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†
York University
daa@yorku.ca

February 25, 2019

Abstract

This paper proposes two consistent in-sample 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 latent factors. The second proposed criterion is a generalization of the bootstrap approximation of the squared error of prediction from Shao (1996) to factor-augmented regressions. We show that these procedures are consistent model selection approaches. 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 stock market excess returns and factors extracted from a large panel of U.S. 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 in-sample predictive power for excess returns.

Keywords: Factor models, 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 since the seminal paper by Stock and Watson (2002). Unlike the traditional regressions, these models allow for the inclusion of a large set of macroeconomic and financial variables as predictors, which are useful for spanning various information sets related to economic agents. Thus, economic variables are considered to be driven by some unobservable factors that are inferred from a large panel of observed data.

*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), the CEA conference in Toronto (May 2015), and seminar participants at the Université de Montréal, the Université Libre de Bruxelles, Queen's University and York University. A special thanks to Marine Carrasco, Xu Cheng, Sílvia Gonçalves, William McCausland and Benoit Perron for useful comments and discussions. I thank the faculty of graduate studies of the Université de Montréal and the Queen's Economics Department for financial support.

†Address: Department of Economics, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada. Email: daa@yorku.ca.

Many empirical studies have been conducted using FAR. Among them, Ludvigson and Ng (2007) look at the risk−return relation in the equity market. With eight selected and estimated factors resuming the information in their macroeconomic and financial datasets using the IC_{p2} criterion of Bai and Ng (2002), they identified, based on the Bayesian information criterion (BIC), three new factors termed "volatility," "risk premium" and "real" factors that predict future excess returns. This paper mainly focuses on this two-step procedure that is widely used in practice. It considers factors estimated by the principal components method in the first step, and studies how to select the smallest basis of these estimated factors that drive a given regressand in the second step.

Considerable research has been devoted to detecting the number of factors that capture the information in a large panel of potential predictors, but very few studies have 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 prediction mean squared error (MSE) approximation, they pointed out that the standard BIC does not incorporate the factor estimation error. Consequently, they suggested a final prediction error (FPE) type of criterion with a penalty term depending on both the time and cross-sectional dimensions of the panel. Nevertheless, estimating consistently the MSE does not by itself ensure a consistent model selection. In fact, Groen and Kapetanios (2013) showed that this is true for the FPE criterion which inconsistently estimates the true factor space. As a result, they provided consistent model selection procedures that minimize the log of the sum of squared residuals and a penalty depending on time and cross-sectional dimensions. Their consistent model selection methods choose the smallest set of estimated factors that span the true factors, with a probability converging to one as the sample sizes grow. Although, these criteria are computationally less costly, in finite sample exercises, they tend to underestimate the true number of estimated factors spanning the true factors. In particular, they found in 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 set of estimated factors that estimates 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 in-sample prediction error term (independent of candidates models), the complexity error term (increases with the model dimension), the model identifiability term (zero for models with estimated factors spanning the true factor space), its parameter estimation and factor estimation errors. When only one set of generated factors spans the true model, this criterion converges to the in-sample regression error variance for this particular set because the identifiability component is zero and the remaining 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 that are 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 due to not only 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 because the factor estimation component dominates both the complexity component and the parameter estimation error component in finite sample. Our simulations show that this factor estimation error while asymptotically negligible, contributes to considerably reducing the probability in finite samples of selecting 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 because they select the smallest set of estimated factors spanning the true model with probability converging to one. The first is the Monte Carlo leave-d-out cross-validation method suggested by Shao (1993) in the context of observed and fixed regressors. The other method uses the bootstrap model 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.[1]

Overall, in comparison with the existing literature, this paper focuses on two-step FAR models widely used by practitioners. It does not assume that all extracted factors from a large panel are relevant for predictive purposes. Furthermore, because our interest is the role played by estimated factors in predicting a given dependent 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 leave-one-out cross-validation often selects a larger set of estimated factors than the smallest relevant one, while the modified BIC from Groen and Kapetanios (2013) tends to under-parameterize for smaller sample sizes. Nevertheless, the Monte Carlo leave-d-out cross-validation and the bootstrap model selection choose, with higher probability, the estimated factors spanning the true factors.

