde{f}_{l,t} - H_{0t}f_l\right\|^2 &= \sup_t\frac{1}{t}\sum_{l=1}^{t}\left\|\tilde{f}_{l,t} - H_tf_l + (H_t - H_{0t})f_l\right\|^2 \\
&\leq 2\left\{\sup_t\frac{1}{t}\sum_{l=1}^{t}\left\|\tilde{f}_{l,t} - H_tf_l\right\|^2 + \sup_t\|H_t - H_{0t}\|^2\frac{1}{R}\sum_{l=1}^{T}\|f_l\|^2\right\} \\
&= O_P\left(1/\delta_{N,R}^2\right) + O_P\left(1/\delta_{N,R}^2\right) = O_P\left(1/\delta_{N,R}^2\right),
\end{aligned}$$

by Lemma A.6 and Theorem 4.1. Since $\frac{1}{T}\sum_{t=R}^{T}\left\|\mathcal{U}_{t,T}\right\|^2 = O_P(1)$ by Lemma A.3(a), it follows that

$$\left\|\mathcal{R}_{2j}\right\| = O_P\left(\frac{1}{\delta_{NT}}\right) \text{ when } j = 1,$$

whereas for $j \geq 2$,

$$\left\|\mathcal{R}_{2j}\right\| = O_P\left(\frac{\sqrt{T}}{\delta_{NT}\sqrt{N}}\right),$$

which is $O_P\left(\frac{1}{\sqrt{N}}\right)$ if $\delta_{NT} = \sqrt{T}$ and $O_P\left(\frac{\sqrt{T}}{N}\right)$ if $\delta_{NT} = \sqrt{N}$. Thus, $\mathcal{A}_{1.1} = O_P\left(\frac{1}{\delta_{NT}}\right) + O_P\left(\frac{\sqrt{T}}{N}\right) = o_P(1)$ if $\sqrt{T}/N \to 0$, which concludes part (a).

Part (b). Given the decomposition of $\hat{\delta}_t - \Phi_t'^{-1}\delta_t$ in (10), we can write

$$\sum_{t=R}^{T-1}u_{t+1}\tilde{z}_{t,t}'\left(\hat{\delta}_t - \Phi_t'^{-1}\ddot{\delta}_t\right) = \mathcal{B}_1 + \mathcal{B}_2 + \mathcal{B}_3,$$

where

$$\mathcal{B}_1 = \sum_{t=R}^{T-1}u_{t+1}\tilde{z}_{t,t}'\Phi_t'^{-1}B(t)\Phi_t^{-1}\left(\hat{V}(t) - \Phi_tV(t)\right),$$

$$\mathcal{B}_2 = \sum_{t=R}^{T-1}u_{t+1}\tilde{z}_{t,t}'\left(\hat{B}(t) - \Phi_t'^{-1}B(t)\Phi_t^{-1}\right)\hat{V}(t),$$

and

$$\mathcal{B}_3 = -\sum_{t=R}^{T-1}u_{t+1}\tilde{z}_{t,t}'\hat{B}(t)t^{-1}\sum_{s=1}^{t-1}\tilde{z}_{s,t}\left(\tilde{z}_{s,t} - \Phi_tz_s\right)'\Phi_t'^{-1}\delta.$$

We can easily show that $\mathcal{B}_2 = O_P\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) = o_P(1)$ if $\sqrt{T}/N \to 0$, given in particular Lemma A.8(b) and (c). For $\mathcal{B}_1$, note that we can show that under our assumptions,

$$\mathcal{B}_1 = \sum_{t=R}^{T-1}u_{t+1}z_t'B\Phi_{0t}\left(\hat{V}(t) - \Phi_tV(t)\right) + O_P\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) \equiv \mathcal{C} + O_P\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right).$$

This follows by first replacing $\tilde{z}_{t,t}$ with $\Phi_tz_t + \left(\tilde{z}_{t,t} - \Phi_tz_t\right)$ and then replacing $B(t)\Phi_t^{-1}$ with $B\Phi_{0t}^{-1} + \left(B(t)\Phi_t^{-1} - B\Phi_{0t}^{-1}\right)$, where we note that $\Phi_{0t}^{-1} = \Phi_{0t}$. The extra terms can be shown to be $O_P\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)$ by applying Theorem 4.1 and Lemma A.8. Notice that we can write

$$\hat{V}(t) - \Phi_tV(t) = t^{-1}\sum_{s=1}^{t-1}\left(\tilde{z}_{s,t} - \Phi_tz_s\right)u_{s+1} = \begin{pmatrix} 0 \\ t^{-1}\sum_{s=1}^{t-1}\left(\tilde{f}_{s,t} - H_tf_s\right)u_{s+1}\end{pmatrix} = \underbrace{\begin{pmatrix} 0_{k_1\times r} \\ I_{r\times r}\end{pmatrix}}_{\equiv J}\tilde{V}_t^{-1}\sum_{j=1}^{4}\left(t^{-1}\sum_{s=1}^{t-1}A_{js,t}u_{s+1}\right).$$

