Queen's Economics Department Working Paper No. 1391

Model Selection in Factor-Augmented Regressions with Estimated Factors

Antoine A. Djogbenou
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

10-2017

Model Selection in Factor-Augmented Regressions with Estimated Factors

Antoine A. Djogbenou*
Queen's University
antoined@econ.queensu.ca

October 26, 2017

Abstract

This paper proposes two consistent model selection procedures for factor-augmented regressions in finite samples. We first demonstrate that the usual cross-validation is inconsistent, but that a generalization, leave-d-out cross-validation, selects the smallest basis for the space spanned by the true factors. The second proposed criterion is a generalization of the bootstrap approximation of the squared error of prediction of Shao (1996) to factor-augmented regressions. We show that this procedure is consistent. Simulation evidence documents improvements in the probability of selecting the smallest set of estimated factors than the usually available methods. An illustrative empirical application that analyzes the relationship between expected stock returns and factors extracted from a large panel of United States macroeconomic and financial data is conducted. Our new procedures select factors that correlate heavily with interest rate spreads and with the Fama-French factors. These factors have strong predictive power for excess returns.

Keywords: Factor model, consistent model selection, cross-validation, bootstrap, excess returns, macroeconomic and financial factors.
JEL classification: C52, C53, C55.

1 Introduction

Factor-augmented regression (FAR) models are now widely used for generating forecasts since the seminal paper of Stock and Watson (2002) on diffusion indices. Unlike the traditional regressions, these models allow the inclusion of a large set of macroeconomic and financial variables as predictors, useful to span various information sets related to economic agents. Thereby, economic variables are considered as driven by some unobservable factors which are inferred from a large panel of observed data. Many empirical studies have been conducted using FAR. Among others, Stock and Watson (2002) forecast the inflation rate assuming some latent factors explain the comovement in their

*I am grateful for comments from participants at the CIREQ Time Series and Financial Econometrics conference at Montreal (May 2015), the SCSE conference at Montreal (May 2015) and the CEA conference in Toronto (May 2015). A special thanks to Marine Carrasco, Xu Cheng, Sílvia Gonçalves, William McCausland and Benoit Perron for useful comments and discussions.

high dimensional macroeconomic dataset. Furthermore, Ludvigson and Ng (2007) look at the risk-return relation in the equity market. From eight estimated factors resuming the information in their macroeconomic and financial datasets using Bai and Ng (2002) IC_{p2} criterion, they identify, based on the bayesian information criterion (BIC), three new factors termed "volatility", "risk premium" and "real" factors that predict future excess returns.

Considerable research has been devoted to detect the number of factors capturing the information in a large panel of potential predictors, but very few addressed the second step selection of relevant estimated factors for a targeted dependent variable. Bai and Ng (2009) addressed this issue and revisited forecasting with estimated factors. Based on the forecast mean squared error (MSE) approximation, they pointed out that the standard BIC criterion does not incorporate the factor estimation error. Consequently, they suggested a final prediction error (FPE) type criterion with a penalty term depending on both the time series and the cross-sectional dimensions of the panel. Nevertheless, estimating consistently the MSE does not by itself ensure the consistent model selection. In fact, Groen and Kapetanios (2013) showed that this is true for the FPE criterion which inconsistently estimates of the true factor space. In consequence, they provided consistent procedures which minimize the log of the sum of squared residuals and a penalty depending on time and cross-sectional dimensions. Their consistent selection methods choose the smallest set of estimated factors that span the true factors with probability converging to one as the sample sizes grow. But in finite sample exercises, these criteria tend to underestimate the true number of estimated factors spanning the true factors. In particular, they found in the simulation experiments that their suggested modified BIC behaves similarly to the standard time series set-up with non-generated regressors using the BIC criterion by under-fitting the true model.

For finite sample improvements, cross-validation procedures have been used for a long time by statisticians to select models with observed regressors and are considered here for factor-augmented regression model selection. As is well known, the leave-one-out cross-validation (CV_1) measures the predictive ability of a model by testing it on a set of regressors and regressand not used in estimation. This model selection procedure is consistent if only one set of generated regressors spans the true factors. Indeed, the CV_1 criterion breaks down into five main terms: the variability of the future observations term (independent of candidates models), the complexity error term (increases with model dimension), the model identifiability term (zero for models with estimated factors spanning the true factor space), its parameter and factor estimation errors. When only one set of generated factors spans the true model, this criterion converges to the forecast error variance for this particular set since the identifiability component is zero and the remainder ones converge to zero. But for the other candidate sets, it is inflated by the positive limit of the identifiability part since they do not span the true latent factor space. These sets of estimated factors called incorrect are therefore excluded with probability converging to one when we minimize the standard cross-validation criterion.

However, when many sets of estimated factors generate the true model, the CV_1 model selection procedure has a positive probability of not choosing the smallest one. The source of this problem is not only due to the well known parameter estimation error when factors are observed but also the factor estimation error in this criterion. The harmful effect of generated regressors is more pronounced when the cross-sectional dimension is much smaller than the time dimension as the factor estimation component dominates in finite sample both the complexity and the parameter estimation ones. Our simulations show that this factor estimation error while asymptotically negligible, contributes to reduce considerably the probability to select in finite samples the smallest set of estimated factors that generate the true factor space.

In this paper, we suggest two alternative model selection procedures with better finite sample properties that are consistent and select the smaller set of estimated factors spanning the true model. The first is the Monte Carlo leave-d-out cross-validation suggested by Shao (1993) in the context of observed and fixed regressors. The other method uses the bootstrap selection procedure studied by Shao (1996) which is implemented with the two-step residual-based bootstrap method suggested by Gonçalves and Perron (2014) when the regressors are generated.

Overall, in comparison with the existing literature, this paper focuses on two-step factor-augmented regression models widely used by practitioner. It does not assume that all latent factors in the large panel are relevant for a prediction purpose. Further, because our interest is the role played by factors in predicting a given variable, we mainly study consistent selection of the estimated factors and do not cover efficient model selection. In addition, the proposed selection rules are designed in order to provide better finite sample performance. In particular, the simulations show that the leave-one-out cross-validation often selects a larger set of estimated factors than the smallest relevant one, while the modified BIC of Groen and Kapetanios (2013) tends to under-parameterize for smaller sample sizes. Nevertheless, the Monte Carlo leave-d-out cross-validation and the bootstrap selection pick with higher probability the estimated factors spanning the true factors. To illustrate the methods, an empirical application that revisits the relationship between macroeconomic and financial factors, and excess stock returns for the U.S. market has been conducted. The factors are extracted from 147 financial series and 130 macroeconomic series. The financial series correspond to the 147 variables in Jurado, Ludvigson, and Ng (2015). The quarterly macroeconomic data set is constructed following McCracken and Ng (2015) and spans the first quarter of 1960 to the third quarter of 2014. After controlling for the consumption-wealth variable (Lettau and Ludvigson, 2005), the lagged realized volatility of the future excess returns and other factors, among the estimated factors from a large panel of U.S. macro and financial data, the factors heavily correlated with interest rate spreads and with the Fama-French factors have strong additional predictive power for excess returns. The out-of-sample performance for predicting excess returns with the new procedures is also compared to existing model selection ones.

The remainder of the paper is organized as follows. In Section 2, we present the settings and assumptions. Section 3 addresses model selection. Section 4 reports the simulation study, and the fifth section presents the empirical application. The last section concludes. Mathematical proofs, figures and tables appear in the Appendix.

2 Settings and assumptions

The econometrician observes $(y_t,\ W_t',\ X_{1t},\ldots,X_{it},\ldots,X_{Nt})$, $t=1,\cdots,T$, and the goal is to predict y_{T+1} using the following factor-augmented regression model

$$y_{t+1}=\delta' Z_t^0+\varepsilon_{t+1},\ t=1,...,\ T-1 \tag{1}$$

with $Z_t^0=(F_t^{0\prime},\ W_t')'$ such that W_t is a $q-$vector of observed regressors, and F_t^0, an r_0-vector of unobserved factors. The latent factors F_t^0 are among the common factors $F_t : r\times 1$ in the large approximate factor model

$$X_{it}=\lambda_i' F_t+e_{it}, i=1,...,N,\ t=1,...,T,$$

where $\lambda_i : r\times 1$ are the factor loadings, and e_{it} an idiosyncratic error term. Because the factors F_t^0 are unobserved, they are replaced by a subset $\tilde{F}_t(m)$ from the r estimated factors $\tilde{F}_t$ from X using principal component estimation. Hence, the estimated regression takes the form

$$y_{t+1}=\alpha(m)'\tilde{F}_t(m)+\beta' W_t+u_{t+1}(m)=\delta(m)'\hat{Z}_t(m)+u_{t+1}(m) \tag{2}$$

where m is any of the 2^r subsets of indices in $\{1,\ \ldots,\ r\}$ denoted $\mathcal{M}$ including the empty set, where no latent factor drives y. The size of $\tilde{F}_t(m)$ is $r(m)\leq r$ and we assume the number of estimated factors selected in the first step is known and equal to r. While Kleibergen and Zhan (2015) guide against the harmful effect of under parameterizing on the true R^2 and test statistics, Kelly and Pruitt (2015) correct for forecast using irrelevant factors by suggesting a three-pass regression filter procedure. Cheng and Hansen (2015) study forecasting using a frequentist model averaging approach. Carrasco and Rossi (2016) also develop regularization methods for in-sample inference and forecasting in misspecified factor models. However, none of these papers study the consistent estimation of the true latent factors space in order to predict y based on the commonly used ordinary least squares of FAR with principal components.

Although there is a large body of literature on selecting the number of factors that resume the information in the factor panel dataset, including the work of Bai and Ng (2002), very few papers have been devoted to the second-step selection. This paper is precisely interested in this second-step selection. Fosten (2017) recently proposes consistent information criteria in cases where a subset of the large panel has a large impact on the dependent variable. We do not allow idiosyncratic error in the FAR, like in Fosten (2017), as we do not impose a finite number of the large set of predictors to predict y_{t+h}. Further, we focus on the two-step FAR where factors affecting potentially few or large subset of the series in X are identified and used for prediction. We denote $Z_t = (F_t', W_t')'$, $t = 1, ..., T$, the vector containing all latent factors and observed regressors, $\|M\| = (\text{Trace}\,(M'M))^{1/2}$, the Euclidean norm, $Q > 0$, the positive definiteness for any square matrix Q, and C, a generic finite constant. The following standard assumptions are made.

Assumption 1. (factor model and idiosyncratic errors)

(a) $\mathrm{E}\,\|F_t\|^4 \leq C$ and $\frac{1}{T}F'F \xrightarrow{P} \Sigma_F > 0$, where $F = (F_1, \cdots, F_T)'$.

(b) $\|\lambda_i\| \leq C$ if λ_i are deterministic, or $\mathrm{E}\,\|\lambda_i\| \leq C$ if not, and $\frac{1}{N}\Lambda'\Lambda \xrightarrow{P} \Sigma_\Lambda > 0$, where $\Lambda = (\lambda_1, \cdots, \lambda_N)'$.

(c) The eigenvalues of the $r \times r$ matrix $(\Sigma_F \times \Sigma_\Lambda)$ are distinct.

(d) $\mathrm{E}\,(e_{it}) = 0$, $E\,|e_{it}|^8 \leq C$.

(e) $\mathrm{E}\,(e_{it}e_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \overline{\sigma}_{ij}$ for all (t,s) and $|\sigma_{ij,ts}| \leq \tau_{st}$ for all (i,j), with $\frac{1}{N}\sum_{i,j=1}^{N} \overline{\sigma}_{ij} \leq C$, $\frac{1}{T}\sum_{t,s=1}^{T} \tau_{st} \leq C$ and $\frac{1}{NT}\sum_{i,j,t,s=1} |\sigma_{ij,ts}| \leq C$.

(f) $\mathrm{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(e_{it}e_{is} - E\,(e_{it}e_{is})\right)\right|^4 \leq C$ for all (t,s).

Assumption 2. (moments and weak dependence among $\{z_t\}$, $\{\lambda_i\}$, $\{e_{it}\}$ and $\{\varepsilon_{t+1}\}$)

(a) $\mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t e_{it}\right\|^2\right) \leq C$, where $\mathrm{E}\,(F_t e_{it}) = 0$ for every (i,t).

(b) For each t, $\mathrm{E}\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N} Z_s\left(e_{it}e_{is} - E\,(e_{it}e_{is})\right)\right\|^2 \leq C$, where $Z_s = (F_s', W_s')'$.

(c) $\mathrm{E}\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T} Z_t e_t'\Lambda\right\|^2 \leq C$ where $\mathrm{E}\,(Z_t\lambda_i' e_{it}) = 0$ for all (i,t).

(d) $\mathrm{E}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_t\right\|^2\right) \leq C$, where $\mathrm{E}\,(\lambda_i e_{it}) = 0$ for all (i,t).

(e) As N, $T \longrightarrow \infty$, $\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j' e_{it}e_{jt} - \Gamma \xrightarrow{P} 0$, where $\Gamma \equiv \lim_{N,\,T\longrightarrow\infty}\frac{1}{T}\sum_{t=1}^{T}\Gamma_t > 0$ and $\Gamma_t \equiv \mathrm{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right)$.

(f) For each t and $h \geq 0$, $\mathrm{E}\left|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_{s+h}\left(e_{it}e_{is} - E\,(e_{it}e_{is})\right)\right| \leq C$.

(g) $\mathrm{E}\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T-h}\lambda_i e_{it}\varepsilon_{t+1}\right\|^2 \leq C$, where $\mathrm{E}\,(\lambda_i e_{it}\varepsilon_{t+1}) = 0$ for all (i,t).

Assumption 3. (moments and Central Limit Theorem for the score vector)

(a) $\mathrm{E}\left(\varepsilon_{t+1}|\mathcal{F}_t\right)=0$, $\mathrm{E}\left(\varepsilon_{t+1}^2|\mathcal{F}_t\right)=\sigma^2$, $\mathrm{E}\left\|Z_t\right\|^8<C$ and $\mathrm{E}\left(\varepsilon_{t+1}^8\right)<C$, where

$$\mathcal{F}_t=\sigma(y_t,\ F_t',\ W_t',\ X_{1t},...,\ X_{Nt},\ y_{t-1},\ F_{t-1}',\ W_{t-1}',\ X_{1,t-1},...,\ X_{N,t-1},...)$$

(b) $\Sigma_Z=\underset{T\longrightarrow\infty}{\mathrm{plim}}\frac{1}{T}\sum_{t=1}^{T}Z_tZ_t'>0$.

(c) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}Z_t\varepsilon_{t+1}\xrightarrow{d}\mathrm{N}\left(0,\ \Omega\right)$, with Ω positive definite.

Assumptions 1 and 2 are the same as in Bai and Ng (2002), Gonçalves and Perron (2014) and Cheng and Hansen (2015) in terms of factor-augmented regression specifications that allow for weak dependence and heteroscedasticity in the idiosyncratic errors. Assumption 3 is useful for deriving the asymptotic distribution of the estimator $\hat{\delta}$ of δ. It assumes that the forecast error is conditionally homoscedastic which is rather strong. This assumption is commonly used in consistent model selection based on cross-validation. It could be relaxed if our interest was efficient model selection (Shao, 1997). We leave this for future research.

The principal component estimate $\tilde{F}$ corresponds to the eigenvectors of $\frac{1}{T}XX'$ associated with the r largest eigenvalues times $\sqrt{T}$, using the normalization $\tilde{F}'\tilde{F}/T=I_r$. As is well known, $\tilde{F}_t$ only consistently estimates a rotation of F_t given by HF_t, with H identifiable asymptotically under Bai and Ng's (2013) assumptions. Note that

$$H=\tilde{V}^{-1}\frac{\tilde{F}'F}{T}\frac{\Lambda'\Lambda}{N},\tag{3}$$

where $\tilde{V}$ contains the r largest eigenvalues of XX'/NT, in decreasing order along the diagonal and is a diagonal matrix with dimension $r\times r$. As it has been argued previously, all of the estimated factors are not necessarily relevant for prediction.

3 Model selection

The aim of this work is to provide an appropriate procedure to select the set of estimated factors that should be used to estimate (2). In practice, we extract estimated factors $\tilde{F}_t$ which summarize information in the large $N\times T$ matrix X. Afterwards, a subvector $\tilde{F}_t\left(m\right)$ is chosen for the prediction of y_{t+1}. Ludvigson and Ng (2007) select $\tilde{F}_t\left(m\right)$ of $\tilde{F}_t$ using the bayesian information criterion (BIC) to predict excess stock returns. Because this criterion does not correct for factor estimation, Bai and Ng (2009) suggest a modified final prediction error (FPE) with an extra penalty to proxy the effect of factor estimation by approximating the mean squared error (MSE). However, as pointed out by Stone (1974), we may have a consistent estimate of the MSE or a loss that does not select the true observed regressors with probability converging to one. This is also true when the variables are latent factors and the goal is to estimate consistently the true factor space. Conditionally on the information up to time t, the true conditional mean is

$$\mathrm{E}\left(y_{t+1}|\mathcal{F}_t\right) = \alpha' F_t^0 + \beta' W_t,\ t = 1,\ \ldots,\ T-1.$$

In consistent model selection literature, it is common to distinguish correct and incorrect sets of predictors. In the usual case with observed factors, Shao (1997) defines a set of regressors m as correct if its conditional mean equals that of the true unknown model, i.e.,

$$\alpha\left(m\right)' F_t\left(m\right) + \beta' W_t = E\left(y_{t+1}|\mathcal{F}_t\right),\ t = 1,\ \ldots,\ T-1.$$

When the smallest set of regressors that generates the true model is picked with probability going to one, the selection procedure is said to be consistent. For FAR models with generated regressors, Groen and Kapetanios (2013) suggest a consistent procedure based on IC type criteria, which select $\tilde{F}_t\left(m\right)$ spanning asymptotically the true unknown factors F_t^0. Formally, $\tilde{F}_t\left(m\right)$ spans F_t^0 or m is correct if $\tilde{F}_t\left(m\right) - F_t\left(m\right) \xrightarrow{P} 0$ and there is a $r_0 \times r\left(m\right)$ matrix $A\left(m\right)$ such that $F_t^0 = A\left(m\right) F_t\left(m\right)$. By definition, $F_t\left(m\right) = H_0\left(m\right) F_t$, where $H_0\left(m\right)$ is a $r\left(m\right) \times r$ sub-matrix of $H_0 = \underset{N,\ T\to\infty}{\operatorname{plim}} H$. If H_0 is diagonal, each estimated factor will identify one and only one unobserved factor. Note that for any m, $F_t(m)$ is subvector of $H_0 F_t$ where we avoid the subscript H_0 to simplify the notation. Further, the only subvector of F_t that will be considered in the paper is the true set of latent factors F_t^0. Bai and Ng (2013) extensively studied conditions that help identify the factors from the first step estimation. We define by $\mathcal{M}_1$, the category of estimated models with set of estimated factors that are incorrect, and by $\mathcal{M}_2$, those which are correct. There is at least one correct set of estimated factors in $\mathcal{M}$ which is the one with all r estimated factors. In remainder of the paper, we will associate one set of estimated factors to the corresponding estimated model. That been said, if we denote m_0 the smallest correct set of generated regressors, a selection procedure will be called consistent if it selects a set of generated regressors $\hat{m}$ such that

$$P\left(\hat{m} = m_0\right) \longrightarrow 1 \text{ as } T,\ N \to \infty.$$

In finite sample experiments, Groen and Kapetanios (2013) information criteria tend to underestimate the true number of factors. In particular, their suggested modified BIC behaves as the BIC for time series with non-generated regressors known to under-fit the true model. In order to obtain a finite sample improvement, this paper proposes alternative consistent selection procedures using cross-validation and bootstrap methods.

The next subsection begins by showing why the usual "naive" leave-one-out cross-validation fails to select the smallest correct set of estimated factors with a probability approaching one, as the sample sizes increase. In addition, a theoretical justification of the Monte Carlo cross-validation and the bootstrap selection procedures in this generated regressors framework is provided.