To illustrate the methods, an empirical application that revisits the relationship between macroe-

[1] Our study of consistent model selection purpose is to identify the estimated factors driving a targeted regressand during our in-sample periods. In situations where the purpose is out-of-sample forecasts, different approaches have been developed. Among others, these methods are the three-pass regression filter procedure by Kelly and Pruitt (2015), the frequentist model averaging approach by Cheng and Hansen (2015) and the regularization devices suggested by Carrasco and Rossi (2016).

conomic and financial factors, and excess stock returns in the U.S. market was 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 dataset 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, from a large panel of U.S. macro and financial data, the estimated factors heavily correlated with interest rate spreads and, along with the Fama−French factors, have strong additional in-sample predictive power for excess returns.

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 and tables appear in the Appendix.

2 Settings and Assumptions

The econometrician observes $(y_t,\ W_t',\ X_{1t},\dots,X_{it},\dots,X_{Nt})\,,t=1,\cdots,T$, and the goal is to model y_t using the following FAR model,

$$y_t=\delta' Z_t^0+\varepsilon_t, t=1,...,T, \tag{1}$$

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} is 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 $X=(X_{it})_{i=1,\dots,N,t=1,\dots,T}$ using principal components estimation. Hence, the estimated regression takes the form

$$y_t=\alpha\,(m)'\,\tilde{F}_t\,(m)+\beta' W_t+u_t\,(m)=\delta\,(m)'\,\hat{Z}_t\,(m)+u_t\,(m)\,, \tag{2}$$

where m is any of the 2^r subsets of indices in $\{1,\dots,r\}$ denoted $\mathcal{M}$ including the empty set, where no latent factor drives y_t. The size of $\tilde{F}_t\,(m)$ is $r\,(m)\le r$, and we assume the number of estimated factors selected in the first step is known and equal to r. The goal of this work is to select the smallest basis of the estimated factors by using the principal components method, which can recover the information in F_t^0. This is in line with applications in Ludvigson and Ng (2007), Ludvigson and Ng (2009) and Groen and Kapetanios (2013), where a smaller set of generated regressors was found to be driving their targeted dependent variable. In the case where none of the strict subset of the extracted factors is sufficient to recover the information in the underlying relevant factors for y_t, all of them is selected.

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 forecasts 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 in FAR with principal component method estimated factors.

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 strongly predicts the dependent variable and, therefore, accounts for the impact of idiosyncratic errors in the model selection process. In this paper, we focus on the two-step FAR, where the factors potentially affecting a large subset of the series in X are identified and used for prediction. We denote $Z_t = (F_t', W_t')'$, $t = 1, \ldots, T$, as the vector containing all latent factors and observed regressors, $\|M\| = (\text{Trace}\,(M'M))^{1/2}$, as the Euclidean norm, $Q > 0$, as the positive definiteness for any square matrix Q, and C, as 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) $\mathrm{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \left(e_{it}e_{is} - E\left(e_{it}e_{is}\right)\right)\right|^4 \leq C$ for all (t,s).

Assumption 2. (moments and weak dependence among $\{z_t\}$, $\{\lambda_i\}$, $\{e_{it}\}$ and $\{\varepsilon_t\}$)

(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 $\mathrm{E}\left(F_t e_{it}\right)=0$ for every (i,t).

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

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

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

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

(a) $\mathrm{E}\left(\varepsilon|X,Z\right)=0$, $\mathrm{E}\left(\varepsilon\varepsilon'|X,F,W\right)=\sigma^2 I_T$, $Z=\left[Z_1\cdots Z_T\right]'$, $\mathrm{E}\left\|Z_t\right\|^8<C$ and $\mathrm{E}\left(\varepsilon_t^8\right)<C$.

(b) $\Sigma_Z=\operatorname*{plim}_{T\longrightarrow\infty}\frac{1}{T}\sum_{t=1}^{T}Z_t Z_t'>0$.

(c) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t\varepsilon_t\xrightarrow{d}\mathrm{N}\left(0,\ \Omega\right)$, 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 FAR 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 regression errors are conditionally homoscedastic and independent, which is rather strong. These assumptions allow us to extend the consistent model selection results of Shao (1993) and Shao (1996) to the context of stochastic and estimated regressors. It could be relaxed if our interest was efficient model selection (Shao, 1997), but we leave this for future research.