With this notation,

$$\mathcal{C} = \sum_{t=R}^{T-1}u_{t+1}z_t'B\Phi_{0t}J\tilde{V}_t^{-1}\sum_{j=1}^{4}\left(t^{-1}\sum_{s=1}^{t-1}A_{js,t}u_{s+1}\right) \equiv \mathcal{C}_1 + \mathcal{C}_2,$$

where

$$\mathcal{C}_1 = \sum_{t=R}^{T-1} u_{t+1} z_t' B \Phi_{0t} J V_0^{-1} \sum_{j=1}^{4} \left(t^{-1} \sum_{s=1}^{t-1} A_{js,t} u_{s+1} \right)$$

and

$$\mathcal{C}_2 = \sum_{t=R}^{T-1} u_{t+1} z_t' \left(B\Phi_{0t} J \right) \left(\tilde{V}_t^{-1} - V_0^{-1} \right) \sum_{j=1}^{4} \left(t^{-1} \sum_{s=1}^{t-1} A_{js,t} u_{s+1} \right).$$

We can easily show that $\mathcal{C}_2 = O_P\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)$ by the usual inequalities, given Lemmas A.6(b) and A.7(a) and (b). Consider $\mathcal{C}_1$. We have that $\mathcal{C}_1 = \sum_{j=1}^4 \mathcal{S}_j$, where

$$\mathcal{S}_j = \sum_{t=R}^{T-1} u_{t+1} z_t' \Psi_{0t} \left(t^{-1} \sum_{s=1}^{t-1} A_{js,t} u_{s+1} \right), \text{ where } \Psi_{0t} \equiv B\Phi_{0t} J V_0^{-1}.$$

Given the definition of $A_{js,t}$, by adding and subtracting appropriately (see (13) with $s = t$), we can write $\mathcal{S}_j = \mathcal{S}_{1j} + \mathcal{S}_{2j}$, where

$$\mathcal{S}_{1j} = \sum_{t=R}^{T-1} u_{t+1} z_t' \Psi_{0t} H_{0t} \frac{1}{t} \sum_{s=1}^{t-1} \frac{1}{t} \sum_{l=1}^{t} f_l \phi_{j,l,s} u_{s+1}$$

and

$$\mathcal{S}_{2j} = \sum_{t=R}^{T-1} u_{t+1} z_t' \Psi_{0t} \frac{1}{t} \sum_{s=1}^{t-1} \frac{1}{t} \sum_{l=1}^{t} \left(\tilde{f}_{l,t} - H_{0t} f_l \right) \phi_{j,l,s} u_{s+1}.$$

It follows that

$$\begin{aligned}
\sup_t \left| \mathcal{S}_{2j} \right| &\le \sup_t \left\| \Psi_{0t} \right\| \frac{1}{R} \sum_{t=R}^{T-1} \left\| u_{t+1} z_t \right\| \\
&\quad \times \sup_t \left\| \frac{1}{t} \sum_{l=1}^{t} \left(\tilde{f}_{l,t} - H_{0t} f_l \right) \sum_{s=1}^{t-1} \phi_{j,l,s} u_{s+1} \right\| \\
&\le \sup_t \left\| \Psi_{0t} \right\| \frac{1}{R} \sum_{t=R}^{T-1} \left\| u_{t+1} z_t \right\| \left(\sup_t \frac{1}{t} \sum_{l=1}^{t} \left\| \tilde{f}_{l,t} - H_{0t} f_l \right\|^2 \right)^{1/2} \\
&\quad \times \left(\sup_t \frac{1}{t} \sum_{l=1}^{t} \left(\sum_{s=1}^{t-1} \phi_{j,l,s} u_{s+1} \right)^2 \right)^{1/2}.
\end{aligned}$$