3.1 Leave-d-out or delete-d cross-validation

This part of the paper studies the factor-augmented model selection based on cross-validation starting with the usual leave-one-out or delete-one cross-validation. As is well known, it estimates the predictive ability of a model by testing it on a set of regressors and regressand not used in estimation. Thereby, the leave-one-out cross-validation minimizes the average squared distance

$$CV_1(m) = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(y_{t+1} - \hat{\delta}_t'(m)\,\hat{Z}_t(m)\right)^2$$

between y_{t+1} and its point forecast using an estimate from the remaining time periods

$$\hat{\delta}_t(m) = \left(\sum_{|j-t|\geq 1}\hat{Z}_j(m)\,\hat{Z}_j(m)'\right)^{-1}\left(\sum_{|j-t|\geq 1}\hat{Z}_j(m)\,y_{j+1}\right).$$

However, by minimizing the CV_1, there is a positive probability that we do not select the smallest possible correct set of generated regressors. In Lemma 3.1, we show that this positive probability to select a larger correct set of estimated factor is not only due to the parameter estimation error but also to the factor estimation one in the CV_1 criterion. We denote $P(m)$, the projection matrix associated with the space spanned by $Z(m) = (F(m),\ W)$, with $F(m)$ the generic limit of $\tilde{F}(m)$ and $\mu = Z^0\delta$, the true conditional mean vector.

Lemma 3.1. *Suppose that Assumptions 1–3 hold. If for any m,*

$$\operatorname*{plim}_{T\to\infty}\sup{}_{1\leq t\leq T-1}\left|Z_t(m)'\left[Z(m)'Z(m)\right]^{-1}Z_t(m)\right| = 0,$$

as $T,\ N\to\infty$, then when m is a correct set of estimated factors,

$$CV_1(m) = \frac{1}{T-1}\varepsilon'\varepsilon + 2\frac{(r(m)+q)}{T-1}\sigma^2 - \frac{1}{T-1}\varepsilon' P(m)\,\varepsilon + V_T(m) + o_P\left(\frac{1}{C_{NT}^2}\right),$$

where $V_T(m) = O_P\left(\frac{1}{C_{NT}^2}\right)$. When m is an incorrect set of estimated factors,

$$CV_1(m) = \sigma^2 + \frac{1}{T-1}\mu'(I-P(m))\mu + o_P(1).$$

From Lemma 3.1, for a correct set of estimated factors,

$$CV_1(m) = \sigma^2 + o_P(1),$$

otherwise

$$CV_1(m) = \sigma^2 + \frac{1}{T-1}\mu'(I-P(m))\mu + o_P(1).$$

Lemma 3.1 extends Equations (3.5) and (3.6) of Shao (1993) to the case where the factors are not observed but estimated. Contrary to that case where the regressors are observed, we have an additional term $V_T(m)$ corresponding to the factor estimation error CV_1, and $P(m)$ is associated with the space spanned by subsets of FH_0', a rotation of the true factor space. Consider two candidates sets m_1 and m_2 such that m_1 is correct and m_2 is incorrect. Assume $\operatorname*{plim\,inf}_{T\to\infty}\frac{1}{T-1}\mu'(I - P(m))\mu > 0$ for incorrect set of estimated factors. The CV_1 will prefer m_1 to m_2 since

$$\operatorname*{plim}_{N,T\to\infty} CV_1(m_1) = \sigma^2 < \sigma^2 + \operatorname*{plim}_{T\to\infty}\frac{1}{T-1}\mu'(I - P(m_2))\mu = \operatorname*{plim}_{N,T\to\infty} CV_1(m_2).$$

as $\frac{1}{T-1}\varepsilon' P(m)\varepsilon = o_P(1)$. Thus, incorrect sets of estimated factors will be excluded with probability approaching one. Therefore, the CV_1 is consistent when $\mathcal{M}_2$ contains only one correct set of estimated factors. When $\mathcal{M}_2$ contains more than one correct set of estimated factors, suppose m_1 and m_2 are two correct set of estimated factors with sizes $r(m_1)$ and $r(m_2)$ $(r(m_1) < r(m_2))$. The leave-one-out cross-validation selects with positive probability the unnecessary large model m_2 when the factors are generated. Indeed, for $m \in \mathcal{M}_2$,

$$CV_1(m) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{(r(m)+q)}{T-1}\sigma^2 + \left(\frac{(r(m)+q)}{T-1}\sigma^2 - \frac{1}{T-1}\varepsilon' P(m)\varepsilon\right) + V_T(m) + o_P\left(\frac{1}{C_{NT}^2}\right)$$

with $V_T(m) = O_P\left(\frac{1}{C_{NT}^2}\right)$. The first term is independent of candidate models. The second term captures the complexity of the model. It is the expected value of $\frac{1}{T-1}\varepsilon' P(m)\varepsilon$ and it increases with the model dimension. The term in parenthesis is a parameter estimation error with mean zero while comparing two competing correct sets of estimated factors. The term $V_T(m)$ contains the factor estimation error in the $CV_1(m)$ which is not reflected by the term in parentheses. Because the complexity component is inflated in finite samples not only by this parameter estimation error but also the factor estimation one, we fail to pick accurately the smallest correct set of estimated factors. In the usual case with observed factors, Shao (1993) already showed that the leave-one-out cross-validation has a positive probability to select a larger model than the consistent one because of the presence of the parameter estimation error. Hence, the consistent model selection crucially depends on the ability to capture the complexity term useful to penalize the over-fitting.

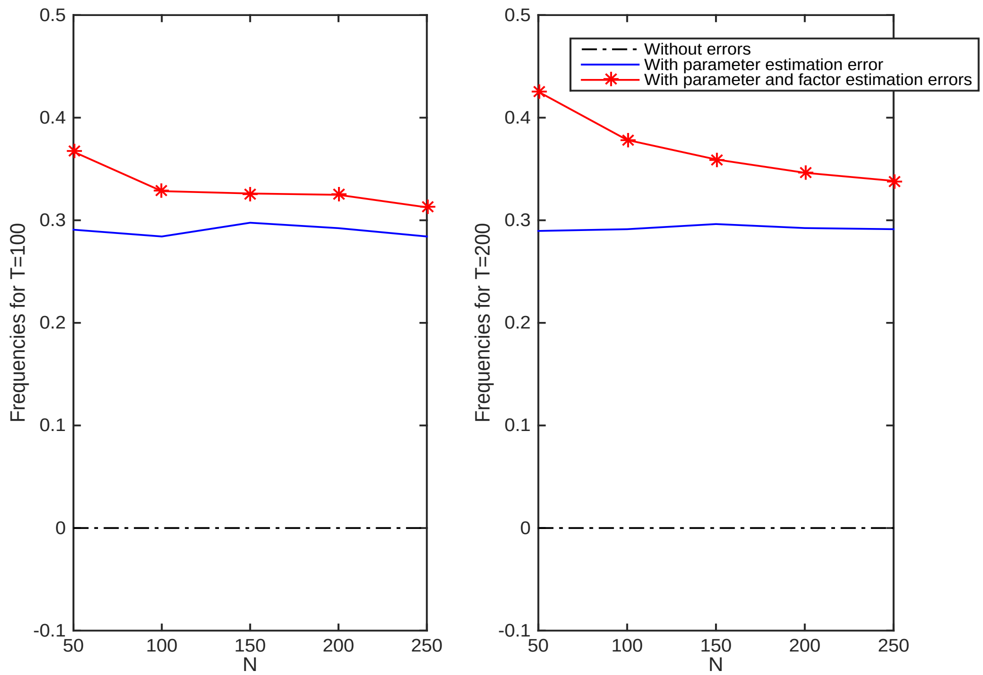
When the factor estimation error in the CV_1 is such that $N = o(T)$, then $V_T(m) = O_P\left(\frac{1}{N}\right)$ and dominates both the complexity term and the parameter estimation error. More precisely, comparing two competing models in $\mathcal{M}_2$ amounts to the comparison of their factor estimation errors in CV_1 instead of the model complexities since

$$CV_1(m) = \frac{1}{T-1}\varepsilon'\varepsilon + V_T(m) + o_P\left(\frac{1}{N}\right).$$

We analyze through a simulation study how the factor estimation error V_T, which is random, contributes to worsen the probability of selecting a consistent model.

We consider the same data generating process (DGP) as the first DGP in the simulation section, where $y_{t+1} = 1 + F_{1t} + 0.5F_{2t} + \varepsilon_{t+1}$, with $\varepsilon_{t+1} \sim N(0,1)$ and $F^0 = (F_1,\ F_2) \subset F = (F_1,\ F_2,\ F_3,\ F_4)$. Given the specification for the latent factors and the factor loadings, the PC1 condition for identifying restrictions provided by Bai and Ng (2013) is asymptotically satisfied and makes possible to identify estimated factors. Hence, we extract four estimated factors and we expect to pick consistently the first two among the $2^4 = 16$ possibilities. The line "with parameter and factor estimation errors" on Figure 1 reports the frequency of selecting a larger set of estimated factors while minimizing the CV_1 criterion which includes the estimation errors.

Figure 1: Frequencies of selecting a larger set of estimated factors minimizing the CV_1 criterion without errors, with the parameter estimation error and both the parameter and the factor estimation errors over 10,000 simulations



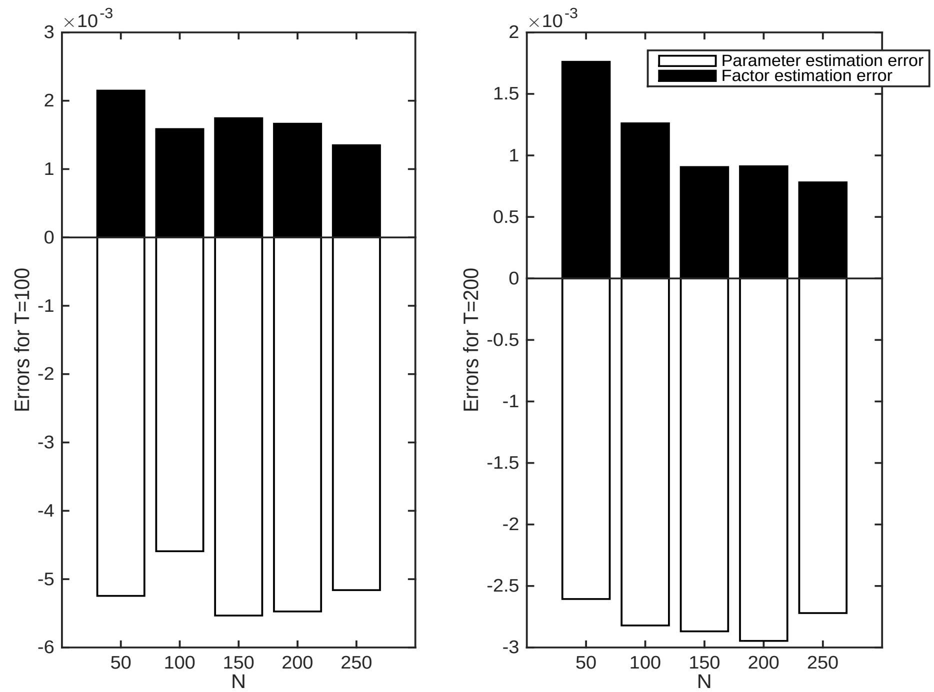
Note: This table reports the frequencies of selecting a larger set of estimated factors than the one that contains the first two factors. The line "without errors" represents the frequencies while minimizing the complexity component and the identifiability one plus $\frac{1}{T-1}\varepsilon'\varepsilon$. The line "with parameter estimation error" corresponds to the frequency when the parameter estimation error is added. The line "with parameter and factor estimation errors" relates to the "naive" leave-one-out cross-validation which includes both the parameter and the factor estimation errors.

Given the different sample sizes, it turns out that the leave-one-out cross-validation selects very often a larger model. To understand how each component in the CV_1 contributes to this overfitting, we will minimize the sum of the complexity and the identifiability terms plus the forecast error in the leave-one-out cross-validation criterion which is

$$CV_{11}(m) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{(r(m)+q)}{T-1}\sigma^2 + \frac{1}{T-1}\mu'(I-P(m))\mu$$

where we omit the parameter and the factor estimation errors. The second and the third terms

Figure 2: Average parameter estimation error and factor estimation error in the CV_1 criterion for selected model over 10,000 simulations



Note: This figure shows the average minimum parameter and factor estimation errors in the leave-one-out cross-validation criterion as N and T vary over the simulations. See also note for Figure 1.

are those important for consistent model selection. The corresponding line "without errors" on Figure 1 shows that we never over-fit through the $10,000$ simulations. Afterwards, we incorporate the parameter estimation error by minimizing

$$CV_{12}(m) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{(r(m)+q)}{T-1}\sigma^2 + \frac{1}{T-1}\mu'(I-P(m))\mu + \left(\frac{(r(m)+q)}{T-1}\sigma^2 - \frac{1}{T-1}\varepsilon' P(m)\varepsilon\right).$$

Once the parameter estimation error is included, the frequency of selecting a larger set increases. Moreover, when we include both the parameter and the factor estimation errors corresponding to the CV_1, that frequency increases more (see, Figure 1). The results show that this factor estimation error while asymptotically negligible, also increases this probability given the different sample sizes. In addition, an increase in the cross-sectional dimension implies a decrease in the factor estimation error (see, Figure 2) which is followed by a drop of the probabilities of over-parametrization.

The sum of the complexity and the identifiability term in the CV_1, helpful to achieve the consistent selection of the estimated factors, corresponds to the conditional mean of the infeasible in-sample squared error

$$L_T(m) = \frac{1}{T-1}(\hat{\mu}(m) - \mu)'(\hat{\mu}(m) - \mu) = \frac{1}{T-1}\varepsilon' P(m)\varepsilon + \frac{1}{T-1}\mu'(I - P(m))\mu,$$

with $\hat{\mu}_t(m) = P(m)y$.

To avoid the selection of larger models, Shao (1993) suggests in observed regressors set-up, a modification of the CV_1 using a smaller construction sample to estimate δ by deleting $d \gg 1$ periods for validation. This consists in splitting the $T-1$ time observations into $\kappa = (T-1) - d$ randomly drawn observations that are used for parameter estimation and d remaining ones that are used for evaluation, while repeating this process b times with b going to infinity. We extend it to FAR and provide conditions for its validity.

Given b random draws of d indexes s in $\{1, \ldots, T-1\}$ called validation samples, for each draw $s = \{s(1), \cdots, s(d)\}$, we define

$$y_s = \begin{pmatrix} y_{s(1)} \\ y_{s(2)} \\ \vdots \\ y_{s(d)} \end{pmatrix}, \; \hat{Z}_s(m) = \begin{pmatrix} \tilde{F}'_{s(1)}(m) & W'_{s(1)} \\ \tilde{F}'_{s(2)}(m) & W'_{s(2)} \\ \vdots & \vdots \\ \tilde{F}'_{s(d)}(m) & W'_{s(d)} \end{pmatrix}.$$

The corresponding construction sample is indexed by $s^c = \{1, \ldots, T-1\} \setminus s$, with y_{s^c} the complement of y_s in y and $\hat{Z}_{s^c}$ the complement of $\hat{Z}_s$ in $\hat{Z}$. We denote $\tilde{y}_s(m) = \hat{Z}_s(m)\hat{\delta}_{s^c}(m)$, $\hat{\delta}_{s^c} = \left(\hat{Z}_{s^c}(m)'\hat{Z}_{s^c}(m)\right)^{-1}\hat{Z}_{s^c}(m)' y_{s^c}$. The Monte Carlo leave-$d$-out cross-validation estimated model is obtained by minimizing

$$CV_d(m) = \frac{1}{d \cdot b}\sum_{s \in \mathcal{R}} \|y_s - \tilde{y}_s(m)\|^2,$$

where $\mathcal{R}$ represents a collection of b subsets of size d randomly drawn from $\{1, \ldots, T-1\}$. This procedure generalizes the leave-one-out cross-validation because when $d = 1$, $s = \{t\}$, $s^c = \{1, \ldots, t-1, \; t+1, \ldots, T-1\}$ and $\mathcal{R} = \{\{1\}, \ldots, \{T-1\}\}$, with $CV_d(m) = CV_1(m)$. Using a smaller construction sample, the next theorem shows that for correct sets of estimated factors,

$$CV_d(m) = \frac{1}{d \cdot b}\sum_{s \in \mathcal{R}}\sum_{t \in s}\varepsilon_{t+1}^2 + \frac{r(m) + q}{\kappa}\sigma^2 + o_P\left(\frac{1}{\kappa}\right)$$

and for incorrect sets of estimated factors,

$$CV_d(m) = \sigma^2 + \frac{1}{T-1}\mu'(I - P(m))\mu + o_P(1).$$

Hence, for correct sets of estimated factors m_1 and m_2 such that $r(m_1) < r(m_2)$,

$$P(CV_d(m_1) - CV_d(m_2) < 0) = P(r(m_2) - r(m_1) > 0 + o_P(1)) = 1 + o(1).$$

Thus, m_1 will be preferred to m_2. To prove the validity of this procedure, we made some additional assumptions.

Assumption 4.

(a) $\operatorname*{plim\,inf}_{T\to\infty}\frac{1}{T-1}\mu'(I-P(m))\mu>0$ for any $m\in\mathcal{M}_1$.

(b) $\operatorname*{plim}_{T\to\infty}\sup_{1\le t\le T-1}\left|Z_t(m)'\left[Z(m)'Z(m)\right]^{-1}Z_t(m)\right|=0$ for all m.

(c) $\operatorname*{plim}_{T\to\infty}\sup_{s\in\mathcal{R}}\left\|\frac{1}{d}Z_s'Z_s-\frac{1}{\kappa}Z_{s^c}'Z_{s^c}\right\|=0$ where $\kappa=T-1-d$.

(d) $\mathrm{E}\left(\mathrm{e_{it}e_{ju}}\right)=\sigma_{\mathrm{ij,tu}}$ with $\frac{1}{\sqrt{T\cdot\kappa}}\sum_{t\,\in\,s^c}\sum_{u=1}^{T}\frac{1}{N}\sum_{i,j}|\sigma_{ij,tu}|\le C$ for all s.

(e) $\frac{1}{\kappa}\mathrm{E}\left(\sum_{\mathrm{t}\,\in\,\mathrm{s^c}}\left\|\frac{1}{\sqrt{\mathrm{N}}}\sum_{\mathrm{i=1}}^{\mathrm{N}}\lambda_{\mathrm{i}}\mathrm{e_t}\right\|^4\right)\le\mathrm{C}$ for all (i,t) and all s.

(f) $\frac{1}{d}Z_s'(m)\,Z_s(m)\xrightarrow{P}\Sigma_Z(m)>0$ for all m and all s.

Assumption 4 (a) is an identifiability assumption in order to distinguish a correct set of estimated factors from an incorrect one. Groen and Kapetanios (2013) also made this assumption. By Assumption 4 (b), for any estimated model, the diagonal elements of the projection matrix vanish asymptotically. This regularity condition can be seen as a form of a stationarity assumption for regressors in the different sub-model, which is typical in cross-validation literature. Assumption 4 (c) argues that the average difference between the Fisher information matrix of the validation and the construction samples are close as $N,T\to\infty$. Assumption 4 (d) complements Assumption 1 (e) as when $s^c=\{1,\ldots,T-1\}$, $\frac{1}{\sqrt{T\cdot\kappa}}\sum_{t\,\in\,s^c}\sum_{u=1}^{T}\frac{1}{N}\sum_{i,j}|\sigma_{ij,tu}|=\frac{1}{TN}\sum_{t,u,i,j}^{T}|\sigma_{ij,tu}|\le C$. Assumption 4 (e) and Assumption 4 (d) strengthen Assumption 2 (d) and Assumption 3 (b), respectively. They are used for proving Lemma 7.2. Next theorem proves the consistency of the Monte Carlo leave-d-out cross-validation for FAR.

Theorem 1. *Suppose that Assumptions 1–4 hold. Suppose further that* $\frac{\kappa}{C_{NT}^2}\to0$, $\frac{T^2}{\kappa^2 b}\to0$ *and* κ, $d\to\infty$, *when* b, T, $N\to\infty$. *Then*

$$P\left(\hat{m}=m_0\right)\to1,$$

where $\hat{m}=\arg\min_m CV_d(m)$, *if* $\mathcal{M}$ *contains at least one correct set of estimated factors.*

The proof of Theorem 1 is given in Appendix. This result is an extension of Shao (1993) to the case with generated regressors. Given the rate conditions, κ, $d\to\infty$ such that $\frac{\kappa}{T-1}\to0$ and $\frac{d}{T-1}\to1$. It follows from Theorem 3.1 that the consistency of the Monte Carlo leave-d-out cross-validation relies on κ much smaller than d. One could consider $\kappa=\min\{T,\ N\}^{3/4}$ and $d=(T-1)-\kappa$ as they are consistent with the conditions in Theorem 1. In particular, Shao (1993) suggests for the observed regressors framework $\kappa=T^{3/4}$. This difference is due to the presence of the factor estimation error which should converge faster to zero relative to the complexity term. An extreme case where this condition is not satisfied is the leave-one-out cross-validation where

$\kappa = (T-1) - 1$ and $d = 1$. Next paragraph studies an alternative selection procedure using the bootstrap methods.

3.2 Bootstrap rule for model selection

It follows from the previous subsection that the improvement in the Monte Carlo leave-d-out cross-validation relies in its ability to capture the complexity and the identifiability component in the conditional mean of the infeasible in-sample squared error L_T. This is obtained by making the complexity component vanishes at a slower rate than the parameter and factor estimation errors. An alternative way to achieve the same purpose is using a bootstrap approach.

The suggested bootstrap model selection procedure generalizes the result of Shao (1996) to the factor-augmented regressions context where we have generated regressors. We define $\hat{\Gamma}_\kappa(m)$, a bootstrap estimator of the prediction error mean conditionally to Z which is $\sigma^2 + \mathrm{E}(L_T(m)|Z)$, based on the two-step residual procedure proposed by Gonçalves and Perron (2014) for FAR. In the case with observed regressors, Shao (1996) considers

$$\hat{\Gamma}_\kappa(m) = \mathrm{E}^*\left(\frac{1}{T-1}\left\|y - Z(m)\hat{\delta}_d^*(m)\right\|^2\right),$$

where $\hat{\delta}_\kappa^*(m) = (Z(m)Z(m))^{-1} Z(m) y^*$ is the bootstrap estimator of δ using a residual bootstrap scheme. E^* represents the expectation in the bootstrap world which is conditional on the data. While fixing $Z^*(m) = Z(m)$, the bootstrap version of y is given by $y^* = Z(m)\hat{\delta} + \varepsilon^*$, with ε^* the i.i.d. resampled version of $\hat{\varepsilon}$ multiplied by $\sqrt{\frac{T-1}{\kappa}}\frac{1}{\sqrt{1-\frac{r+q}{T-1}}}$, where $\kappa \to \infty$ such that $\frac{\kappa}{T-1} \to 0$. When $\kappa = T-1$, we obtain to the usual residual bootstrap. In fact, the factor $\sqrt{\frac{T-1}{\kappa}}$ ensures $\hat{\delta}_d^*(m)$ to converge to δ at a slower rate $\sqrt{\kappa}$ useful for consistent model selection rather than the usual $\sqrt{T}$. As for the leave-d-out cross-validation, $\kappa = o(T)$ such that $\frac{\kappa}{T-1} \to 0$ and $\frac{d}{T-1} \to 1$. If $\kappa = O(T)$, we have similarly to the leave-one-out cross-validation, a naive estimator of L_T up to the constant σ^2 which does not choose the smallest model in $\mathcal{M}_2$ with probability going to one. In our set-up, to mimic the estimation of F by $\tilde{F}$ from X, $\tilde{F}^*$ is extracted from the bootstrap sample X^* and $\hat{Z}^* = \left(\tilde{F}^*,\ W\right)$.Subsets of $\tilde{F}^*$ are denoted by $\tilde{F}^*(m)$. We also define $\hat{Z}^*(m) = \left(\tilde{F}^*(m),\ W\right)$ and

$$\hat{\Gamma}_\kappa(m) = \mathrm{E}^*\left(\frac{1}{T-1}\left\|y - \hat{Z}^*(m)\hat{\delta}_\kappa^*(m)\right\|^2\right),$$

where

$$\hat{\delta}_\kappa^*(m) = \left(\hat{Z}^{*\prime}(m)\hat{Z}^*(m)\right)^{-1}\hat{Z}^{*\prime}(m) y^*(m) \qquad (4)$$

with $\hat{Z}^*(m)$ and $y^*(m)$ the bootstrap analog of $\hat{Z}(m)$ and $y(m)$, respectively, obtained through the following algorithm.

Algorithm

A) Estimate $\tilde{F}$ and $\tilde{\Lambda}$ from X.

B) For each m:

1. Compute $\hat{\delta}(m)$ by regressing y on $\hat{Z}(m)$.
2. Generate B bootstrap samples such that $X^*_{it} = \tilde{F}'_t\tilde{\lambda}_i + e^*_{it}$, $y^*(m) = \hat{Z}(m)\hat{\delta}(m) + \varepsilon^*$ where $\{e^*_{it}\}$ and $\{\varepsilon^*_{t+1}\}$ are re-sampled residual based respectively on $\{\hat{e}_{it}\}$ and $\{\hat{\varepsilon}_{t+1}\}$, with $\hat{\varepsilon}_{t+1} = \hat{\varepsilon}_{t+1}(M)$ and M is the residual when all the estimated factors are used.
   (a) $\{e^*_{it}\}$ are obtained by multiplying $\{\hat{e}_{it}\}$ i.i.d.$(0,1)$ external draws η_{it} for $i = 1,\ldots,N$ and $t = 1,\ldots,T$.
   (b) $\{\varepsilon^*_{t+1}\}_{t=1,\ldots,T-1}$ are i.i.d. draws of $\left\{\sqrt{\frac{T-1}{\kappa}}\frac{1}{\sqrt{1-\frac{r+q}{T-1}}}\left(\hat{\varepsilon}_{t+1}(M) - \overline{\hat{\varepsilon}(M)}\right)\right\}_{t=1,\ldots,T-1}$.
3. For each bootstrap sample, extract $\tilde{F}^*$ from X^* and estimate $\hat{\delta}^*_\kappa(m)$ based on $\hat{Z}^*(m) = \left(\tilde{F}^*(m), W\right)$ and $y^*(m)$ using (3.4).

C) Obtain $\hat{m}$ as the model that minimizes the average of $\hat{\Gamma}^j_\kappa(m) = \frac{1}{T-1}\left\|y - \hat{Z}^{*j}(m)\hat{\delta}^{*j}_\kappa(m)\right\|^2$ over the B samples indexed by j, where

$$\hat{\Gamma}_\kappa(m) = \sum_{j=1}^{B}\hat{\Gamma}^j_\kappa(m).$$

By multiplying the second-step i.i.d. bootstrap residuals by $\frac{\sqrt{T-1}}{\sqrt{\kappa}}$, we obtain $\hat{\Gamma}_\kappa(m) = \frac{\varepsilon'\varepsilon}{T-1} + \frac{(r(m)+q)}{\kappa}\sigma^2 + o_P\left(\frac{1}{\kappa}\right)$ for m in $\mathcal{M}_2$ and $\hat{\Gamma}_\kappa(m) = \sigma^2 + \frac{1}{T-1}\mu'(I - P(m))\mu + o_P(1)$ for m in $\mathcal{M}_1$, which achieves a consistent selection. The next theorem proves the validity of the described bootstrap scheme.

Theorem 2. *Suppose that Assumptions 1–3 hold. Suppose further that Assumptions 6-8 of Gonçalves and Perron (2014) and* $E^*|\eta_{it}|^4 \leq C < \infty$ *hold. If* N, $T \longrightarrow \infty$ *and* $\kappa \longrightarrow \infty$ *such that* $\frac{\kappa}{C^2_{NT}} \longrightarrow 0$ *then*

$$\sqrt{\kappa}\left(\hat{\delta}^*_\kappa(m) - \Phi^*_0(m)\hat{\delta}(m)\right) \rightarrow^{d^*} N\left(0,\ \Sigma_{\delta^*(m)}\right)$$

for any m *with* $\Sigma_{\delta^*(m)} = \sigma^2\left[\Phi^*_0(m)\Sigma_Z(m)\Phi^{*\prime}_0(m)\right]^{-1}$ *and* $\Sigma_Z(m) = \text{plim}_{T\to\infty}\frac{1}{T}Z(m)'Z(m)$.

From Theorem 2, it follows that $\hat{\delta}^*_\kappa(m)$ converges to the limit of $\Phi^*_0(m)\hat{\delta}(m)$ at a lower rate $\sqrt{\kappa} = o\left(\sqrt{T}\right)$. The proof in Appendix shows that our bootstrap scheme satisfies the high level conditions provided by Gonçalves and Perron (2014). This result allows us to use our new bootstrap scheme for the following optimality results.

Theorem 3. *Suppose that Assumptions 1–3 and Assumption 4 (a) complemented by Assumptions 6-8 of Gonçalves and Perron (2014) hold. Suppose further that* $\kappa \to \infty$ *such that* $\frac{\kappa}{C^2_{NT}} \to 0$ *as* T,

$N \to \infty$ and $E^* |\eta_{it}|^4 \leq C < \infty$. Then if $\mathcal{M}_2$ is not empty, it holds that

$$\lim_{N,\, T \to \infty} P(\hat{m} = m_0) = 1$$

where $\hat{m} = \arg\min_m \hat{\Gamma}_\kappa(m)$.

This bootstrap result is the analog of Theorem 1. The following section compares the different procedures through a simulation study.

4 Simulation experiment

To investigate the finite sample properties of the proposed model selection methods, Monte Carlo simulations are conducted. We consider the following model

$$y_{t+1} = \alpha' F_t^0 + \alpha_0 + \varepsilon_{t+1},$$

where $\alpha_0 = 1$, $F_t^0 \subset F_t \sim i.i.d.\mathrm{N}(0, \mathrm{I}_4)$ and $\varepsilon_{t+1} \sim i.i.d.\mathrm{N}(0,1)$. Three data generating process (DGP) are used.

- For DGP1, $r_0 = 2$, $F_t^0 = (F_{t,1},\ F_{t,2})'$ and $\alpha = (1,\ 1/2)'$.
- For DGP2, $r_0 = 3$, $F_t^0 = (F_{t,1},\ F_{t,2},\ F_{t,3})'$ and $\alpha = (1,\ 1/2,\ -1)'$.
- For DGP3, $r_0 = 4$, $F_t^0 = (F_{t,1},\ F_{t,2},\ F_{t,3},\ F_{t,4})'$ and $\alpha = (1,\ 1/2,\ -1,\ 2)'$.

There are 4 factors, but only DGP 3 uses them all. DGP 1 and 2 only use a subset of them to generate the variable of interest y_{t+1}. The panel factor model is a matrix of dimension $N \times T$, with elements

$$X_{it} = \lambda_i' F_t + e_{it},$$

where $\lambda_{1i} \sim 12\mathrm{U}[0,1]$, $\lambda_{2i} \sim 8\mathrm{U}[0,1]$, $\lambda_{3i} \sim 4\mathrm{U}[0,1]$ and $\lambda_{4i} \sim \mathrm{U}[0,1]$. The factor loadings are labelled in decreasing order of importance to explain the dynamics of the panel X_{it}. The specification for the unobserved factors and the factor loadings satisfies asymptotically PC1 identifying restrictions provided by Bai and Ng (2013). Indeed, $\operatorname{plim}_{T\to\infty} \frac{1}{T} F'F = I_4$ and $\operatorname{plim}_{N\to\infty} \frac{1}{N}\Lambda'\Lambda$ is diagonal with distinct entries, and make possible to identify estimated factors as N, $T \to \infty$ go to infinity. As in Djogbenou, Gonçalves, and Perron (2015), $e_{it} \sim \mathrm{N}(0, \sigma_i^2)$ with $\sigma_i^2 \sim \mathrm{U}[.5,\ 1.5]$. We consider 1000 replications, for bootstrap and Monte Carlo, 399 simulations and for sample sizes $T \in \{100,\ 200\}$, $N \in \{50,\ 100,\ 150,\ 200, 250\}$. The construction data size for the CV_d and for the bootstrap is $\kappa = (\min\{T, N\})^{3/4}$. The first step bootstrap residual are obtained by the wild bootstrap using i.i.d. normal with mean 0 and variance 1 external draws.

We compare the ability of the proposed procedures to select consistently the true model to the leave-one-out cross-validation

$$CV_1(m) = \frac{1}{T-1} \sum_{t=1}^{T-1} \left(y_{t+1} - \hat{\delta}_t'(m) \tilde{F}_t(m) - \hat{\alpha}_0 \right)^2$$

and the modified bayesian information criteria (BICM) suggested by Groen and Kapetanios (2013)

$$BICM(m) = \frac{T}{2} \ln \left(\hat{\sigma}^2(m) \right) + r(m) \ln(T) \left(1 + \frac{T}{N} \right),$$

where $\hat{\sigma}^2(m) = \frac{1}{T-1-r(m)} \left\| y - \tilde{F}(m) \hat{\delta}(m) - \hat{\alpha}_0 \right\|^2$, is made by considering subsets of the first four principal component estimated factors.

Table 1 present the average number of selected estimated factors whereas Tables 2–4 show the frequencies of selecting the consistent set of estimated factors over the $2^4 = 16$ possibilities including the case of no factor. Except for the largest estimated model, where the average number of estimated factors tends to be close to four, the CV_1 tends to overestimate the true number of factors. The BICM very often selects a smaller set of estimated factors than the true one. The leave-d-out cross-validation and the bootstrap procedure select in average a number of factor close to the true number.

The suggested procedures offer a higher frequency of selecting factor estimates that span the true model for DGP 1 and 2. In particular, when $N = 100$ and $T = 200$, for DGP 1, the frequency of selecting the first two estimated factors is 68.30 using the modified BIC and 64.20 using the leave-one-out cross-validation. The bootstrap selection method increases the frequency of the CV_1 by 27.5 points of percentage and the CV_d increases it by 29.2 points of percentage. These frequencies increase with the sample sizes. In general, the leave-one-out cross-validation very often selects a larger model than the true one and the modified BIC tends to pick smaller subset of the consistent model. As DGP 3 corresponds to the largest model, CV_1 unsurprisingly performs well.

5 Empirical application

This section revisits the factor analysis of excess risk premia of Ludvigson and Ng (2007). The data set contains 147 quarterly financial series and 130 quarterly macroeconomic series from the first quarter of 1960 to the third quarter of 2014. The variables in the financial dataset are constructed using Jurado, Ludvigson, and Ng (2015) financial dataset and variables from Kenneth R. French website as described in the Supplemental Appendix [1]. The quarterly macro data are downloaded from the St. Louis Federal Reserve website and correspond to the monthly series constructed by McCracken and Ng (2015). Some of the quarterly data are also constructed based on McCracken and Ng (2015) data as explained in the Supplemental Appendix. We examine how economic

[1]We gratefully thank Sydney C. Ludvigson who provided us their data set.

information summarized through a few numbers of estimated factors from real economic activities data and those related to financial markets can explain the future excess returns using various selection procedures. Recently, Gonçalves, McCracken, and Perron (2017) also study the predictive ability of estimated factors from the macroeconomic data provided by McCracken and Ng (2015) to forecast excess returns to the S&P 500 Composite Index. They detect the interest rate factor as the strongest predictor of the equity premium. Indeed, as argued by Ludvigson and Ng (2007), restricting attention to a few sets of observed factors may not span all information related to financial market participants. Unlike Gonçalves, McCracken, and Perron (2017), they considered both financial and macroeconomic data. Using the BIC, they found three new estimated factors termed "volatility", "risk premium" and "real" factors that have predictive power for the market excess returns after controlling for usual observed factors.

Following Ludvigson and Ng (2007), we define m_{t+1} as the continuously compounded one-quarter-ahead excess returns in period t +1 obtained by computing the log return on the Center for Research in Security Prices (CRSP) value-weighted price index for NYSE, AMEX and NASDAQ minus the three-month Treasury bill rate. The factor-augmented regression model used by Ludvigson and Ng (2007) takes the form,

$$m_{t+1} = \alpha_1' F_t + \alpha_2' G_t + \beta' W_t + \varepsilon_{t+1}.$$

The variables F_t and G_t are latent and represent respectively the macroeconomic and the financial factors. The vector W_t contains commonly used observable predictors that may help predict excess returns and the constant. The observed predictors are essentially those studied by Ludvigson and Ng (2007). We have the dividend price ratio (d-p) introduced by Campbell and Shiller (1989), the relative T-bill (RREL) from Campbell (1991) and the consumption-wealth variable suggested by Lettau and Ludvigson (2001). In addition, the lagged realized volatility is computed over each quarter (see, Ludvigson and Ng (2007)) and included. The factors are estimated by $\tilde{F}_t$ and $\tilde{G}_t$ using principal components based respectively on the macro factor panel model

$$X_{1it} = \lambda_i' F_t + e_{1it}$$

and the financial factor panel model

$$X_{2it} = \gamma_i' G_t + e_{2it}.$$

Like Ludvigson and Ng (2007), we use the IC_{p2} information criterion of Bai and Ng (2002) and select six estimated factors from each set that summarize 54.87% of the information in our macroeconomic series and 83.64% of the financial information. Despite the imperfection of naming an estimated factor, it turns to be interesting as it helps us understand the economic message revealed by the data. Figures in the Supplemental Appendix represent the marginal R^2 of each variables to each estimated factor, obtained by regressing each estimated factor on the variables.

In the panel, similarly to McCracken and Ng (2015), $\tilde{F}_1$ is revealed as a real factor because variables related to production and labor market are highly correlated to it. The third factors $\tilde{F}_3$ represents an interest rate spread factor. The estimated financial factors $\tilde{G}_2$ and $\tilde{G}_3$ are market risk factors. The market excess returns and the High Minus Low Fama-French factors have a marginal R^2 greater than 0.7 with $\tilde{G}_2$ whereas the the Small Minus Big Fama-French factor and Cochrane-Piazzesi factor have the highest correlation to $\tilde{G}_3$. The estimated factor $\tilde{G}_4$ is dominated by oil industry portfolio return and $\tilde{G}_6$ is mostly related to utility industry portfolio return.

The next two subsections study the in-sample and out-of-sample excess returns prediction while picking consistently the estimated factors in the second step.

5.1 In-sample prediction of excess returns

The estimated regression takes the form

$$m_{t+1} = \alpha_1' (m) \tilde{F}_t (m) + \alpha_2' (m) \tilde{G}_t (m) + \beta' Z_t + u_{t+1} (m)$$

for $m = 1, \ldots, 2^r$ including the possibility that no factor is selected, with r the number of selected factors in the first step. The selected model and the estimated regression results are reported in Table 5.

The Monte Carlo cross-validation and the bootstrap selection procedures select smaller set of generated regressors than the leave-one-out cross-validation. On the other hand, BICM selects the model with no financial or macro factor. Our cross-validation method selects three factors: the third macro factor $\left(\tilde{F}_{3t}\right)$, the second financial factor $\left(\tilde{G}_{2t}\right)$ and the third financial factor $\left(\tilde{G}_{3t}\right)$. Investors care about the spread between interest rates and effective federal funds rate motivating interventions by the Federal Reserve to impulse economic expansion. Estimated risk factors also play an important role in predicting the equity premium associated to U.S. stock market as in Ludvigson and Ng (2007). We can deduce that the important estimated factors which investors in the U.S. financial market should care about are interest rate spread factor $\left(\tilde{F}_{3t}\right)$, and market risk factors $\left(\tilde{G}_{2t}\right)$ and $\left(\tilde{G}_{3t}\right)$. These factors are significant and simultaneously picked by the leave-d-out cross-validation and bootstrap model selection approach. We also study the joint significativity of the estimated factors using the Fisher test. The constrained model is the one estimated with only observed regressors and the volatility factor, whereas the unconstrained model is $\hat{m}_j$, $j = 1, \cdots, 4$. The estimated models $\hat{m}_1$, $\hat{m}_2$, $\hat{m}_3$ and $\hat{m}_4$ correspond respectively to those selected by the CV_1, the BICM, the CV_d and the $\hat{\Gamma}_\kappa$. The F-test statistic is always greater than the 5% critical value, implying additional significant information in the unconstrained model for the different procedures except the BICM where no factor is selected.

5.2 Out-of-sample prediction of excess returns

This subsection studies how the new procedures behave in out-of-sample forecasting. Parameters and factors are estimated recursively with an initial sample of data from 1960:1 through 2004:4. The forecasts are generated for each subsample based on the model selected over that subsample by each criterion. The forecast sample corresponds to the period 2005:1-2014:3. This forecasts are obtained by regressing the dependent variable from 1960:2-2005:1 on the independent variables from 1960:1-2004:12. The used estimated factors are extracted from the large data set covering 1960:1 to 2004:4. From this estimated factors, a subset is selected using each of the four criteria. This procedure is repeated by expanding data sample each quarter, re-estimating the factors and selecting a new set of factors to forecast next quarter's excess returns. The number of estimated factors that summarizes information in X_1 and X_2 are selected using Bai and Ng (2002) criterion from the first estimation sample and maintained in each recursion.