By Lemma A.7, the last factor is $O_P(1)$ when $j = 1$, which implies that $\sup_t \left| \mathcal{S}_{2j} \right| = O_P(1/\delta_{NT})$ for this value of j. For $j = 2, 3$, and 4, Lemma A.1 implies that the last factor is $O_P\left(\sqrt{T/N}\right)$, which implies that $\sup_t \left| \mathcal{S}_{2j} \right| = O_P(1/\delta_{NT}) O_P\left(\sqrt{T/N}\right) = o_P(1)$ if $\sqrt{T}/N \to 0$. Hence, we conclude that

$$\mathcal{C}_1 = \sum_{j=1}^{4} \mathcal{S}_{1j} + o_P(1),$$

where we can write

$$\mathcal{S}_{1j} = \frac{1}{T} \sum_{t=R}^{T-1} s_{j,t+1}$$

with $s_{j,t+1} \equiv u_{t+1} z_t' Z_{jt}$ and $Z_{jt} = \Psi_{0t} H_{0t} \left(\frac{T}{t}\right) \frac{1}{t} \sum_{s=1}^{t-1} \sum_{l=1}^{t} f_l \phi_{j,l,s} u_{s+1}$. Noting that Z_{jt} is a function of $\sigma\left(z_t, z_{t-1}, \ldots; u_t, u_{t-1}, \ldots; e_t, e_{t-1}, \ldots\right)$ and that this information set is identical to $\mathcal{F}^t = \sigma\left(z_t, z_{t-1}, \ldots; y_t, y_{t-1}, \ldots; X_t, X_{t-1}, \ldots\right)$, given that $y_{t+1} = z_t'\delta + u_{t+1}$ and that $X_t = \Lambda f_t + e_t$, it follows that $E\left(s_{j,t+1} | \mathcal{F}^t\right) = 0$ by Assumption 5. Thus,

$$\begin{aligned}
Var\left(\mathcal{S}_{1j}\right) &= \frac{1}{T^2} \sum_{t=R}^{T-1} E\left(s_{j,t+1}^2\right) = \frac{1}{T^2} \sum_{t=R}^{T-1} E\left[\left(u_{t+1} z_t' Z_{jt}\right)^2\right] \\
&\le \frac{1}{T^2} \sum_{t=R}^{T-1} E\left[\left\| u_{t+1} z_t' \right\|^2 \left\| Z_{jt} \right\|^2\right] \\
&\le \frac{1}{T^2} \sum_{t=R}^{T-1} \left(E\left\| u_{t+1} z_t' \right\|^4\right)^{1/2} \left(E\left\| Z_{jt} \right\|^4\right)^{1/2} \\
&\le \frac{1}{T} \left(\sup_t E\left\| Z_{jt} \right\|^4\right)^{1/2} \frac{1}{T} \sum_{t=R}^{T-1} \left(E\left\| u_{t+1} z_t' \right\|^4\right)^{1/2}.
\end{aligned}$$

We can check that $\sup_t E\left\| Z_{1t} \right\|^4 = O_P(1)$ and $\sup_t E\left\| Z_{jt} \right\|^4 = O_P\left((T/N)^2\right)$ for $j = 2, 3, 4$ by Lemma A.2. Since $E\left\| u_{t+1} z_t' \right\|^4 \le M$ by Assumption 6, it follows that $Var\left(\mathcal{S}_{1j}\right)$ is equal to $O(1/T)$ when $j = 1$ and $O(1/N)$ when $j = 2, 3$, or 4. This shows that $\mathcal{S}_{1j} = o_P(1)$ and concludes the proof that $\mathcal{B}_1 = o_P(1)$. Finally, we show that $\mathcal{B}_3 = o_P(1)$. This term can be written as

$$\mathcal{B}_3 = -\sum_{t=R}^{T-1} u_{t+1} \tilde{z}_{t,t}' \hat{B}(t) t^{-1} \sum_{s=1}^{t-1} \tilde{z}_{s,t} \left(\tilde{f}_{s,t} - H_t f_s\right)' H_t'^{-1} \frac{\alpha}{\sqrt{T}}.$$

By adding and subtracting appropriately, we can show that

$$\begin{aligned}
\mathcal{B}_3 &= -\sum_{t=R}^{T-1} u_{t+1} z_t' B \Phi_{0t} t^{-1} \sum_{s=1}^{t-1} z_s \left(\tilde{f}_{s,t} - H_t f_s\right)' H_t'^{-1} \frac{\alpha}{\sqrt{T}} \\
&\quad + O_P\left(\frac{\sqrt{T}}{\delta_{N,T}^2}\right).
\end{aligned}$$

By using the decomposition for $\tilde{f}_{s,t} - H_t f_s$, the leading term of $\mathcal{B}_3$ is asymptotically equal to

$$-\sum_{t=R}^{T-1} u_{t+1} z_t' B \Phi_{0t} \left(t^{-1} \sum_{s=1}^{t-1} z_s \left(A_{1s,t} + \cdots + A_{4s,t}\right)' \right) V_0^{-1} H_{0t}'^{-1} \frac{\alpha}{\sqrt{T}}$$

We can then use the triangle inequality and Lemma A.7(b) to bound this term by a term of order $O_P\left(\sqrt{T}/\delta_{N,T}^2\right) = o_P(1)$ given the rate restriction $\sqrt{T}/N \to 0$. This completes the proof of part (b).

Part (c). We can write

$$\begin{aligned}
&\sum_{t=R}^{T-1} u_{t+1} \left(\tilde{z}_{t,t} - \Phi_t z_t\right)' \Phi_t'^{-1} \delta \\
&= \sum_{t=R}^{T-1} u_{t+1} \left(\tilde{f}_{t,t} - H_t f_t\right)' H_t'^{-1} \frac{\alpha}{\sqrt{T}} \\
&= \sum_{t=R}^{T-1} u_{t+1} \left(\tilde{f}_{t,t} - H_t f_t\right)' H_{0t}'^{-1} \frac{\alpha}{\sqrt{T}} \\
&\quad + \sum_{t=R}^{T-1} u_{t+1} \left(\tilde{f}_{t,t} - H_t f_t\right)' \left(H_t^{-1} - H_{0t}^{-1}\right)' \frac{\alpha}{\sqrt{T}}.
\end{aligned}$$

The second term can be bounded by $O_P\left(\sqrt{T}/\delta_{NT}^2\right) = o_P(1)$ if $\sqrt{T}/N \to 0$ by using the same arguments as those used in part (a) of this lemma to show that $\mathcal{A}_2 = o_P(1)$. Similarly, we can prove that the first term is $o_P(1)$ by relying on the same arguments as those used in part (a) to show that $\mathcal{A}_1 = o_P(1)$. ■

A.3. Proof of results in Section 4

Proof of Theorem 4.1. Part (a). By (9), it follows that

$$\frac{1}{P}\sum_{t=R}^{T-1}\left\|\tilde{f}_{t,t}-H_t f_t\right\|^2=\frac{1}{P}\sum_{t=R}^{T-1}\left\|\tilde{V}_t^{-1}\left(A_{1t,t}+A_{2t,t}+A_{3t,t}+A_{4t,t}\right)\right\|^2$$
$$\leq 4\sup_t\left\|\tilde{V}_t^{-1}\right\|^2\left(\frac{1}{P}\sum_{t=R}^{T-1}\left\|A_{1t,t}\right\|^2+\frac{1}{P}\sum_{t=R}^{T-1}\left\|A_{2t,t}\right\|^2\right.$$
$$\left.+\frac{1}{P}\sum_{t=R}^{T-1}\left\|A_{3t,t}\right\|^2+\frac{1}{P}\sum_{t=R}^{T-1}\left\|A_{4t,t}\right\|^2\right)$$

where $\sup_t\left\|\tilde{V}_t^{-1}\right\|^2 = O_P(1)$ by Lemma A.5. The result now follows by Lemma A.4.(a), which shows that the first term inside the parentheses is $O_P(1/T)$, whereas the remaining terms are $O_P(1/N)$.

Part (b). The proof follows exactly the same reasoning as part (a), given Lemma A.4.(b). ■

Proof of Lemma 4.1. The proof is immediate given Lemma A.8. ■

Proof of Lemma 4.2. The proof is immediate given (11) and Lemma A.9. ■

Proof of Theorem 4.2. The proof follows from Lemma 2.1, given Lemmas 4.1 and 4.2 and the arguments described in the main text. ■