We compare the different sets of model by computing the MSE relative to the benchmark which only contains the constant. The alternative model contains generated regressors selected using the CV_1, the BICM, the CV_d or the $\hat{\Gamma}_\kappa$. Hence, we compute MSE_u/MSE_r the out-of-sample mean squared error of each unconstrained model relative to MOD_0, where no factor is present. The forecast error is smaller in the bootstrap selection procedure as we obtain $MSE_u/MSE_r = 0.8552$ for $\hat{\Gamma}_\kappa$, $MSE_u/MSE_r = 0.9800$ for the CV_d, $MSE_u/MSE_r = 1.3168$ for the CV_1 and $MSE_u/MSE_r = 1.0000$ for the BICM. Another method of gauging the out-of-sample is to test the equal predictive ability of out-of-sample forecasts as considered by Gonçalves, McCracken, and Perron (2017). Because in each recursion, a new number of estimated factors is selected, there are no available critical values for such a situation. Figure 3 indicates how the number of selected factors varies while the estimation sample changes. During the forecast exercise, while the BICM never selects an estimated factor, the leave-one-out cross-validation always chooses a larger model than our proposed methods. As is argued in the previous sections, the new consistent model selection approaches prevent against too much under-fitting and over-fitting.

6 Conclusion

This paper suggests and provides conditions for the validity of two consistent model selection procedures for the factor-augmented regression models. It is the Monte Carlo leave-d-out cross-validation and the bootstrap selection approach. In finite samples, the simulations document improvement in the probability of selecting the smallest set of estimated factors that span the true model comparatively to other existing methods. The procedures in this paper have been used to select estimated factors for in-sample and out-of-sample predictions of one-quarter-ahead excess stock returns on U. S. market. The in-sample analysis reveals that the estimated factor highly correlated to interest rate spreads and the generated regressor highly correlated to the Fama-French factors drive the underlying unobserved factors, and strongly predict the excess returns. Moreover, the out-of-sample forecasts lead to a smaller forecast error using our suggested procedures. For future research, an important extension of the results in this paper is to allow the inclusion of the non linear factors and the possibility of interaction between the factors.

7 Appendix: Proofs of results in Section 3, Simulation Results and Empirical Application Details

7.1 Proofs of results in Section 3

Denote $\tilde{P}(m) = \tilde{Z}(m)\left(\tilde{Z}(m)'\tilde{Z}(m)\right)^{-1}\tilde{Z}(m)'$ and $C_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$. We state the following two auxiliary results.

Lemma 7.1. *Under Assumptions 1–3,*

$$\frac{1}{T-1}\mu'\tilde{P}(m)\mu = \mu' P(m)\mu + O_P\left(\frac{1}{C_{NT}^2}\right)$$

and

$$\frac{1}{T-1}\varepsilon'\tilde{P}(m)\varepsilon = \varepsilon' P(m)\varepsilon + O_P\left(\frac{1}{C_{NT}^2}\right)$$

for any $m \in \mathcal{M}$.

Lemma 7.2. *Under Assumptions 1–4, as* b, T, $N \to \infty$,

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\tilde{F}_t - HF_t\right\|^4 = O_P\left(\frac{\kappa}{T}\right) + O_P\left(\frac{\kappa}{N^2}\right).$$

We now present the proofs of the auxiliary results followed by those of the main results.

Proof of Lemma 7.1. We have the following decomposition

$$\begin{aligned}
\frac{1}{T-1}\mu'\tilde{P}(m)\mu &= \delta'\left(\frac{1}{T-1}Z^{0\prime}\tilde{Z}(m)\right)\left(\frac{1}{T-1}Z(m)'Z(m)\right)^{-1}\left(\frac{1}{T-1}\tilde{Z}(m)'Z^0\right)\delta \\
&\quad +\delta'\left(\frac{1}{T-1}Z^{0\prime}\tilde{Z}(m)\right)\left[\left(\frac{1}{T-1}\tilde{Z}(m)'\tilde{Z}(m)\right)^{-1} - \left(\frac{1}{T-1}Z(m)'Z(m)\right)^{-1}\right] \\
&\quad \times\left(\frac{1}{T-1}\tilde{Z}(m)'Z^0\right)\delta \\
&\equiv \frac{1}{T-1}\mu' P(m)\mu + L_{1T}(m) + L_{2T}(m) + L_{3T}(m),
\end{aligned}$$

where we use $\tilde{Z}(m) = Z(m) + \left(\tilde{Z}(m) - Z(m)\right)$ obtain

$$\begin{aligned}
L_{1T}(m) &= \delta'\left(\frac{1}{T-1}Z^{0\prime}\left[\tilde{Z}(m) - Z(m)\right]\right)\left(\frac{1}{T-1}Z(m)'Z(m)\right)^{-1}\left(\frac{1}{T-1}\left[\tilde{Z}(m) - Z(m)\right]'Z^0\right)\delta, \\
L_{2T}(m) &= 2\delta'\left(\frac{1}{T-1}Z^{0\prime}Z(m)\right)\left(\frac{1}{T-1}Z(m)'Z(m)\right)^{-1}\left(\frac{1}{T-1}\left[\tilde{Z}(m) - Z(m)\right]'Z^0\right)\delta
\end{aligned}$$

and

$$L_{3T}(m) = \delta' \left(\frac{1}{T-1} Z^{0\prime} \tilde{Z}(m)\right) \left[\left(\frac{1}{T-1} \tilde{Z}(m)' \tilde{Z}(m)\right)^{-1} - \left(\frac{1}{T-1} Z(m)' Z(m)\right)^{-1}\right] \left(\frac{1}{T-1} \tilde{Z}(m)' Z^{0}\right) \delta.$$

To find the order of $L_{1T}(m)$, it will be sufficient to study $\frac{1}{T-1} Z' \left[\tilde{F}(m) - F(m)\right]$ as $\left(\frac{1}{T-1} Z(m)' Z(m)\right)^{-1} = O_P(1)$ using Assumption 3 (b). From Gonçalves and Perron (2014, Lemma A.2), $\frac{1}{T-1} F' \left[\tilde{F} - FH'\right] = O_P\left(\frac{1}{C_{NT}^2}\right)$ and $\frac{1}{T-1} W' \left[\tilde{F} - FH'\right] = O_P\left(\frac{1}{C_{NT}^2}\right)$, and it follows that $\frac{1}{T-1} Z' \left[\tilde{F}(m) - F(m)\right] = O_P\left(\frac{1}{C_{NT}^2}\right)$. Indeed, from their proof of Lemma A.2 (b),

$$\frac{1}{T-1} F' \left[\tilde{F} - FH'\right] = (b_{f1} + b_{f2} + b_{f3} + b_{f4}) \tilde{V}^{-1}, \tag{5}$$

where $b_{f1} = O_P\left(\frac{1}{C_{NT} T^{1/2}}\right)$, $b_{f2} = O_P\left(\frac{1}{N^{1/2} T^{1/2}}\right)$, $b_{f3} = O_P\left(\frac{1}{N^{1/2} T^{1/2}}\right)$, $b_{f4} = O_P\left(\frac{1}{N}\right) + O_P\left(\frac{1}{N^{1/2} T^{1/2}}\right)$ and $\tilde{V}^{-1} = O_P(1)$. Hence, $\frac{1}{T-1} F' \left[\tilde{F} - FH'\right] = O_P\left(\frac{1}{C_{NT}^2}\right)$, similarly $\frac{1}{T-1} W' \left[\tilde{F} - FH'\right] = O_P\left(\frac{1}{C_{NT}^2}\right)$, thus $L_{1T}(m) = O_P\left(\frac{1}{C_{NT}^4}\right)$ for any m. Since $\frac{1}{T-1} Z^{0\prime} Z(m) = O_P(1)$, using similar arguments as in the proof of L_{1T}, we have $L_{2T}(m) = O_P\left(\frac{1}{C_{NT}^2}\right)$, for any m. To finish, $L_{3T}(m) = O_P\left(\frac{1}{C_{NT}^2}\right)$ as

$$\left(\frac{1}{T-1} \tilde{Z}(m)' \tilde{Z}(m)\right)^{-1} - \left(\frac{1}{T-1} Z(m)' Z(m)\right)^{-1} = O_P\left(\frac{1}{C_{NT}^2}\right).$$

Indeed, $\left(\frac{1}{T-1} \tilde{Z}(m)' \tilde{Z}(m)\right)^{-1} - \left(\frac{1}{T-1} Z(m)' Z(m)\right)^{-1}$, for any m, can be decomposed as

$$\left(\frac{1}{T-1} \tilde{Z}(m)' \tilde{Z}(m)\right)^{-1} (A_{01}(m) + A_{02}(m)) \left(\frac{1}{T-1} Z(m)' Z(m)\right)^{-1},$$

which is $O_P\left(\frac{1}{C_{NT}^2}\right)$, since $A_{01}(m) = \frac{1}{T-1} \left(Z(m) - \tilde{Z}(m)\right)' \tilde{Z}(m) = O_P\left(\frac{1}{C_{NT}^2}\right)$ and $A_{02}(m) = \frac{1}{T-1} Z(m)' \left(Z(m) - \tilde{Z}(m)\right) = O_P\left(\frac{1}{C_{NT}^2}\right)$, using Gonçalves and Perron (2014, Lemma A.2). Using the order in probability of $L_{1T}(m)$, $L_{2T}(m)$ and $L_{3T}(m)$, we conclude that $\frac{1}{T-1} \mu' \tilde{P}(m) \mu = \frac{1}{T-1} \mu' P(m) \mu + O_P\left(\frac{1}{C_{NT}^2}\right)$. The proof of the second part of Lemma 7.1 follows identical steps. □

Proof of Lemma 7.2. The proof uses the following identity

$$\tilde{F}_t - HF_t = \tilde{V}^{-1} (A_{1t} + A_{2t} + A_{3t} + A_{4t})$$

$$A_{1t} = \frac{1}{T} \sum_{u=1}^{T} \tilde{F}_u \gamma_{ut}, \; A_{2t} = \frac{1}{T} \sum_{u=1}^{T} \tilde{F}_u \zeta_{ut}, \; A_{3t} = \frac{1}{T} \sum_{u=1}^{T} \tilde{F}_u \eta_{ut}, \; A_{4t} = \frac{1}{T} \sum_{u=1}^{T} \tilde{F}_u \xi_{ut}$$

where $\gamma_{ut} = \mathrm{E}\left(\frac{1}{N} \sum_{i=1}^{N} e_{iu} e_{it}\right)$, $\zeta_{ut} = \frac{1}{N} \sum_{i=1}^{N} \left(e_{iu} e_{it} - \mathrm{E}\left(\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{e}_{\mathrm{iu}} \mathrm{e}_{\mathrm{it}}\right)\right)$, $\eta_{ut} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i' F_u e_{it}$, and $\xi_{ut} = \frac{1}{N}\sum_{i=1}^{N} \lambda_i' F_t e_{iu}$. By the c-r inequality,

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\tilde{F}_t - HF_t\right\|^4 \leq 4^3\left\|\tilde{V}^{-1}\right\|^4\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\left(\sum_{t\,\in\,s^c}\|A_{1t}\|^4 + \sum_{t\,\in\,s^c}\|A_{2t}\|^4 + \sum_{t\,\in\,s^c}\|A_{3t}\|^4 + \sum_{t\,\in\,s^c}\|A_{4t}\|^4\right).$$

Using the Cauchy-Schwartz inequality, we have

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\in\,s^c}\|A_{1t}\|^4 = \frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\frac{1}{T}\sum_{u=1}^{T}\tilde{F}_u\gamma_{ut}\right\|^4 \leq \frac{\kappa}{T}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_u\right\|^2\right)^2\frac{1}{b}\sum_{s\in\,\mathcal{R}}\left[\frac{1}{\sqrt{\kappa\cdot T}}\sum_{t\,\in\,s^c}\sum_{u=1}^{T}\gamma_{ut}^2\right]^2$$

In addition, $\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2 = O_P(1)$ and $\frac{1}{\sqrt{T\cdot\kappa}}\sum_{t\,\in\,s^c}\sum_{u=1}^{T}\gamma_{ut}^2 \leq C$ for any $s \in \mathcal{R}$ (because $\frac{1}{\sqrt{T\cdot\kappa}}\sum_{t\,\in\,s^c}\sum_{u=1}^{T}\gamma_{ut}^2 \leq C$ using the proof of Bai and Ng (2002, Lemma 1 (i))). In consequence,

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{1t}\|^4 = O_P\left(\frac{\kappa}{T}\right). \tag{6}$$

By repeated application of Cauchy-Schwarz inequality,

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{2t}\|^4 = \frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\frac{1}{T}\sum_{u=1}^{T}\tilde{F}_u\zeta_{ut}\right\|^4 \leq \left[\frac{1}{T^2}\sum_{u=1}^{T}\sum_{u_1=1}^{T}\left(\tilde{F}_u'\tilde{F}_{u_1}\right)^2\right]\left[\frac{1}{T^2}\sum_{u=1}^{T}\sum_{u_1=1}^{T}\left(\sum_{t\,\in\,s^c}\zeta_{ut}^2\zeta_{u_1t}^2\right)\right].$$

Hence,

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t=1}^{T}\|A_{2t}\|^4 \leq \left[\frac{1}{T}\sum_{u_1=1}^{T}\left\|\tilde{F}_{u_1}\right\|^2\right]^2\left[\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\frac{1}{T^2}\sum_{u_1=1}^{T}\sum_{u=1}^{T}\left(\sum_{t\,\in\,s^c}\zeta_{u_1t}^2\zeta_{ut}^2\right)\right],$$

where $\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2 = O_P(1)$ and $\mathrm{E}\left[\frac{1}{\mathrm{b}}\sum_{\mathrm{s}\,\in\,\mathcal{R}}\frac{1}{\mathrm{T}^2}\sum_{\mathrm{u_1=1}}^{\mathrm{T}}\sum_{\mathrm{u=1}}^{\mathrm{T}}\left(\sum_{\mathrm{t}\,\in\,\mathrm{s}^c}\zeta_{\mathrm{u_1t}}^2\zeta_{\mathrm{ut}}^2\right)\right] = \mathrm{O}\left(\left(\frac{\sqrt{\kappa}}{\mathrm{N}}\right)^2\right)$. Indeed,

$$\begin{aligned}\mathrm{E}\left[\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\frac{1}{T^2}\sum_{u_1=1}^{T}\sum_{u=1}^{T}\left(\sum_{t\,\in\,s^c}\zeta_{u_1t}^2\zeta_{ut}^2\right)\right] &\leq \frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\frac{1}{T^2}\sum_{u_1=1}^{T}\sum_{u=1}^{T}\sum_{t\,\in\,s^c}\left[\mathrm{E}\left(\zeta_{u_1t}^4\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left(\zeta_{ut}^4\right)\right]^{\frac{1}{2}}\\ &\leq \kappa\left[\max_{u,t}\mathrm{E}\left(\zeta_{ut}^4\right)\right] = O\left(\frac{\kappa}{N^2}\right),\end{aligned}$$

since $\max_{u,t}\mathrm{E}\left(\zeta_{ut}^4\right) = O\left(\frac{1}{N^2}\right)$ by Assumption 1 (e). Thus,

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{2t}\|^4 = O_P\left(\frac{\kappa}{N^2}\right). \tag{7}$$

Thirdly, as $\frac{1}{b\cdot\kappa}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\frac{1}{N^{1/2}}\Lambda e_t\right\|^4 = O_P(1)$ by Assumption 4 (e), we can write

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{3t}\|^4 = \frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\frac{1}{T}\frac{1}{N}\sum_{u=1}^{T}\tilde{F}_uF_u'\Lambda e_t\right\|^4 \leq \frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\frac{1}{N}\Lambda e_t\right\|^4\left\|\frac{1}{T}\sum_{u=1}^{T}\tilde{F}_uF_u'\right\|^4,$$

which implies that $\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{3t}\|^4$ is bounded by

$$\frac{\kappa}{N^2}\left[\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\frac{1}{\kappa}\sum_{t\,\in\,s^c}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_ie_t\right\|^4\right]\left(\frac{1}{T}\sum_{u=1}^{T}\left\|\tilde{F}_u\right\|^2\right)^2\left(\frac{1}{T}\sum_{u=1}^{T}\|F_u\|^2\right)^2 = O_P\left(\frac{\kappa}{N^2}\right),$$

since $\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2 = O_P(1)\,,\ \frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2 = O_P(1)$. Hence,

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{3t}\|^4 = O_P\left(\frac{\kappa}{N^2}\right) \tag{8}$$

The proof that

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{4t}\|^4 = O_P\left(\frac{\kappa}{N^2}\right) \tag{9}$$

uses $\frac{1}{T}\sum_{u=1}^{T}\left\|\tilde{F}_u\right\|^2 = O_P(1)\,,\ \frac{1}{b\cdot\kappa}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|F_t\|^4 = O_P(1)\,,\ \frac{1}{T}\sum_{u=1}^{T}\left\|\frac{1}{\sqrt{N}}\Lambda' e_u\right\|^2 = O_P(1)$ and the bound of $\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{4t}\|^4$ by

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|F_t\|^4\left[\frac{1}{T}\sum_{u=1}^{T}\left\|\tilde{F}_u\right\|\left\|\frac{1}{N}\Lambda' e_u\right\|\right]^4 \leq \frac{\kappa}{N^2}\frac{1}{b\cdot\kappa}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|F_t\|^4\left(\frac{1}{T}\sum_{u=1}^{T}\left\|\tilde{F}_u\right\|^2\right)^2\left(\frac{1}{T}\sum_{u=1}^{T}\left\|\frac{1}{\sqrt{N}}\Lambda' e_u\right\|^2\right)^2.$$

Finally, from Equations (7.1),(6), (7), (8) and (9), $\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\tilde{F}_t - HF_t\right\|^4 = O_P\left(\frac{\kappa}{T}\right) + O_P\left(\frac{\kappa}{N^2}\right)$. □

Proof of Lemma 3.1. To prove Lemma 3.1, we will first need to show that

$$\max_{1\leq t\leq T-1}\left\|\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right\| - \max_{1\leq t\leq T-1}\left\|Z_t(m)'\left(Z'(m)Z(m)\right)^{-1}Z_t(m)\right\| = o_P(1).$$

We have the following decomposition

$$\begin{aligned}
& \hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m) \\
= \ & \frac{1}{T-1}\hat{Z}_t'(m)\left[\left(\frac{1}{T-1}\hat{Z}'(m)\hat{Z}(m)\right)^{-1} - \left(\frac{1}{T-1}Z'(m)Z(m)\right)^{-1}\right]\hat{Z}_t(m) \\
& + \frac{1}{T-1}\left(\hat{Z}_t(m) - Z_t(m)\right)'\left(\frac{1}{T-1}Z'(m)Z(m)\right)^{-1}\left(\hat{Z}_t(m) - Z_t(m)\right) \\
& + \frac{2}{T-1}Z_t(m)'\left(\frac{1}{T-1}Z'(m)Z(m)\right)^{-1}\left(\hat{Z}_t(m) - Z_t(m)\right) \\
& + Z_t(m)'\left(Z'(m)Z(m)\right)^{-1}Z_t(m).
\end{aligned}$$

Hence, we can write

$$\begin{aligned}
& \max_{1\leq t\leq T-1}\left\|\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right\| \\
\leq \ & \frac{1}{T-1}\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)\right\|^2\left\|\left(\frac{1}{T-1}\hat{Z}'(m)\hat{Z}(m)\right)^{-1} - \left(\frac{1}{T-1}Z'(m)Z(m)\right)^{-1}\right\| \\
& + \frac{1}{T-1}\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m) - Z_t(m)\right\|^2\left\|\left(\frac{1}{T-1}Z'(m)Z(m)\right)^{-1}\right\| \\
& + \frac{2}{T-1}\max_{1\leq t\leq T-1}\|Z_t(m)\|\left\|\left(\frac{1}{T-1}Z'(m)Z(m)\right)^{-1}\right\|\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m) - Z_t(m)\right\| \\
& + \max_{1\leq t\leq T-1}\left\|Z_t(m)'\left(Z'(m)Z(m)\right)^{-1}Z_t(m)\right\|.
\end{aligned}$$

From that bound, it follows that

$$\left| \max_{1\leq t\leq T-1} \left\| \hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right\| - \max_{1\leq t\leq T-1}\left\| Z_t(m)'\left(Z'(m)Z(m)\right)^{-1}Z_t(m)\right\|\right|$$

is lower than $A_1(m)+A_2(m)+A_3(m)$ where

$$\begin{aligned} A_1(m) &= \frac{1}{T-1}\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)\right\|^2\left\|\left(\frac{1}{T-1}\hat{Z}'(m)\hat{Z}(m)\right)^{-1}-\left(\frac{1}{T-1}Z'(m)Z(m)\right)^{-1}\right\|, \\ A_2(m) &= \frac{1}{T-1}\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)-Z_t(m)\right\|^2\left\|\left(\frac{1}{T-1}Z'(m)Z(m)\right)^{-1}\right\| \end{aligned}$$

and

$$A_3(m) = \frac{2}{T-1}\max_{1\leq t\leq T-1}\|Z_t(m)\|\left\|\left(\frac{1}{T-1}Z'(m)Z(m)\right)^{-1}\right\|\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)-Z_t(m)\right\|.$$

Since $\frac{1}{T-1}\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)\right\|^2 \leq \frac{1}{T-1}\sum_{t=1}^{T-1}\left\|\hat{Z}_t(m)\right\|^2 = O_P(1)$ and $\left(\frac{1}{T-1}\hat{Z}'(m)\hat{Z}(m)\right)^{-1} - \left(\frac{1}{T-1}Z'(m)Z(m)\right)^{-1} = O_P\left(\frac{1}{C_{NT}^2}\right)$, we obtain that $A_1(m) = O_P\left(\frac{1}{C_{NT}^2}\right)$. Because, we have the bound

$$\frac{1}{T-1}\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)-Z_t(m)\right\|^2 \leq \frac{1}{T-1}\sum_{t=1}^{T-1}\left\|\hat{Z}_t(m)-Z_t(m)\right\|^2, \tag{10}$$

which is lower than $\frac{1}{T-1}\sum_{t=1}^{T-1}\left\|\hat{Z}_t(M)-Z_t(M)\right\|^2 = O_P\left(\frac{1}{C_{NT}^2}\right)$ (using Bai and Ng (2002, Theorem 1) with M denoting the set with all estimated factors), $A_2(m) = O_P\left(\frac{1}{C_{NT}^2}\right)$. From Bai (2003, Proposition 2), $\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)-Z_t(m)\right\| = O_P\left(\frac{1}{T^{1/2}}\right)+O_P\left(\frac{T^{1/2}}{N^{1/2}}\right)$, it follows that $A_3(m) = O_P\left(\frac{1}{T}\right)+O_P\left(\frac{1}{N^{1/2}}\right)$ as $\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)\right\| = O_P\left(T^{1/2}\right)$ (since $\frac{1}{T-1}\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)\right\|^2 = O_P(1)$). From the bounds of $A_1(m)$, $A_2(m)$ and $A_3(m)$, we deduce

$$\left| \max_{1\leq t\leq T-1} \left\| \hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right\| - \max_{1\leq t\leq T-1}\left\| Z_t(m)'\left(Z'(m)Z(m)\right)^{-1}Z_t(m)\right\|\right| = o_P(1). \tag{11}$$

This implies that

$$\max_{1\leq t\leq T-1}\left\|\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right\| = o_P(1), \tag{12}$$

as $\max_{1\leq t\leq T-1}\left\|Z_t(m)'\left(Z'(m)Z(m)\right)^{-1}Z_t(m)\right\| = o_P(1)$ given Assumption 4 (b).

The remaining part of the proof goes similarly as the proof of Shao (1993, Equation 3.4). Noting that, $CV_1(m) = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(1-\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right)^{-2}\hat{\varepsilon}_{t+1}^2$ (see, Shao (1993)), we rely on Taylor expansion to have that for any m, $\left(1-\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right)^{-2}$ is equal to

$$1+2\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)+O_P\left[\left(\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right)^2\right]. \tag{13}$$

Hence $CV_1(m) = A_4(m) + 2A_5(m) + o_P(A_5(m))$, where

$$A_4(m) = \frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^2(m) \text{ and } A_5(m) = \frac{1}{T-1}\sum_{t=1}^{T-1}\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\hat{\varepsilon}_{t+1}^2(m),$$

since $\max_{1\le t\le T-1}\left\|\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right\| = o_P(1)$. We next study $A_4(m)$ and $A_5(m)$.

Given the decomposition $\hat{\varepsilon}(m) = \varepsilon + \mu - \tilde{\mu}(m)$ where $\mu = F^0\alpha + W\beta$ and $\tilde{\mu}(m) = \tilde{P}(m)\mu + \tilde{P}(m)\varepsilon$, we have $A_4(m) = \frac{1}{T-1}\varepsilon'\varepsilon + \tilde{L}_T(m) - 2r_{1T}(m)$, with

$$\tilde{L}_T(m) = \frac{1}{T-1}\mu'\left(I - \tilde{P}(m)\right)\mu + \frac{1}{T-1}\varepsilon'\tilde{P}(m)\varepsilon. \tag{14}$$

and

$$r_{1T}(m) = \frac{1}{T-1}(\tilde{\mu}(m) - \mu)'\varepsilon = \frac{1}{T}\left[\tilde{P}(m)\varepsilon - \left(I - \tilde{P}(m)\right)\mu\right]'\varepsilon = \frac{1}{T-1}\varepsilon'\tilde{P}(m)\varepsilon - \frac{1}{T-1}\mu'\left(I - \tilde{P}(m)\right)\varepsilon. \tag{15}$$

From the definition of $A_4(m)$, $\tilde{L}_T(m)$ and $r_{1T}(m)$, we find

$$A_4(m) = \frac{1}{T-1}\varepsilon'\varepsilon - \frac{1}{T-1}\varepsilon'\tilde{P}(m)\varepsilon + \frac{1}{T-1}\mu'\left(I - \tilde{P}(m)\right)\mu + 2\frac{1}{T-1}\mu'\left(I - \tilde{P}(m)\right)\varepsilon. \tag{16}$$

Under our Assumptions 1–3, for any m,

$$\frac{1}{T-1}\varepsilon'\tilde{P}(m)\varepsilon = \frac{1}{T-1}\varepsilon'P(m)\varepsilon + O_P\left(\frac{1}{C_{NT}^2}\right),$$

$$\frac{1}{T-1}\mu'\left(I - \tilde{P}(m)\right)\mu = \mu'(I - P(m))\mu + O_P\left(\frac{1}{C_{NT}^2}\right),$$

and

$$\frac{1}{T-1}\mu'\left(I - \tilde{P}(m)\right)\varepsilon = \frac{1}{T-1}\mu'(I - P(m))\varepsilon + O_P\left(\frac{1}{C_{NT}^2}\right)$$

given Lemma 7.1. Hence, it follows that

$$A_4(m) = \frac{1}{T-1}\varepsilon'\varepsilon - \frac{1}{T-1}\varepsilon'P(m)\varepsilon + \frac{1}{T-1}\mu'(I - P(m))\mu + 2\frac{1}{T-1}\mu'(I - P(m))\varepsilon + O_P\left(\frac{1}{C_{NT}^2}\right).$$

To complete the study of $A_4(m)$ and $A_5(m)$, we now consider the case where $m \in \mathcal{M}_1$ and the case where $m \in \mathcal{M}_2$. Let start with the first case. For any $m \in \mathcal{M}_1$,

$$A_4(m) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{1}{T-1}\mu'(I - P(m))\mu + o_P(1) \tag{17}$$

since $\frac{1}{T-1}\varepsilon'P(m)\varepsilon = o_P(1)$ and $\frac{1}{T-1}\mu'(I - P(m))\varepsilon = o_P(1)$ (see, Groen and Kapetanios (2013, Proof of Theorem 1)). Moreover, we have

$$|A_5(m)| \le \max_{1\le t\le T-1}\left\{\left|\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right|\right\}\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^2(m), \tag{18}$$

implying $A_5(m) = o_P(1)$, as the first term in the right hand side of (18) is $o_P(1)$ given Assumption 4

(b) and the second term is equal to $A_4(m)$, which is $O_P(1)$. Hence, for $m \in \mathcal{M}_1$,

$$CV_1(m) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{1}{T-1}\mu'(I - P(m))\mu + o_P(1) = \sigma^2 + \frac{1}{T-1}\mu'(I - P(m))\mu + o_P(1).$$

We now turn our attention to the second case. Because $\mu'(I - P(m))\mu = 0$, $\mu'(I - P(m))\varepsilon = 0$ for $m \in \mathcal{M}_2$,

$$A_4(m) = \frac{1}{T-1}\varepsilon'\varepsilon - \frac{1}{T-1}\varepsilon' P(m)\varepsilon + O_P\left(\frac{1}{C_{NT}^2}\right). \tag{19}$$

Further, $A_5(m) = \frac{(r(m)+q)}{T-1}\sigma^2 + o_P\left(\frac{1}{T-1}\right)$ for $m \in \mathcal{M}_2$. Indeed, as

$$A_5(m) = \frac{1}{T-1}\text{Trace}\left[\left(\frac{1}{T-1}\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{Z}_t(m)\hat{Z}_t'(m)\hat{\varepsilon}_{t+1}^2(m)\right]$$

and $\frac{1}{T-1}\hat{Z}'(m)\hat{Z}'(m) = \Sigma_Z(m) + o_P(1)$, it holds that

$$A_5(m) = \frac{1}{T-1}\text{Trace}\left[\left(\Sigma_Z(m)^{-1} + o_P(1)\right)\left(\sigma^2\Sigma_Z(m) + o_P(1)\right)\right] = \frac{(r(m)+q)}{T-1}\sigma^2 + o_P\left(\frac{1}{T-1}\right).$$

In consequence, for $m \in \mathcal{M}_2$

$$\begin{aligned} CV_1(m) &= \frac{1}{T-1}\varepsilon'\varepsilon + 2\frac{(r(m)+q)}{T-1}\sigma^2 - \frac{1}{T-1}\varepsilon' P(m)\varepsilon + O_P\left(\frac{1}{C_{NT}^2}\right) + o_P\left(\frac{1}{C_{NT}^2}\right) \\ &= \frac{1}{T-1}\varepsilon'\varepsilon + 2\frac{(r(m)+q)}{T-1}\sigma^2 - \frac{1}{T-1}\varepsilon' P(m)\varepsilon + O_P\left(\frac{1}{C_{NT}^2}\right). \end{aligned}$$

Because, in the usual case where the factors are observed, $CV_1(m) = \frac{1}{T-1}\varepsilon'\varepsilon + 2\frac{(r(m)+q)}{T-1}\sigma^2 - \frac{1}{T-1}\varepsilon' P(m)\varepsilon + o_P\left(\frac{1}{T}\right)$ for $m \in \mathcal{M}_2$ (see, Shao (1993)). In consequence, we denote $V_T(m) = CV_1(m) - \left(\frac{1}{T-1}\varepsilon'\varepsilon + 2\frac{(r(m)+q)}{T-1}\sigma^2 - \frac{1}{T-1}\varepsilon' P(m)\varepsilon\right) = O_P\left(\frac{1}{C_{NT}^2}\right)$ the additional terms due the factor estimation when $m \in \mathcal{M}_2$. □

Proof of Theorem 1. We have the following decomposition

$$\begin{aligned} CV_d(m) &= \frac{1}{d\times b}\sum_{s\in\mathcal{R}}\left\|(y_s - P_{s^c}(m)y_{s^c}) + \left(P_{s^c}(m) - \tilde{P}_{s^c}(m)\right)y_{s^c}\right\|^2 \\ &\equiv B_1(m) + B_2(m) + B_3(m) \end{aligned}$$

where

$$\begin{aligned} B_1(m) &= \frac{1}{d\times b}\sum_{s\in\mathcal{R}}\left\|\left(P_{s^c}(m) - \tilde{P}_{s^c}(m)\right)y_{s^c}\right\|^2, \\ B_2(m) &= 2\frac{1}{d\times b}\sum_{s\in\mathcal{R}}(y_s - P_{s^c}(m)y_{s^c})'\left(P_{s^c}(m) - \tilde{P}_{s^c}(m)\right)y_{s^c}, \\ B_3(m) &= \frac{1}{d\times b}\sum_{s\in\mathcal{R}}\|(y_s - P_{s^c}(m)y_{s^c})\|^2, \end{aligned}$$

with $P_{s^c}(m) = Z_s(m)\left(Z_{s^c}(m)'Z_{s^c}(m)\right)^{-1}Z_{s^c}(m)'$ and $\tilde{P}_{s^c}(m) = \hat{Z}_s(m)\left(\hat{Z}_{s^c}(m)'\hat{Z}_{s^c}(m)\right)^{-1}\hat{Z}_{s^c}(m)'$.

The proofs will be done into two parts. The first shows that $B_1(m) = o_P\left(\frac{1}{\kappa}\right)$ and $B_2(m) = o_P\left(\frac{1}{\kappa}\right)$, while the second studies $B_3(m)$ and concludes.

Part 1: Using a decomposition of $\left(P_{s^c}(m) - \tilde{P}_{s^c}(m)\right) y_{s^c}$ and the c-r inequality, we obtain that $B_1(m)$ is lower than

$$\underbrace{4\frac{1}{d\times b}\sum_{s\in\mathcal{R}}\|B_{11s}(m)\|^2}_{B_{11}(m)} + \underbrace{4\frac{1}{d\times b}\sum_{s\in\mathcal{R}}\|B_{12s}(m)\|^2}_{B_{12}(m)} + \underbrace{4\frac{1}{d\times b}\sum_{s\in\mathcal{R}}\|B_{13s}(m)\|^2}_{B_{13}(m)} + \underbrace{4\frac{1}{d\times b}\sum_{s\in\mathcal{R}}\|B_{14s}(m)\|^2}_{B_{14}(m)}.$$

with

$$B_{11s}(m) = \left(\hat{Z}_s(m)\left[\left(Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1} - \left(\hat{Z}'_{s^c}(m)\hat{Z}_{s^c}(m)\right)^{-1}\right]\hat{Z}'_{s^c}(m)\right) y_{s^c},$$

$$B_{12s}(m) = \left(\left(Z_s(m) - \hat{Z}_s(m)\right)\left[\left(Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1}\right]\left(\hat{Z}_{s^c}(m) - Z_{s^c}(m)\right)\right) y_{s^c},$$

$$B_{13s}(m) = \left[Z_s(m)\left(Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1}\left(Z_{s^c}(m) - \hat{Z}_{s^c}(m)\right)\right] y_{s^c}$$

and

$$B_{14s}(m) = \left[\left(Z_s(m) - \hat{Z}_s(m)\right)\left(Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1} Z_{s^c}(m)\right] y_{s^c}.$$

Starting with $B_{11}(m)$, we have that for any m,

$$B_{11}(m) \le \frac{1}{d\times b}\sum_{s\in\mathcal{R}}\left\|\hat{Z}_s(m)\right\|^2\left\|\left(Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1} - \left(\hat{Z}'_{s^c}(m)\hat{Z}_{s^c}(m)\right)^{-1}\right\|^2\left\|\hat{Z}'_{s^c}(m) y_{s^c}\right\|^2.$$

Using the fact that $\left\|\hat{Z}_s(m)\right\| \le \left\|\hat{Z}(m)\right\|$ and the Cauchy-Schwarz inequality, it follows that

$$B_{11}(m) \le \frac{1}{d}\left\|\hat{Z}(m)\right\|^2\left[\frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\left(\frac{1}{\kappa}Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1} - \left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\hat{Z}_{s^c}(m)\right)^{-1}\right\|^4\frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\frac{1}{\kappa}\hat{Z}'_{s^c}(m) y_{s^c}\right\|^4\right]^{1/2}.$$

Because $\frac{1}{d}\left\|\hat{Z}(m)\right\|^2 = O_P(1)$, to find the order of $B_{11}(m)$, we next show that

$$B_{111}(m) = \frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\left(\frac{1}{\kappa}Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1} - \left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\hat{Z}_{s^c}(m)\right)^{-1}\right\|^4 = o_P\left(\frac{1}{\kappa^2}\right)$$

and

$$B_{112}(m) = \frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\frac{1}{\kappa}\hat{Z}'_{s^c}(m) y_{s^c}\right\|^4 = O_P(1).$$

To bound $B_{111}(m)$, we first write that $\left(\frac{1}{\kappa}Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1} - \left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\hat{Z}_{s^c}(m)\right)^{-1}$ is equal to $B_{1111s}(m) + B_{1112s}(m)$ where

$$B_{1111s}(m) = \left(\frac{1}{\kappa}Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1}\left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\left(\hat{Z}_{s^c}(m) - Z_{s^c}(m)\right)\right)\left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\hat{Z}_{s^c}(m)\right)^{-1}$$

and

$$B_{1112s}(m) = \left(\frac{1}{\kappa}Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1}\left(\frac{1}{\kappa}\left(\hat{Z}_{s^c}(m) - Z_{s^c}(m)\right)' Z_{s^c}(m)\right)\left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\hat{Z}_{s^c}(m)\right)^{-1}.$$

Hence by the c-r inequality, $\|B_{111}(m)\|$ is bounded by $2^3\left(\frac{1}{b}\sum_{s\in\mathcal{R}}\|B_{1111s}(m)\|^4+\frac{1}{b}\sum_{s\in\mathcal{R}}\|B_{1112s}(m)\|^4\right)$. In particular,

$$\frac{1}{b}\sum_{s\in\mathcal{R}}\|B_{1111s}(m)\|^4\leq 2^3\left\|(\Sigma_Z(m))^{-1}\right\|^8\frac{1}{b\cdot\kappa^4}\sum_{s\in\mathcal{R}}\left\|\hat{Z}'_{s^c}(m)\left(\hat{Z}_{s^c}(m)-Z_{s^c}(m)\right)\right\|^4(1+o_P(1))$$

and

$$\frac{1}{b}\sum_{s\in\mathcal{R}}\|B_{1112s}(m)\|^4\leq 2^3\left\|(\Sigma_Z(m))^{-1}\right\|^8\frac{1}{b\cdot\kappa^4}\sum_{s\in\mathcal{R}}\left\|\left(\hat{Z}_{s^c}(m)-Z_{s^c}(m)\right)'Z_{s^c}(m)\right\|^4(1+o_P(1))$$

since $\left(\frac{1}{\kappa}Z'_{s^c}(m)Z_{s^c}(m)\right)^{-1}=(\Sigma_Z(m))^{-1}+o_P(1)$ given Assumption 4 (c) and (f). Combining the arguments in Lemma 7.2 and Gonçalves and Perron (2014, Lemma A2 (b)-(c)), we can show that

$$\frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\frac{1}{\kappa}\left(\hat{Z}_{s^c}(m)-Z_{s^c}(m)\right)'Z_{s^c}(m)\right\|^4=o_P\left(\frac{1}{\kappa^2}\right)$$

and

$$\frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\left(\hat{Z}_{s^c}(m)-Z_{s^c}(m)\right)\right\|^4=o_P\left(\frac{1}{\kappa^2}\right).$$

Thus, $B_{111}(m)=o_P\left(\frac{1}{\kappa^2}\right)$. Further, by Cauchy-Scharwz inequality and Jensen inequality, we have

$$B_{112}(m)=\frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\,y_{s^c}\right\|^4\leq\left(\frac{1}{b\kappa}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\hat{Z}_t(m)\right\|^8\right)^{1/2}\left(\frac{1}{b\kappa}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\|y_t\|^8\right)^{1/2}=O_P(1),\tag{20}$$

since $\frac{1}{b\kappa}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\|y_t\|^8=O_P(1)$ from Assumption 3 and $\frac{1}{b\kappa}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\hat{Z}_t(m)\right\|^8=O_P(1)$. Using $\frac{1}{d}\left\|\hat{Z}(m)\right\|^2=O_p(1)$, $B_{111}(m)=o_P\left(\frac{1}{\kappa^2}\right)$ and $B_{112}(m)=O_P(1)$, we find for any m that

$$B_{11}(m)=o_P\left(\frac{1}{\kappa}\right).\tag{21}$$

We now look at $B_{12}(m)$. Since $\left\|Z_s(m)-\hat{Z}_s(m)\right\|\leq\left\|Z(m)-\hat{Z}(m)\right\|$ and $\left(\frac{1}{T_c}Z'_{s^c}(m)Z_{s^c}(m)\right)^{-1}=(\Sigma_Z(m))^{-1}+o_p(1)$, it follows that

$$B_{12}(m)\leq\frac{1}{d}\left\|Z(m)-\hat{Z}(m)\right\|^2\left\|(\Sigma_Z(m))^{-1}\right\|^2(1+o_p(1))\frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\frac{1}{\kappa}\left(\hat{Z}_{s^c}(m)-Z_{s^c}(m)\right)'y_{s^c}\right\|^2.$$