References

Andreou, Elena, Ghysels, Eric, Kourtellos, Andros, 2013. Should macroeconomic forecasters use daily financial data? J. Bus. Econom. Statist. 31, 1–12.
Bai, Jennie, (2010). Equity Premium Predictions with Adaptive Macro Indices, manuscript.
Bai, Jushan, 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.
Bai, Jushan, 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229–1279.
Bai, Jushan, Li, Kunpeng, 2012. Statistical analysis of factor models of high dimension. Ann. Statist. 40, 436–465.
Bai, Jushan, Ng, Serena, 2006. Confidence intervals for diffusion index forecast and inference with factor-augmented regressions. Econometrica 74, 1133–1155.
Bai, Jushan, Ng, Serena, 2013. Principal components estimation and identification of static factors. J. Econometrics 176, 178–189.
Batje, Fabian, Menkhoff, Lukas, (2012). Macro Determinants of U.S. Stock Market Risk Premia in Bull and Bear Markets, manuscript.
Cakmakli, Cem, van Dijk, Dick, 2016. Getting the most out of macroeconomic information for predicting stock returns and volatility. Int. J. Forecast. 32, 650–668.
Chamberlain, Gary, Rothschild, Michael, 1983. Arbitrage, factor structure and mean–variance analysis in large asset markets. Econometrica 51, 1305–1324.
Cheng, Xu, Hansen, Bruce E., 2015. Forecasting with factor-augmented regression: A frequentist model averaging approach. J. Econometrics 186, 280–293.
Ciccarelli, Matteo, Mojon, Benoit, 2010. Global inflation. Rev. Econ. Stat. 92, 524–535.
Clark, Todd E., McCracken, Michael W., 2001. Tests of equal forecast accuracy and encompassing for nested models. J. Econometrics 105, 85–110.
Clark, Todd E., McCracken, Michael W., 2005. Evaluating direct multistep forecasts. Econometric Rev. 24, 369–404.
Clark, Todd E., McCracken, Michael W., 2012. Reality checks and comparisons of nested predictive models. J. Bus. Econom. Statist. 30, 53–66.
Fan, Jianqing, Liao, Yuan, Mincheva, Martina, 2013. Large covariance estimation by thresholding principal orthogonal complements. J. R. Stat. Soc. Ser. B Stat. Methodol. 75, 603–680.
Fosten, Jack, (2015). Forecast Evalutation with Factor-Augmented Models, University of Surrey, manuscript.
Ghysels, Eric, Horan, Casidhe, Moench, Emanuel, (2014). Forecasting through the Rear-view Mirror: Data Revisions and Bond Return Predictability, Federal Reserve Bank of New York Staff Reports No. 581.
Gonçalves, Silvia, Perron, Benoit, 2014. Bootstrapping factor-augmented regression models. J. Econometrics 182, 156–173.
Han, Xu, Inoue, Atsushi, 2015. Tests of parameter instability in dynamic factor models. Econometric Theory 31, 1117–1152.
Hansen, Peter Reinhard, Timmermann, Allan, 2015. Equivalence between out-of-sample forecast comparisons and wald statistics. Econometrica 83, 2485–2505.
Hjalmarsson, Erik, 2010. Predicting global stock returns. J. Financ. Quant. Anal. 45, 49–80.
Hofmann, Boris, 2009. Do monetary indicators lead euro area inflation? J. Int. Money Finan. 28, 1165–1181.
Horn, Roger A., Johnson, Charles R., 1985. Matrix Analysis. Cambridge University Press, Cambridge UK.
Inoue, Atsushi, Kilian, Lutz, 2007. In-sample or out-of-sample tests of predictability: which one should we use? Econometric Rev. 23, 371–402.
Kelly, Brian, Pruitt, Seth, 2013. The three-pass regression filter: a new approach to forecasting with many predictors. J. Econometrics 186, 294–316.
Li, Jia, Patton, Andrew J., (2015). Asymptotic inference about predictive accuracy using high frequency data, manuscript.
Ludvigson, Sydney, Ng, Serena, 2009. A Factor analysis of bond risk premia. In: Ullah, A., Giles, D. (Eds.), Handbook of Empirical Economics and Finance. Chapman and Hall, pp. 313–372.
McCracken, Michael W., 2007. Asymptotics for out-of-sample tests of granger causality. J. Econometrics 140, 719–752.
McCracken, Michael W., Ng, Serena, 2016. FRED-MD: A monthly database for macroeconomic research. J. Bus. Econom. Statist. 34, 574–589.
Murphy, Kevin M., Topel, Robert H., 1985. Estimation and inference in two-step econometric models. J. Bus. Econom. Statist. 3, 370–379.
Pagan, Adrian, 1984. Econometric issues in the analysis of regressions with generated regressors. Internat. Econom. Rev. 25, 183–209.
Pagan, Adrian, 1986. Two stage and related estimators and their applications. Rev. Econom. Stud. 53, 517–538.
Randles, Ronald H., 1982. On the asymptotic normality of statistics with estimated parameters. Ann. Statist. 10, 463–474.
Shintani, Mototsugu, 2005. Nonlinear forecasting analysis using diffusion indexes: An application to Japan. J. Money Credit Bank. 37, 517–538.
Stock, James H., Watson, Mark W., 2002. Macroeconomic forecasting using diffusion indexes. J. Bus. Econom. Statist. 20, 147–162.
Stock, James H., Watson, Mark W., 2003. Forecasting output and inflation: The role of asset prices. J. Econ. Lit. 41, 788–829.
West, Kenneth D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.