As $\frac{1}{b}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\hat{Z}_t(m)-Z_t(m)\right\|^4=o_P(1)$ from Lemma 7.2, we deduce

$$\frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\left(\hat{Z}_{s^c}(m)-Z_{s^c}(m)\right)'y_{s^c}\right\|^2\leq\left(\frac{1}{\kappa\cdot b}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\hat{Z}_t(m)-Z_t(m)\right\|^4\right)^{1/2}\left(\frac{1}{\kappa\cdot b}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\|y_{t+1}\|^4\right)^{1/2}$$

is $o_P\left(\frac{1}{\kappa^{1/2}}\right)$. Hence, using $\frac{1}{d}\left\|Z(m)-\hat{Z}(m)\right\|^2=O_p\left(\frac{1}{C_{NT}^2}\right)$, for any m, we obtain

$$B_{12}(m)=O_p\left(\frac{1}{C_{NT}^2\kappa^{1/2}}\right).\tag{22}$$

For $B_{13}(m)$, we have for any m, the bound

$$B_{13}(m) \leq \frac{1}{d} \sum_{s \in \mathcal{R}} \|Z_s(m)\|^2 \left\| \left(\frac{1}{\kappa} Z'_{s^c}(m) Z_{s^c}(m) \right)^{-1} \right\|^2 \frac{1}{b} \left\| \frac{1}{\kappa} \left(Z_{s^c}(m) - \hat{Z}_{s^c}(m) \right)' y_{s^c} \right\|^2$$

Since for any m, $\frac{1}{d}\|Z_s(m)\|^2 \leq \frac{1}{d}\|Z(m)\|^2 = O_P(1)$, $\left(\frac{1}{\kappa} Z'_{s^c}(m) Z_{s^c}(m)\right)^{-1} = \Sigma_Z(m) + o_P(1)$ and $\frac{1}{b}\sum_{s \in \mathcal{R}} \left\| \frac{1}{\kappa} \left(Z_{s^c}(m) - \hat{Z}_{s^c}(m) \right)' y_{s^c} \right\|^2 = o_P\left(\frac{1}{\kappa}\right)$ (using the same arguments as in Lemma 7.2 and Gonçalves and Perron (2014, Lemma A2 (c)), it follows that

$$B_{13}(m) = o_P\left(\frac{1}{\kappa}\right). \tag{23}$$

To finish, we have that

$$B_{14}(m) = O_p\left(\frac{1}{C_{NT}^2}\right), \tag{24}$$

using the bound

$$B_{14}(m) \leq \frac{1}{d} \left\| Z(m) - \hat{Z}(m) \right\|^2 \left\| (\Sigma_Z(m))^{-1} \right\|^2 (1 + o_p(1)) \frac{1}{d \cdot b} \sum_{s \in \mathcal{R}} \left\| \frac{1}{\kappa} Z_{s^c}(m)' y_{s^c} \right\|^2,$$

where $\frac{1}{d}\left\| Z(m) - \hat{Z}(m) \right\|^2 = O_P\left(\frac{1}{C_{NT}^2}\right)$ and

$$\frac{1}{b} \sum_{s \in \mathcal{R}} \left\| \frac{1}{\kappa} Z_{s^c}(m)' y_{s^c} \right\|^2 \leq \left(\frac{1}{b \cdot \kappa} \sum_{s \in \mathcal{R}} \sum_{t \in s^c} \|Z_t(m)\|^4 \right)^{1/2} \left(\frac{1}{b \cdot \kappa} \sum_{s \in \mathcal{R}} \sum_{t \in s^c} \|y_{t+1}\|^4 \right)^{1/2} = O_P(1).$$

Finally, from Equations (21), (22), (23) and (24), we conclude that $B_1(m) = o_P\left(\frac{1}{\kappa}\right)$, for any m. By similar arguments, we can also prove that $B_2(m) = o_P\left(\frac{1}{\kappa}\right)$.

Part 2:

We first show in this part that

$$B_3(m) = \frac{1}{d \cdot b} \sum_{s \in \mathcal{R}} \|(y_s - P_{s^c}(m) y_{s^c})\|^2 = \frac{1}{d \cdot b} \sum_{s \in \mathcal{R}} \left\| (I_d - Q_s(m))^{-1} \left(y_s - Z_s(m) \tilde{\delta}(m) \right) \right\|^2$$

with $Q_s(m) = Z_s(m)(Z'(m) Z(m))^{-1} Z'_s(m)$ and $\tilde{\delta}(m)$, the OLS estimator by regressing y_s on $Z_s(m)$. We use the following identity

$$y_s - P_{s^c}(m) y_{s^c} = (I_d - Q_s(m))^{-1} [y_s - Q_s(m) y_s - (I_d - Q_s(m)) P_{s^c}(m) y_{s^c}].$$

Because

$$
\begin{aligned}
& (I_d - Q_s(m)) P_{s^c}(m) \\
= \;& P_{s^c}(m) - Z_s(m)(Z'(m) Z(m))^{-1} Z_s'(m) Z_s(m) \left(Z_{s^c}(m)' Z_{s^c}(m)\right)^{-1} Z_{s^c}(m)' \\
= \;& P_{s^c}(m) - Z_s(m)(Z'(m) Z(m))^{-1} \times \left[Z'(m) Z(m) - Z_{s^c}(m)' Z_{s^c}(m)\right] \left(Z_{s^c}(m)' Z_{s^c}(m)\right)^{-1} Z_{s^c}(m)' \\
= \;& P_{s^c}(m) - P_{s^c}(m) - Z_s(m)(Z'(m) Z(m))^{-1} Z_{s^c}(m)' = -Z_s(m)(Z'(m) Z(m))^{-1} Z_{s^c}(m)',
\end{aligned}
$$

it follows that $(I_d - Q_s(m))(y_s - P_{s^c}(m) y_{s^c})$ is equal to

$$
\begin{aligned}
& y_s - Z_s(m)(Z'(m) Z(m))^{-1} Z_s(m) y_s + Z_s(m)(Z'(m) Z(m))^{-1} Z_{s^c}(m)' y_{s^c} \\
= \;& y_s - Z_s(m)(Z'(m) Z(m))^{-1} Z(m) y \\
= \;& y_s - Z_s(m) \tilde{\delta}(m).
\end{aligned}
$$

Thus $y_s - \tilde{y}_s(m) = y_s - P_{s^c}(m) y_{s^c} = (I_d - Q_s(m))^{-1} \left(y_s - Z_s(m) \tilde{\delta}(m)\right)$ and

$$
B_3(m) = \frac{1}{d \cdot b} \sum_{s \in \mathcal{R}} \left\| (I_d - Q_s(m))^{-1} \left(y_s - Z_s(m) \tilde{\delta}(m)\right) \right\|^2.
$$

Because $Z_s(m)$ can be treated as non generated regressors and $\tilde{\delta}(m)$ the associated estimator, we next apply Shao (1993, Theorem 2). Hence for $m \in \mathcal{M}_1$,

$$
B_3(m) = \frac{1}{d \cdot b} \sum_{s \in \mathcal{R}} \|\varepsilon_s\|^2 + \frac{1}{T-1} \delta' Z^{0\prime} \left(I - Z(m)(Z'(m) Z(m))^{-1} Z'(m)\right) Z^0 \delta + o_P(1) + R_T(m), \tag{25}
$$

where $R_T(m) \geq 0$ and $m \in \mathcal{M}_2$,

$$
B_3(m) = \frac{1}{d \cdot b} \sum_{s \in \mathcal{R}} \|\varepsilon_s\|^2 + \frac{r(m) + q}{\kappa} \sigma^2 + o_P\left(\frac{1}{\kappa}\right). \tag{26}
$$

Finally, using $B_1(m) = o_P\left(\frac{1}{\kappa}\right)$ and $B_2(m) = o_P\left(\frac{1}{\kappa}\right)$, we deduce that

$$
CV_d(m) = B_3(m) + o_P\left(\frac{1}{\kappa}\right).
$$

Furthermore, the result follows from Shao (1993, Theorem 2). □

Proof of Theorem 2. The proof follows similarly as the one of Djogbenou, Gonçalves, and Perron (2015, Theorem 3.1) by showing that the high level conditions of Gonçalves and Perron (2014) are satisfied. We use the identity

$$
\sqrt{\kappa}\left(\hat{\delta}_d^*(m) - \Phi_0^{*-1}(m) \hat{\delta}(m)\right) = \left(\frac{1}{T-1} \hat{Z}^{*\prime}(m) \hat{Z}^*(m)\right)^{-1} \left[A^*(m) + B^*(m) - C^*(m)\right] \tag{27}
$$

where $A^*(m) = \Phi_0^*(m) \sqrt{\kappa} \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{Z}_t(m) \varepsilon_{t+1}^*$, $B^*(m) = \sqrt{\kappa} \frac{1}{T-1} \sum_{t=1}^{T-1} \left(\tilde{F}_t^*(m) - H_0^*(m) \tilde{F}_t(m)\right) \varepsilon_{t+1}^*$ and $C^*(m) = \sqrt{\kappa} \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{Z}_t^*(m) \left(\tilde{F}_t^*(m) - H_0^*(m) \tilde{F}_t(m)\right)' (H_0^*(m))^{-1\prime} \hat{\alpha}(m)$ where $\underset{N,T \to \infty}{\text{plim}} \Phi^*(m) =$

$\Phi_0^*(m)$ and $\underset{N,T\to\infty}{\text{plim}}\, H^*(m) = H_0^*(m)$ are diagonal. Note that in the bootstrap world Φ_0^* is diagonal (see Gonçalves and Perron (2014)) and $H_0^*(m)$ is an $r(m)\times r(m)$ squared submatrix of H_0^*. Note also that given this fact, we treat $\tilde{F}_t^*(m)$ as estimating $H_0^*(m)\tilde{F}_t(m)$. Hence, we can use the properties of H_0^* as a diagonal and nonsingular matrix in order to identify the rotation matrix associated with subvectors $\tilde{F}_t^*(m)$ of $\tilde{F}_t^*$. We will establish the result in three steps proving that $A^*(m)$ converges in distribution whereas $B^*(m)$ and $C^*(m)$ converge in probability to zero.

Part 1: One can write that

$$B^*(m) = \frac{\sqrt{\kappa}}{T-1}\sum_{t=1}^{T-1}\left(\tilde{F}_t^*(m) - H_0^*(m)\tilde{F}_t(m)\right)\varepsilon_{t+1}^* = \frac{1}{\sqrt{T-1}}\sum_{t=1}^{T-1}\left(\tilde{F}_t^*(m) - H_0^*(m)\tilde{F}_t(m)\right)\xi_{t+1}^* \tag{28}$$

with $\xi_{t+1}^* = \frac{1}{\sqrt{1-\frac{r+q}{T-1}}}\left(\hat{\varepsilon}_{t+1} - \overline{\hat{\varepsilon}}\right)$. Given Gonçalves and Perron (2014, Lemma B2), $B^*(m) = O_P\left(\frac{1}{C_{NT}}\right)$ as long as $B^* = \frac{1}{\sqrt{T-1}}\sum_{t=1}^{T-1}\left(\tilde{F}_t^* - H_0^*\tilde{F}_t\right)\xi_{t+1}^* = O_P\left(\frac{1}{C_{NT}}\right)$ if their Conditions A^*-D^* are verified with ξ_{t+1}^* replacing ε_{t+1}^*. Indeed, A^* and B^* are satisfied since e_{it}^* relies on the wild bootstrap and we only need to check Conditions C^* and D^*. Starting with Condition $C^*(a)$, since e_{it}^* and ε_{s+1}^* are independent and e_{it}^* is independent across (i,t),

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}^*\left|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T-1}\sum_{i=1}^{N}\xi_{s+1}^*\left(e_{it}^*e_{is}^* - \mathrm{E}\left(e_{it}^*e_{is}^*\right)\right)\right|^2 &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\sum_{s=1}^{T-1}\mathrm{E}^*\left(\xi_{s+1}^{*2}\right)\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\tilde{e}_{is}^2\mathrm{Var}^*\left(\eta_{it}\eta_{is}\right)\\
&\leq C\left(\frac{1}{T-1-r-q}\sum_{l=1}^{T-1}\hat{\varepsilon}_{l+1}^2\right)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{e}_{it}^4,
\end{aligned}$$

where the first equality uses the fact that $\mathrm{Cov}^*\left(e_{it}^*e_{is}^*, e_{jt}^*e_{jl}^*\right) = 0$ for $i \neq j$ or $s \neq l$, and the inequality uses the fact that $\mathrm{E}^*\left(\varepsilon_{s+h}^{*2}\right) = \frac{1}{1-\frac{r+q}{T-1}}\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^2 - \overline{\hat{\varepsilon}}^2\right) \leq \frac{1}{T-1-r-q}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^2$.and that $\mathrm{Var}^*(\eta_{it}\eta_{is})$ is bounded under the assumptions of Theorem 2. Since $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{e}_{it}^4 = O_P(1)$ and $\frac{1}{T-1-r-q}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^2 = O_P(1)$ under Assumptions 1–3 (see, Gonçalves and Perron (2014)), the result follows. We now verify Condition $C^*(b)$. We have

$$\begin{aligned}
\mathrm{E}^*\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T-1}\sum_{i=1}^{N}\tilde{\lambda}_i e_{it}^*\xi_{t+h}^*\right\|^2 &= \frac{1}{TN}\left[\sum_{t=1}^{T-1}\mathrm{E}^*\left(\xi_{t+1}^{*2}\right)\left(\sum_{i=1}^{N}\tilde{\lambda}_i'\tilde{\lambda}_i\mathrm{E}^*\left(e_{it}^{*2}\right)\right)\right]\\
&\leq \left(\frac{1}{T-1-r-q}\sum_{s=1}^{T-1}\hat{\varepsilon}_{s+1}^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i\right\|^4\right)^{1/2}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\tilde{e}_{it}^4\right)^{1/2}
\end{aligned}$$

where the first equality uses the fact that $\mathrm{E}^*\left(e_{it}^*e_{js}^*\right) = 0$ whenever $i \neq j$ or $t \neq s$, and the third equality the fact that $\mathrm{E}^*\left(\varepsilon_{t+1}^{*2}\right) \leq \frac{1}{T-1-r-q}\sum_{s=1}^{T-1}\hat{\varepsilon}_{s+1}^2$ and $\left(\frac{1}{T}\sum_{t=1}^{T-1}\tilde{e}_{it}^2\right)^2 \leq \frac{1}{T}\sum_{t=1}^{T-1}\tilde{e}_{it}^4$. The result, that $C^*(b)$ holds, follows since each term of the last inequality is $O_P(1)$ (see, Gonçalves and Perron (2014)). To prove Condition $C^*(c)$, We follow closely the proof in Gonçalves and Perron (2014) and it will be sufficient to show that $\frac{1}{T-1}\sum_{t=1}^{T-1}\xi_{t+1}^{*4} = O_{p^*}(1)$ in probability. Using the definition of

$\mathrm{E}^*\left(\xi_{t+1}^{*4}\right)$ and the c-r inequality,

$$\begin{aligned}\mathrm{E}^*\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\xi_{t+1}^{*4}\right) &= \mathrm{E}^*\left(\xi_{t+1}^{*4}\right) = \frac{T-1}{(T-1-r-q)^2}\sum_{s=1}^{T-1}\left(\hat{\varepsilon}_{s+1}-\bar{\hat{\varepsilon}}\right)^4 \\ &\leq 2^3\frac{(T-1)^2}{(T-1-r-q)^2}\frac{1}{T-1}\sum_{s=1}^{T-1}\hat{\varepsilon}_{s+1}^4 + 2^3\frac{(T-1)^2}{(T-1-r-q)^2}\bar{\hat{\varepsilon}}^4.\end{aligned}$$

Because $\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^4 = O_P(1)$ and $\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1} = O_P(1)$ under Assumptions 1–3, $\mathrm{E}^*\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\varepsilon_{t+1}^{*4}\right) = O_P(1)$ and $C^*(c)$ follows. For Condition $D^*(a)$, we have $\mathrm{E}^*\left(\xi_{t+1}^*\right) = \frac{T-1}{T-1-r-q}\frac{1}{T-1}\sum_{s=1}^{T-1}\left(\hat{\varepsilon}_{s+1}-\bar{\hat{\varepsilon}}\right) = 0$ and

$$\frac{1}{T}\sum_{t=1}^{T-1}\mathrm{E}^*\left(\xi_{t+1}^{*2}\right) = \frac{T-1}{T}\mathrm{E}^*\left(\xi_{t+1}^{*2}\right) \leq \frac{T-1}{T}\frac{1}{T-1-r-q}\sum_{s=1}^{T-1}\hat{\varepsilon}_{s+1}^2 = O_P(1).$$

To finish, we also verify Condition $D^*(b)$. To show that condition, which is $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\hat{Z}_t\xi_{t+1}^* \xrightarrow{d^*} N(0,\ \Omega^*)$, we rely on Lyapunov Theorem by proving that the required conditions are satisfied. Noting $\Psi_t^* \equiv \Omega^{*-\frac{1}{2}}\hat{Z}_t\xi_{t+1}^*$ and $\Omega^* = \operatorname*{plim}_{N,T\to\infty}\mathrm{E}^*\left(\frac{1}{T}\sum_{t=1}^{T-1}\hat{Z}_t\hat{Z}_t'\xi_{t+1}^{*2}\right)$, we can write $\Omega^{*-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\hat{Z}_t\xi_{t+1}^* = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\Psi_t^*$, where Ψ_t^* are conditionally independent for $t = 1,\ldots,T-1$, with $\mathrm{E}^*\left(\Psi_t^*\right) = \Omega^{*-\frac{1}{2}}\hat{Z}_t\mathrm{E}^*\left(\xi_{t+1}^*\right) = 0$ and

$$\operatorname*{plim}_{N,T\to\infty}\mathrm{Var}^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Psi_t^*\right) = \Omega^{*-\frac{1}{2}}\left(\operatorname*{plim}_{N,T\to\infty}\mathrm{E}^*\left(\frac{1}{T}\sum_{t=1}^{T-1}\hat{Z}_t\hat{Z}_t'\xi_{t+1}^{*2}\right)\right)\Omega^{*-\frac{1}{2}} = I_{q+r}.$$

It only remains to show that for some $1 < s < 2$, $\Upsilon_T \equiv \frac{1}{T^d}\sum_{t=1}^{T-1}\mathrm{E}^*\left\|\Psi_t^*\right\|^{2s} = o_P(1)$. Using the bound

$$\Upsilon_T = \frac{1}{T^s}\sum_{t=1}^{T-1}\mathrm{E}^*\left(\left\|\Omega^{*-\frac{1}{2}}\hat{Z}_t\xi_{t+1}^*\right\|^2\right)^s \leq \left\|\Omega^{*-\frac{1}{2}}\right\|^{2s}\frac{1}{T^{s-1}}\frac{1}{T}\sum_{t=1}^{T-1}\left\|\hat{Z}_t\right\|^{2s}\mathrm{E}^*\left\|\xi_{t+1}^*\right\|^{2s}$$

and the fact that $\mathrm{E}^*\left\|\xi_{t+1}^*\right\|^{2s} = \frac{(T-1)^{s-1}}{(T-1-r-q)^s}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1}-\bar{\hat{\varepsilon}}\right)^{2s}$, we obtain

$$\Upsilon_T \leq \left\|\Omega^{*-\frac{1}{2}}\right\|^{2s}\left(\frac{1}{(T-1-r-q)^s}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1}-\bar{\hat{\varepsilon}}\right)^{2s}\right)\frac{1}{T}\sum_{t=1}^{T-1}\left\|\hat{Z}_t\right\|^{2s}. \tag{29}$$

To find the order in probability of Υ_T, we note $\frac{1}{T}\sum_{t=1}^{T-1}\left\|\hat{Z}_t\right\|^{2s} \leq \left(\frac{1}{T}\sum_{t=1}^{T-1}\left\|\hat{Z}_t\right\|^4\right)^{s/2} = O_P(1)$, as $\frac{1}{T}\sum_{t=1}^{T-1}\left\|\hat{Z}_t\right\|^4$ is $O_P(1)$. In addition, by an application of a $c-r$ inequality, we have

$$\frac{1}{(T-1-r-q)^s}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1}-\bar{\hat{\varepsilon}}\right)^{2s} \leq 2^{2s-1}\frac{T-1}{(T-1-r-q)^s}\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^{2s}+\bar{\hat{\varepsilon}}^{2s}\right) = O_P\left(\frac{1}{T^{s-1}}\right)$$

where $\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^{2s} \leq \left(\frac{1}{T}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^4\right)^{s/2} = O_P(1)$ and $\bar{\hat{\varepsilon}} = O_P(1)$. Hence, we deduce from (29) that $\boldsymbol{\Upsilon}_T = O_P\left(\frac{1}{T^{s-1}}\right) = o_P(1)$ since $s > 1$. Thus $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\hat{Z}_t\xi_{t+1}^* \xrightarrow{d^*} \mathrm{N}(0,\ \Omega^*)$.

Part 2: By Gonçalves and Perron (2014, Lemma B4),

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\hat{Z}^{*}(m)\left(\tilde{F}_{t}^{*}(m)-H_{0}^{*}(m)\tilde{F}_{t}(m)\right)'\left(H_{0}^{*}(m)\right)^{-1\prime}\hat{\alpha}(m)=O_{P}\left(\frac{\sqrt{T}}{N}\right)$$

for any m, as it does not involve the residual bootstrap for the time series dimension. Hence, we have any m that

$$C^{*}(m)=\sqrt{\kappa}\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{Z}^{*}(m)\left(\tilde{F}_{t}^{*}(m)-H_{0}^{*}(m)\tilde{F}_{t}(m)\right)'\left(H_{0}^{*}(m)\right)^{-1\prime}\hat{\alpha}(m)=O_{P}\left(\frac{\sqrt{\kappa}}{N}\frac{T}{T-1}\right)=o_{P}(1).$$

Part 3: By similar steps to condition $D^{*}(b)$, $\Omega^{*}(m)^{-\frac{1}{2}}A^{*}(m)\xrightarrow{d^{*}}\mathrm{N}\left(0,\ \mathrm{I}_{\mathrm{r(m)+q}}\right)$ where

$$\Omega^{*}(m)^{-\frac{1}{2}}A^{*}(m)=\Omega^{*}(m)^{-\frac{1}{2}}\Phi_{0}^{*}(m)\sqrt{\kappa}\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{Z}_{t}(m)\varepsilon_{t+1}^{*}=\Omega^{*}(m)^{-\frac{1}{2}}\Phi_{0}^{*}(m)\frac{1}{\sqrt{T-1}}\sum_{t=1}^{T-1}\hat{Z}_{t}(m)\xi_{t+1}^{*},\tag{30}$$

with $\Omega^{*}(m)=Var^{*}\left(\Phi_{0}^{*}(m)\frac{1}{\sqrt{T-1}}\sum_{t=1}^{T-1}\hat{Z}_{t}(m)\xi_{t+1}^{*}\right)$ and $\Psi_{t}^{*}(m)\equiv\Omega^{*-\frac{1}{2}}(m)\hat{Z}_{t}(m)\xi_{t+1}^{*}$. Finally,

$$A^{*}(m)\xrightarrow{P}N\left(0,\ \sigma^{2}\Phi_{0}^{*}(m)\left(\underset{N,T\to\infty}{\text{plim}}\frac{1}{T-1}Z(m)'Z(m)\right)\Phi_{0}^{*}(m)'\right).$$

Indeed, $\Omega^{*}(m)=\Phi_{0}^{*}(m)\left(\frac{1}{T-1-(r+q)}\sum_{t=1}^{T-1}\hat{Z}_{t}(m)\hat{Z}_{t}(m)'\left(\hat{\varepsilon}_{t+1}-\bar{\hat{\varepsilon}}\right)^{2}\right)\Phi_{0}^{*}(m)'$ with

$$\frac{1}{T-1-(r+q)}\sum_{t=1}^{T-1}\hat{Z}_{t}(m)\hat{Z}_{t}(m)'\left(\hat{\varepsilon}_{t+1}-\bar{\hat{\varepsilon}}\right)^{2}=\frac{1}{T-1-(r+q)}\sum_{t=1}^{T-1}\hat{Z}_{t}(m)\hat{Z}_{t}(m)'\hat{\varepsilon}_{t+1}^{2}+o_{P}(1)$$

and $\frac{1}{T-1-(r+q)}\sum_{t=1}^{T-1}\hat{Z}_{t}(m)\hat{Z}_{t}(m)'\hat{\varepsilon}_{t+1}^{2}$ an estimate of $\sigma^{2}\underset{N,T\to\infty}{\text{plim}}\frac{1}{T-1}Z(m)'Z(m)$ given Assumption 3. Hence, we have that $\Omega^{*}(m)\xrightarrow{P}\sigma^{2}\Phi_{0}^{*}(m)\left(\underset{N,T\to\infty}{\text{plim}}\frac{1}{T-1}Z(m)'Z(m)\right)\Phi_{0}^{*}(m)'$.

From **Part 1, Part 2 and Part 3,** and the fact that

$$\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{Z}_{t}^{*}(m)\hat{Z}_{t}^{*}(m)=\Phi_{0}^{*}(m)\left[\text{plim}\frac{1}{T-1}Z(m)'Z(m)\right]\Phi_{0}^{*}(m)'+o_{p^{*}}(1)$$

and

$$\sqrt{\kappa}\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*}(m)^{-1}\hat{\delta}(m)\right)=\left(\frac{1}{T-1}\hat{Z}^{*\prime}(m)\hat{Z}^{*}(m)\right)^{-1}\left[A^{*}(m)+o_{p^{*}}(1)\right],$$

by the asymptotic equivalence Lemma,

$$\sqrt{\kappa}\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*}(m)^{-1\prime}\hat{\delta}(m)\right)\xrightarrow{d^{*}}\mathrm{N}\left(0,\Phi_{0}^{*}(m)^{-1\prime}\left[\underset{N,T\to\infty}{\text{plim}}\frac{1}{T-1}Z(m)'Z(m)\right]^{-1}\Phi_{0}^{*}(m)^{-1}\right).$$

□

Proof of Theorem 3. We start by recalling that $F_{t}(m)$ is a generic limit of the candidate set of estimated factors $\tilde{F}_{t}(m)$. The proof begins showing first that if there is an $r_{0}\times r(m)$ matrix $Q(m)$ such that $F_{t}^{0}=Q(m)F_{t}(m)$ and no set of estimated factors such that the $r_{0}\times r(\check{m})$ matrix $Q(\check{m})$

such that $F_t^0 = Q(\check{m}) F_t(\check{m})$, then $P\left(\hat{\Gamma}_\kappa(m) < \hat{\Gamma}_\kappa(\check{m})\right) \longrightarrow 1$ as $T,\ N \to \infty$. Second, we show that if it exists an $r_0 \times r(m)$ matrix $Q(m)$ such that $F_t^0 = Q(m) F_t(m)$ and an $r_0 \times r(\check{m})$ matrix $Q(\check{m})$ such that $F_t^0 = Q(m) F_t(m)$, with $r(m) < r(\check{m})$, then $P\left(\hat{\Gamma}_\kappa(m) < \hat{\Gamma}_\kappa(\check{m})\right) \longrightarrow 1$. The first part corresponds to the case where only one set of estimated factor belongs to $\mathcal{M}_2$. However, in the second situation, our bootstrap selection rule picks the smaller set of estimated factor in $\mathcal{M}_2$.

Part 1: We observe that for any m,

$$\hat{\Gamma}_\kappa(m) = \mathrm{E}^*\left(\frac{1}{T-1}\left\|\left(\mathbf{y} - \hat{Z}(m)\hat{\delta}(m)\right) + \left(\hat{Z}(m)\hat{\delta}(m) - \hat{Z}^*(m)\hat{\delta}_\kappa^*(m)\right)\right\|^2\right) \equiv D_1(m)+D_2(m)+D_3(m),$$

where

$$\begin{aligned} D_1(m) &= \frac{1}{T-1}\left\|y - \hat{Z}(m)\hat{\delta}(m)\right\|^2 \\ D_2(m) &= \mathrm{E}^*\left(\frac{1}{T-1}\left\|\hat{Z}(m)\hat{\delta}(m) - \hat{Z}^*(m)\hat{\delta}_\kappa^*(m)\right\|^2\right) \end{aligned} \tag{31}$$

and

$$D_3(m) = 2\frac{1}{T-1}\left(y - \hat{Z}(m)\hat{\delta}(m)\right)'\mathrm{E}^*\left(\hat{Z}(m)\hat{\delta}(m) - \hat{Z}^*(m)\hat{\delta}_\kappa^*(m)\right).$$

Using the decomposition $\hat{Z}^*(m)\hat{\delta}_\kappa^*(m) - \hat{Z}(m)\hat{\delta}(m)$ as

$$\hat{Z}^*(m)\left(\hat{\delta}_\kappa^*(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right) + \left(\hat{Z}^*(m) - \hat{Z}(m)\Phi_0^*(m)'\right)\Phi_0^{*-1\prime}(m)\hat{\delta}(m),$$

where $\Phi_0^*(m)$ is an $(r(m)+q)\times(r(m)+q)$ submatrix of $\Phi_0^* = diag(\pm 1)$ the limit in probability of Φ^* conditionally on the sample (see, Gonçalves and Perron (2014)), we can write that $D_2(m) = D_{21}(m) + D_{22}(m) + 2D_{23}(m)$, with

$$\begin{aligned} D_{21}(m) &= \mathrm{E}^*\left(\left(\hat{\delta}_\kappa^*(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right)'\frac{1}{T-1}\hat{Z}^{*\prime}(m)\hat{Z}^*(m)\left(\hat{\delta}_\kappa^*(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right)\right), \\ D_{22}(m) &= \mathrm{E}^*\left(\hat{\delta}'(m)\Phi_0^{*-1}(m)\frac{1}{T-1}\left(\hat{Z}^*(m) - \hat{Z}(m)\Phi_0^*(m)'\right)'\left(\hat{Z}^*(m) - \hat{Z}(m)\Phi_0^*(m)'\right)\Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right) \end{aligned}$$

and

$$D_{23}(m) = \frac{1}{T-1}\mathrm{E}^*\left(\hat{\delta}'(m)\Phi_0^{*-1}(m)\left(\hat{Z}^*(m) - \hat{Z}(m)\Phi_0^*(m)'\right)'\hat{Z}(m)\left(\hat{\delta}_\kappa^*(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right)\right).$$

Starting with $D_{21}(m)$, we can show that

$$\begin{aligned} D_{21}(m) &= \mathrm{Trace}\left(\mathrm{E}^*\left(\frac{1}{T-1}\hat{Z}^{*\prime}(m)\hat{Z}^*(m)\left(\hat{\delta}_\kappa^*(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right)\right)\left(\hat{\delta}_\kappa^*(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right)'\right) \\ &= \frac{1}{\kappa}\mathrm{Trace}\left(\left(\Phi_0^*(m)\Sigma_Z(m)\Phi_0^{*\prime}(m)\right)\mathrm{Avar}^*\left(\sqrt{\kappa}\left(\hat{\delta}_\kappa^*(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right)\right)\right) + o_P\left(\frac{1}{\kappa}\right). \end{aligned}$$

Since Theorem 2 implies that as $\frac{\sqrt{\kappa}}{N} \to 0$,

$$\sqrt{\kappa}\left(\hat{\delta}_\kappa^*(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right) \xrightarrow{d^*} \mathrm{N}\left(0,\ \sigma^2\Phi_0^{*-1\prime}(m)\Sigma_Z(m)^{-1}\Phi_0^{*-1}(m)\right),$$

if m is in $\mathcal{M}_2$, we deduce that

$$\underset{N,T\to\infty}{\text{plim}} \text{ Avar}^*\left(\sqrt{\kappa}\left(\hat{\delta}^*_\kappa(m) - \Phi_0^{*-1\prime}(m)\,\hat{\delta}(m)\right)\right) = \sigma^2\Phi_0^{*-1\prime}(m)\,\Sigma_Z(m)^{-1}\,\Phi_0^{*-1}(m)\,.$$

Thereby, for any set m of estimated factor in $\mathcal{M}_2$,

$$D_{21}(m) = \frac{\sigma^2}{\kappa}\text{Trace}\left(\Phi_0^*(m)\,\mathbf{\Sigma}_Z(m)\,\Phi_0^{*\prime}(m)\,\Phi_0^{*-1\prime}(m)\,\Sigma_Z(m)^{-1}\,\Phi_0^{*-1}(m)\right)+o_P\left(\frac{1}{\kappa}\right) = \frac{\sigma^2\left(r(m)+q\right)}{\kappa}+o_P\left(\frac{1}{\kappa}\right).$$

For D_{22}, we use the fact that

$$D_{22}(m) = \hat{\delta}(m)'\,\Phi_0^{*-1}(m)\,\text{E}^*\left[\frac{1}{T-1}\left(\hat{Z}^*(m) - \hat{Z}(m)\,\Phi_0^*(m)\right)'\left(\hat{Z}^*(m) - \hat{Z}(m)\,\Phi_0^*(m)\right)\right]\Phi_0^{*-1\prime}(m)\,\hat{\delta}(m) \tag{32}$$

and $\text{E}^*\left(\frac{1}{T-1}\left(\hat{Z}^*(m) - \hat{Z}(m)\,\Phi_0^*(m)\right)'\left(\hat{Z}^*(m) - \hat{Z}(m)\,\Phi_0^*(m)\right)\right)$ is a submatrix of

$$D_{221} = \frac{1}{T-1}\sum_{t=1}^{T-1}\text{E}^*\left(\tilde{F}_t^* - H_0^*\tilde{F}_t\right)\left(\tilde{F}_t^* - H_0^*\tilde{F}_t\right)'. \tag{33}$$

Because, we treat $\tilde{F}_t^*$ as estimating $H_0^*\tilde{F}_t$, we can use the step of the proof of Gonçalves and Perron (2014, Lemma 3.1), and have that

$$\|D_{221}\| \le \frac{1}{T}\sum_{t=1}^{T}\text{E}^*\left\|\tilde{F}_t^* - H_0^*\tilde{F}_t\right\|^2 \le C\frac{4}{T}\sum_{t=1}^{T}\left(\text{E}^*\|A_{1t}^*\|^2 + \text{E}^*\|A_{2t}^*\|^2 + \text{E}^*\|A_{3t}^*\|^2 + \text{E}^*\|A_{4t}^*\|^2\right),$$

where

$$A_{1t}^* = \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^*\gamma_{st}^*,\ A_{2t}^* = \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^*\zeta_{st}^*,\ A_{3t}^* = \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^*\eta_{st}^*,\ A_{4t}^* = \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^*\xi_{st}^*,$$

with $\gamma_{st}^* = \text{E}^*\left(\frac{1}{N}\sum_{i=1}^N e_{is}^*e_{it}^*\right)$, $\zeta_{st}^* = \frac{1}{N}\sum_{i=1}^N\left(e_{is}^*e_{it}^* - \text{E}^*\left(\frac{1}{N}\sum_{i=1}^N e_{is}^*e_{it}^*\right)\right)$, $\eta_{st}^* = \frac{1}{N}\sum_{i=1}^N\tilde{\lambda}_i'\tilde{F}_s e_{it}^*$, and $\xi_{st}^* = \frac{1}{N}\sum_{i=1}^N\tilde{\lambda}_i'\tilde{F}_t e_{is}^*$. Note that, we ignore $\left\|V_0^{*-1}\right\|$, with V_0^* the limit of the matrix containing the first r eigenvalues of $X^*X^{*\prime}/(NT)$ in decreasing order, as it is bounded. Consequently, we find the order in probability of $\|D_{221}\|$ deriving those of $\frac{1}{T}\sum_{t=1}^T\text{E}^*\|A_{1t}^*\|^2$, $\frac{1}{T}\sum_{t=1}^T\text{E}^*\|A_{2t}^*\|^2$, $\frac{1}{T}\sum_{t=1}^T\text{E}^*\|A_{3t}^*\|^2$ and $\frac{1}{T}\sum_{t=1}^T\text{E}^*\|A_{4t}^*\|^2$.

First, by the Cauchy-Schwarz inequality, it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\text{E}^*\|A_{1t}^*\|^2 \le \frac{1}{T}\text{E}^*\left(\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s^*\right\|^2\right)\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\gamma_{st}^*\|^2\right)\right) = \frac{r}{T}\text{E}^*\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\gamma_{st}^*\|^2\right) = O_P\left(\frac{1}{T}\right) \tag{34}$$

using $\frac{1}{T}\sum_{s=1}^T\left\|\tilde{F}_s^*\right\|^2 = \text{Trace}\left(\frac{1}{T}\sum_{s=1}^T\tilde{F}_s^*\tilde{F}_s^{*\prime}\right) = \text{Trace}(I_r) = r$ and the fact that the high level condition $A^*(b): \frac{1}{T}\sum_{s=1}^T\sum_{t=1}^T\|\gamma_{st}^*\|^2 = O_P(1)$ of Gonçalves and Perron (2014) follows under our assumptions. Second, we also have by an application of the Cauchy-Schwartz inequality that

$$\frac{1}{T}\sum_{t=1}^{T}\text{E}^*\|A_{2t}^*\|^2 \le \text{E}^*\left(\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s^*\right\|^2\right)\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\zeta_{st}^{*2}\right)\right) \le r\cdot\text{E}^*\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\zeta_{st}^{*2}\right) = O_P\left(\frac{1}{N}\right) \tag{35}$$

given condition $A^*(c)$ of Gonçalves and Perron (2014). Thirdly, using the same arguments as Gonçalves and Perron (2014),

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}^*\left\|A_{3t}^*\right\|^2 \leq \mathrm{E}^*\left(\left\|\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^*\tilde{F}_s'\right\|^2\frac{1}{T^2}\sum_{s=1}^{T}\left\|\frac{1}{N}\tilde{\Lambda}'e_t^*\right\|^2\right) \leq r\cdot\mathrm{E}^*\left(\frac{1}{T^2}\sum_{s=1}^{T}\left\|\frac{1}{N}\tilde{\Lambda}'e_t^*\right\|^2\right) = O_P\left(\frac{1}{N}\right). \tag{36}$$

Similarly,

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}^*\left\|A_{4t}^*\right\|^2 = O_P\left(\frac{1}{N}\right), \tag{37}$$

and we can deduce from (34), (35), (36) and (37) that, $\|D_{221}\| = O_P\left(C_{NT}^{-2}\right)$. From (32), we deduce that $D_{22}(m) = O_P\left(C_{NT}^{-2}\right)$. Finally, we can write by an application of Cauchy-Schwartz inequality that $|D_{23}(m)| \leq \sqrt{D_{21}(m)}\sqrt{D_{22}(m)} = O_P\left(\frac{1}{\sqrt{\kappa}}\right)O_P\left(C_{NT}^{-1}\right) = O_P\left(\frac{1}{\sqrt{\kappa}C_{NT}}\right)$. Given the bound for $D_{21}(m)$, $D_{22}(m)$ and $D_{23}(m)$, it follows that,

$$D_2(m) = \frac{\sigma^2\left(r\left(m\right)+q\right)}{\kappa} + O_P\left(\frac{1}{\sqrt{\kappa}C_{NT}}\right). \tag{38}$$

We also have that

$$D_3(m) = \frac{-2}{T-1}\left(y-\hat{Z}\left(m\right)\hat{\delta}\left(m\right)\right)'\hat{Z}\left(m\right)\Phi_0^{*\prime}\left(m\right)\left(1+o_P\left(1\right)\right)E^*\left(\hat{\delta}_d^*\left(m\right)\right)+o_P\left(\frac{1}{T-1}\right) = o_P\left(\frac{1}{T-1}\right), \tag{39}$$

as $\left(y-\hat{Z}\left(m\right)\hat{\delta}\left(m\right)\right)'\hat{Z}\left(m\right) = 0$.

We now turn our attention to $D_1(m)$. Denoting $\tilde{M}\left(m\right) = I_{T-1} - \tilde{P}\left(m\right)$ and $M\left(m\right) = I_{T-1} - P\left(m\right)$, we have that from Lemma 7.1 that

$$\frac{1}{T-1}y'\tilde{M}\left(m\right)y = \frac{1}{T-1}y'M\left(m\right)y + O_P\left(\frac{1}{C_{NT}^2}\right),\quad \frac{1}{T-1}\varepsilon'M\left(m\right)\varepsilon = \frac{1}{T-1}\varepsilon'\varepsilon + O_P\left(\frac{1}{C_{NT}^2}\right).$$

Therefore, $D_1(m) = \frac{1}{T-1}y'M\left(m\right)y + O_P\left(\frac{1}{C_{NT}^2}\right)$, which is equal to

$$\frac{1}{T-1}\varepsilon'M\left(m\right)\varepsilon + \frac{1}{T-1}\delta'Z^{0\prime}M\left(m\right)Z^0\delta + 2\frac{1}{T-1}\delta'Z^{0\prime}M\left(m\right)\varepsilon + O_P\left(\frac{1}{C_{NT}^2}\right).$$

Using $\frac{1}{T-1}\delta'Z^{0\prime}M\left(m\right)\varepsilon = o_P\left(1\right)$ (see, Groen and Kapetanios (2013)), we consequently deduce that

$$D_1(m) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{1}{T-1}\delta'Z^{0\prime}M\left(m\right)Z^0\delta + O_P\left(\frac{1}{C_{NT}^2}\right). \tag{40}$$

From (38), (39) and (40), it follows that

$$\hat{\Gamma}_\kappa\left(m\right) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{1}{T-1}\delta'Z^{0\prime}M\left(m\right)Z^0\delta + \frac{1}{T-1}\delta'Z^{0\prime}M\left(m\right)\varepsilon + \frac{\sigma^2\left(r\left(m\right)+q\right)}{\kappa} + o_P\left(\frac{1}{\kappa}\right).$$

Given the assumptions that there exists matrix $Q\left(m\right)$ such that $F_t^0 = Q\left(m\right)F_t\left(m\right)$ and no matrix

$Q(\check{m})$ such that $F_t^0 = Q(\check{m}) F_t(\check{m})$, it follows that $M(m) Z^0 = 0$ and $M(\check{m}) Z^0 \neq 0$. Therefore,

$$\hat{\Gamma}_\kappa(m) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{\sigma^2(r(m)+q)}{\kappa} + o_P\left(\frac{1}{\kappa}\right) = \frac{1}{T-1}\varepsilon'\varepsilon + o_P\left(\frac{1}{\kappa}\right).$$

and

$$\hat{\Gamma}_\kappa(m') = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{1}{T-1}\delta' Z^{0\prime} M(m') Z^0\delta + o_P(1) = \sigma^2 + \frac{1}{T-1}\delta' Z^{0\prime} M(m') Z^0\delta + o_P(1)$$

as $\frac{1}{T-1}\delta' Z^{0\prime} M(m) Z^0\delta = \frac{1}{T-1}\delta' Z^{0\prime} M(m)\varepsilon = 0$, $\frac{1}{T-1}\delta' Z^{0\prime} M(\check{m})\varepsilon = o_P(1)$. Since

$$\operatorname*{plim\,inf}_{N,T\to\infty} \frac{1}{T-1}\delta' Z^{0\prime} M(\check{m}) Z^0\delta > 0,$$

if $M(\check{m}) Z^0 \neq 0$ given Assumption 4 (a), we have that

$$P\left(\hat{\Gamma}_\kappa(m) < \hat{\Gamma}_\kappa(\check{m})\right) = P\left(\sigma^2 < \sigma^2 + \frac{1}{T-1}\delta' Z^{0\prime} M(m') Z^0\delta + o_P(1)\right) \longrightarrow 1. \qquad (41)$$

Part 2: In this part of our proof, we show that if it exists a matrix $Q(m)$ such that $F_t^0 = Q(m) F_t(m)$ and a matrix $Q(\check{m})$ such that $F_t^0 = Q(m) F_t(m)$, with $r(m) < r(\check{m})$ then $P\left(\hat{\Gamma}_\kappa(m) < \hat{\Gamma}_\kappa(\check{m})\right)$ converges to 1. In this case,

$$\hat{\Gamma}_\kappa(m) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{\sigma^2(r(m)+q)}{\kappa} + o_P\left(\frac{1}{\kappa}\right) \text{ and } \hat{\Gamma}_\kappa(\check{m}) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{\sigma^2(r(\check{m})+q)}{\kappa} + o_P\left(\frac{1}{\kappa}\right).$$

Hence,

$$P\left(\hat{\Gamma}_\kappa(\check{m}) - \hat{\Gamma}_\kappa(m) > 0\right) = P\left(\sigma^2(r(\check{m}) - r(m)) > o + o_P(1) > 0\right) = 1 + o(1). \qquad (42)$$

From (41) and (42), we have the proof of Theorem 3. □

7.2 Simulation Results

Table 1: Average number of selected estimated factors

True latent factors	$T =$	CV_1 100	200	BICM 100	200	CV_d 100	200	$\hat{\Gamma}_\kappa$ 100	200
$(F_{t,1},\ F_{t,2})$	$N = 100$	2.34	2.38	1.55	1.68	2.05	2.00	2.16	2.03
	$N = 200$	2.32	2.40	1.73	1.87	2.05	2.08	2.14	2.14
$(F_{t,1},\ F_{t,2},\ F_{t,3})$	$N = 100$	3.10	3.17	2.54	2.63	2.92	2.93	3.01	2.96
	$N = 200$	3.10	3.16	2.67	2.81	2.95	3.01	3.01	3.03
$(F_{t,1},\ F_{t,2},\ F_{t,3},\ F_{t,4})$	$N = 100$	3.89	3.95	3.45	3.60	3.82	3.89	3.86	3.91
	$N = 200$	3.90	3.96	3.57	3.73	3.83	3.93	3.88	3.94

Note: This table reports the average number of selected estimated factors over 1000 simulations.

Table 2: Frequencies for DGP 1 in percentage (there are $2^4 = 16$ different possibilities)

	CV_1		BICM		CV_d		$\hat{\Gamma}_\kappa$	
$T =$	100	200	100	200	100	200	100	200
Selected estimated factors	$N = 100$		$N = 100$		$N = 100$		$N = 100$	
$\tilde{F}_{t,1}$	01.80	00.50	43.90	31.50	06.00	03.20	03.20	02.70
$\tilde{F}_{t,2}$	00.00	00.00	00.80	00.20	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3}\right)$	00.30	00.10	00.00	00.00	00.30	00.10	00.60	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,4}\right)$	00.40	00.10	00.00	00.00	00.30	00.00	00.30	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2}\right)^\bigstar$	64.90	64.20	55.30	68.30	83.10	93.40	77.70	91.70
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	00.10	00.00	00.00	00.00	00.10	0.00	00.10	0.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)$	17.50	18.80	00.00	00.00	05.90	02.40	09.20	04.70
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	11.20	12.50	00.00	00.00	03.90	00.90	07.40	00.90
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.10	00.00	00.00	00.00	00.10	00.00	00.10	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	03.70	03.80	00.00	00.00	00.30	00.00	01.40	00.00
Selected estimated factors	$N = 200$		$N = 200$		$N = 200$		$N = 200$	
$\tilde{F}_{t,1}$	01.60	00.20	27.30	12.70	04.50	0.60	03.00	00.40
$\tilde{F}_{t,2}$	0.00	00.00	00.10	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3}\right)$	00.30	00.00	00.00	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,4}\right)$	00.10	00.00	00.00	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2}\right)^\bigstar$	67.70	63.50	72.50	87.30	85.70	90.80	80.70	85.40
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)$	15.70	19.70	00.00	00.00	05.60	06.40	08.80	09.60
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	11.80	13.10	00.10	00.00	04.10	02.10	06.80	04.10
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.00	00.00	00.00	00.00	00.00	00.00	00.10	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	02.80	03.50	00.00	00.00	00.10	00.10	00.60	00.50

Note: The table reports the frequency of selecting each subset. ★ indicates the consistent set.

Table 3: Frequencies for DGP 2 in percentage (there are $2^4 = 16$ different possibilities)

	CV_1		BICM		CV_d		$\hat{\Gamma}_\kappa$	
$T =$	100	200	100	200	100	200	100	200
Selected estimated factors	$N = 100$		$N = 100$		$N = 100$		$N = 100$	
$\tilde{F}_{t,3}$	00.00	00.00	00.10	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t1},\ \tilde{F}_{t,3}\right)$	05.20	02.20	44.50	36.40	12.10	07.50	08.30	05.90
$\left(\tilde{F}_{t2},\ \tilde{F}_{t,3}\right)$	00.10	00.00	01.10	00.20	00.40	00.00	00.20	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)^\star$	77.70	77.30	54.30	63.40	82.70	91.60	81.30	92.10
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.00	01.10	00.00	00.00	00.40	00.00	01.00	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	01.80	00.00	00.00	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.10	00.00	00.00	00.00	00.10	00.00	00.10	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	15.10	19.40	00.00	00.00	04.30	00.90	09.10	02.00
Selected estimated factors	$N = 200$		$N = 200$		$N = 200$		$N = 200$	
$\tilde{F}_{t,3}$	00.00	00.00	00.10	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t1},\ \tilde{F}_{t,3}\right)$	03.50	01.20	32.80	19.10	09.00	02.80	05.60	02.50
$\left(\tilde{F}_{t2},\ \tilde{F}_{t,3}\right)$	00.00	00.00	00.60	00.10	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)^\star$	82.10	81.50	66.40	80.80	86.60	93.60	87.10	92.20
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	01.00	00.50	00.00	00.00	00.40	00.10	00.40	00.20
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	13.40	16.80	00.10	00.00	04.00	03.50	06.90	05.10

Note: See note for Table 2.

Table 4: Frequencies for DGP 3 in percentage (there are $2^4 = 16$ different possibilities)

	CV_1		BICM		CV_d		$\hat{\Gamma}_\kappa$	
T=	100	200	100	200	100	200	100	200
Selected estimated factors	$N = 100$		$N = 100$		$N = 100$		$N = 100$	
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,4}\right)$	00.00	00.00	01.60	00.20	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	00.00	00.00	00.30	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.00	00.00	01.30	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	00.10	00.00	01.00	00.10	00.10	00.00	00.10	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	10.30	04.60	45.00	38.80	17.80	11.00	13.40	09.30
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.40	00.00	02.40	00.60	00.50	00.10	00.50	00.10
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)^\star$	89.20	95.40	48.40	60.30	81.60	88.90	86.00	90.60
Selected estimated factors	$N = 200$		$N = 200$		$N = 200$		$N = 200$	
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,4}\right)$	00.00	00.00	00.40	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.00	00.00	00.50	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	00.10	00.00	00.80	00.00	00.30	00.00	00.10	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	09.20	03.80	38.20	27.20	15.80	06.90	11.60	05.60
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.30	00.00	02.00	00.00	00.60	00.00	00.40	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)^\star$	90.40	96.20	58.10	72.80	83.30	93.10	87.90	94.40

Note: See note for Table 2.

7.3 Empirical Application Details

We present here the empirical results.

Table 5: Variation explained by estimated macro in X_1 and financial factors in X_2

	Macro factors $\left(\tilde{F}\right)$		Financial factors $\left(\tilde{G}\right)$	
N°	Percentage (%)	Cumulative (%)	Percentage (%)	Cumulative (%)
1	24.06	24.06	71.56	71.56
2	9.52	33.58	4.10	75.66
3	8.04	41.62	3.62	79.28
4	5.87	47.49	1.72	81.00
5	4.13	51.62	1.47	82.47
6	3.25	54.87	1.17	83.64

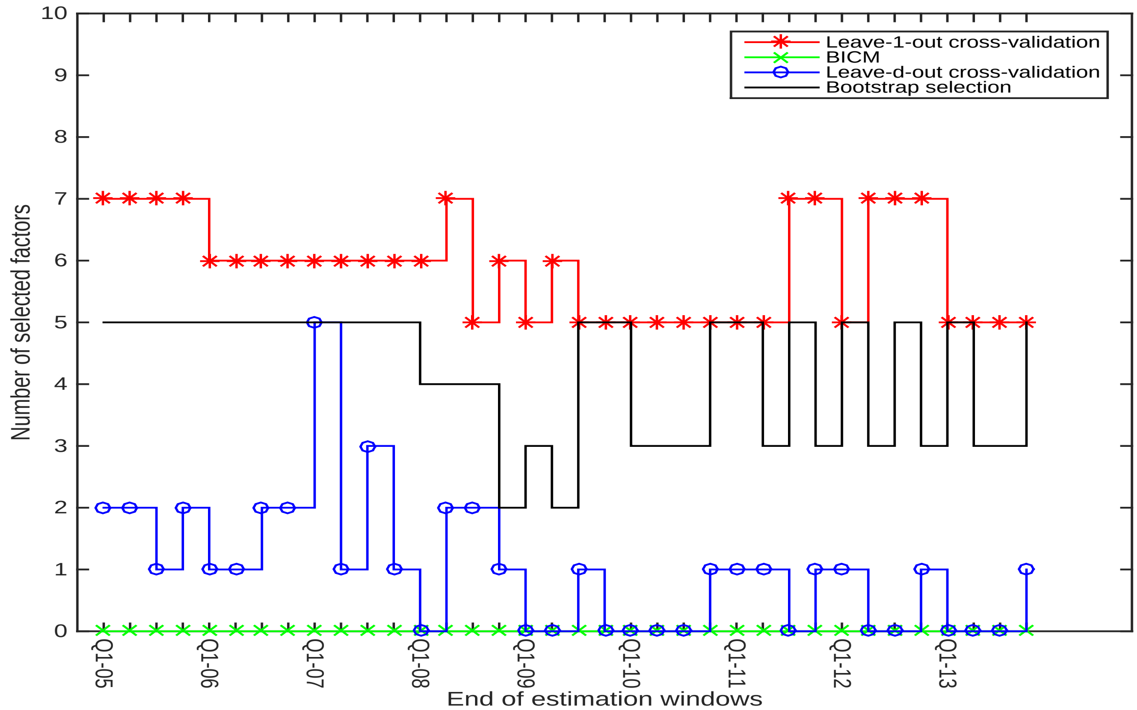
Note: The percentage of variation explained by each estimated factors is measured by the associated eigenvalue relative to the sum of the overall eigenvalues.

Table 6: Estimation results for $m_{t+1} = \alpha_1' (m) \tilde{F}_t (m) + \alpha_2' (m) \tilde{G}_t (m) + \beta Z_t + u_{t+1} (m)$

Regressors	CV_1	BICM	CV_d	$\hat{\Gamma}_\kappa$
constant	10.90★★	6.75	11.96★★	10.21★★
$(t - stat)$	(2.65)	(1.52)	(3.03)	(2.27)
CAY_t	21.05★	29.01★★	21.90★	23.17★
$(t - stat)$	(1.70)	(2.43)	(1.79)	(2.02)
$RREL_t$	0.50★	−0.33★	0.05	−0.17
$(t - stat)$	(1.59)	(−1.76)	(0.22)	(0.19)
$d - p_t$	1.86★★	1.06	2.00★★	1.75★★
$(t - stat)$	(2.69)	(1.44)	(2.98)	(3.03)
VOL_t	0.15★	0.05	0.12	0.16
$(t - stat)$	(1.83)	(0.47)	(1.16)	(0.83)
$\tilde{F}_{1t}$	−0.72★★			
$(t - stat)$	(−2.05)			
$\tilde{F}_{3t}$	1.34★★		1.01★★	0.97★★
$(t - stat)$	(3.63)		(3.08)	(2.53)
$\tilde{F}_{4t}$	−0.66★★			
$(t - stat)$	(−2.38)			
$\tilde{G}_{2t}$	0.59★★		0.64★★	0.63★★
$(t - stat)$	(2.02)		(2.69)	(2.44)
$\tilde{G}_{3t}$	0.49★★		0.59★★	0.61★★
$(t - stat)$	(2.01)		(2.01)	(2.51)
$\tilde{G}_{4t}$	−0.72★★			−0.71★★
$(t - stat)$	(−2.41)			(−2.37)
$\tilde{G}_{6t}$	−0.55★★			0.55★★
$(t - stat)$	(−2.13)			(1.98)
R^2	0.219	0.048	0.143	0.19
$F - test$	6.25		7.41	7.08
$F - cv$	2.05		3.04	2.26

Note: The estimated coefficients are reported. The student test statistic are presented into parenthesis. ★★ indicates the significant coefficients at 5% whereas those significant at 10% are indicated by ★. MOD_0 represents estimation results with usual factors that are not estimated from our economics data. These regressors are the consumption-wealth ratio (CAY), the relative T-bill (RREL), the dividend price ratio (d-p) and the sample volatility (VOL) of one-quarter-ahead excess returns. The other columns show estimates by selecting generated regressors and those in MOD_0. We tested whether the additional estimated factors are jointly significant. The Fisher test statistic corresponds to the difference between the sum squared residuals of MOD_0 and $\hat{m}_j$, $j = 1$, 2, 3 and 4, divided by the sum squared residuals of MOD_0 and corrected by the degrees of freedom. The critical values are based on the asymptotic result that the statistic follows a Fisher distribution with the number of additional parameters $r(\hat{m}_j)$ and $(T - 6) - r(\hat{m}_j)$ as degree of freedom.

Figure 3: Number of selected factors



Note: This figure reports the number of selected factors over the forecast sample 2005-Q1 to 2014-Q3. These numbers of estimated factors correspond to the number of estimated factors picked using the different methods.

References

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

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

Bai, J. and S. Ng (2009). Boosting diffusion indices. *Journal of Applied Econometrics 24*(4), 607–629.

Bai, J. and S. Ng (2013). Principal component estimation and identificaton of static factor model. *Journal of Econometrics 176*(1), 18–29.

Campbell, J. Y. (1991). A Variance Decomposition for Stock Returns. *Economic Journal 101*(405), 157–179.

Carrasco, M. and B. Rossi (2016). In-sample inference and forecasting in misspecified factor models. *Journal of Business & Economic Statistics 34*(3), 313–338.

Cheng, X. and B. E. Hansen (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics 186*(2), 280–293.

Djogbenou, A., S. Gonçalves, and B. Perron (2015). Bootstrap inference in regressions with estimated factors and serial correlation. *Journal of Time Series Analysis 36*(3), 481–502.

Fosten, J. (2017). Model selection with estimated factors and idiosyncratic components. *Journal of Applied Econometrics 32*(6), 1087–1106. jae.2567.

Gonçalves, S., M. W. McCracken, and B. Perron (2017). Tests of equal accuracy for nested models with estimated factors. *Journal of Econometrics 198*(2), 231–252.

Gonçalves, S. and B. Perron (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics 182*(1), 156–173.

Groen, J. and G. Kapetanios (2013). Model selection criteria for factor-augmented regressions. *Oxford bulletin of economics and statistics 75*(1), 37–63.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring Uncertainty. *American Economic Review 105*(3), 1177–1216.

Kelly, B. and S. Pruitt (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics 186*(2), 294–316.

Kleibergen, F. and Z. Zhan (2015). Unexplained factors and their effects on second pass R-squared's. *Journal of Econometrics 189*(1), 101–116.

Lettau, M. and S. C. Ludvigson (2001). Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance 56*(3), 815–849.

Lettau, M. and S. C. Ludvigson (2005). Expected returns and expected dividend growth. *Journal of Financial Economics 76*(3), 583–626.

Ludvigson, S. C. and S. Ng (2007). The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics 83*(1), 171–222.

McCracken, M. W. and S. Ng (2015). FRED-MD: A Monthly Database for Macroeconomic Research. Working Papers 2015-12, Federal Reserve Bank of St. Louis.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association 88*(422), 486–494.

Shao, J. (1996). Bootstrap model selection. *Journal of the American Statistical Association 91*(434), 655–665.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica 7*(2), 221–242.

Shiller, R. J. (1989). Comovements in Stock Prices and Comovements in Dividends. *Journal of*

Finance 44(3), 719–730.

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

Stone, M. (1974). Cross-validation and multinomial prediction. *Biometrika 61*(3), 509–515.