The principal component method estimated $\tilde{F}$ 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 the identification conditions of Bai and Ng (2013). 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 previously argued, 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 $\tilde{F}_t$ that summarize the information in the large $T \times N$ matrix X. Afterwards, a subvector $\tilde{F}_t(m)$ is chosen for the prediction of y_t. Ludvigson and Ng (2007) select $\tilde{F}_t(m)$ from $\tilde{F}_t$ using the BIC to predict excess stock returns. Because this criterion does not correct for factor estimation, Bai and Ng (2009) suggest a modified FPE with an extra penalty to proxy the effect of factor estimation by approximating the 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 latent factors and the observed regressors at time t, the true conditional mean is

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

In the 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 m of regressors $F_t(m)$ as correct if the associated conditional mean equals that of the true unknown model, meaning

$$\alpha(m)' F_t(m) + \beta' W_t = E\left(y_t | F_t, W_t\right), \ t = 1, \ \ldots, \ T.$$

When the smallest set of regressors that generates the true model is selected 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 $\tilde{F}_t(m)$ spanning asymptotically the true unknown factors F_t^0. Formally, $\tilde{F}_t(m)$ spans F_t^0 or m is correct if $\tilde{F}_t(m) - F_t(m) \xrightarrow{P} 0$ and there is an $r_0 \times r(m)$ matrix $A(m)$ such that $F_t^0 = A(m) F_t(m)$. By definition, $F_t(m) = H_0(m) F_t$, where $H_0(m)$ is a $r(m) \times r$ sub-matrix of $H_0 = \underset{N,\, T \to \infty}{\mathrm{plim}} H$. 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 a subvector of $H_0 F_t$, where we avoid the subscript H_0 to simplify the notation. In the remainder of the paper, the only subvector of F_t that will be considered in this 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 sets 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. For the remainder of the paper, we will associate a set m of estimated factors $F_t(m)$ to the corresponding estimated model. That 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 Inconsistency of the Leave-one-out or Delete-one 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,

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

between y_t and its point fitted value using an estimate from the remaining time periods

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

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 the columns of $Z(m) = (F(m),\ W)$, with $F(m) : T \times r(m)$ the generic limit of $\tilde{F}(m) : T\times r(m)$ $\mu = Z^0\delta : T\times 1$, $W : T\times q$, the matrix of observed regressors, the true conditional mean vector, and Z^0 a $T\times(r_0+q)$ matrix with a typical element $Z_t^0 = (F_t^{0\prime}, W_t')'$.

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

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

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

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

where $V_T(m) = O_P\left(\frac{1}{\min\{N,T\}}\right)$, with $V_T(m)$ defined in the proof in the Appendix. When m is an *incorrect set of estimated factors,*

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

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

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

otherwise

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

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 the case where the regressors are observed, we have an additional term $V_T(m)$ corresponding to the factor estimation error in CV_1, and $P(m)$ is the projection matrix associated with the space spanned by the subset m of FH_0', a rotation of the true factor space. Consider two candidate sets m_1 and m_2 such that m_1 is incorrect and m_2 is correct. Assume $\underset{T\to\infty}{\text{plim inf}}\frac{1}{T}\mu'(I - P(m))\mu > 0$ for incorrect sets of estimated factors. The CV_1 will prefer m_2 to m_1 because

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

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

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

with $V_T(m) = O_P\left(\frac{1}{\min\{N,T\}}\right)$. The first term is independent of candidate models. The second term captures the complexity of the estimated model. It is the expected value of $\frac{1}{T}\varepsilon' P(m)\varepsilon$, and it increases with the model dimension. The term in parentheses 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 $CV_1(m)$ that is not reflected by the term in parentheses. Because the complexity component is inflated in finite samples by not only this parameter estimation error but also the factor estimation one, we fail to accurately select 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 $V_T(m) = O_P\left(\frac{1}{min\{N,T\}}\right) = O_P\left(\frac{1}{N}\right)$ and dominates both the complexity term and the parameter estimation error of order $O_P\left(\frac{1}{T}\right)$. More precisely, comparing two competing estimated models with estimated factors in $\mathcal{M}_2$ amounts to the comparison of their factor estimation errors in CV_1 instead of the model complexities because $V_T(m) = O_P\left(\frac{1}{N}\right)$ and

$$CV_1(m) = \frac{1}{T}\varepsilon'\varepsilon + V_T(m).$$

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

We consider the same data generating process (DGP) as the first DGP in the simulation section, where $y_t = 1 + F_{1t} + 0.5F_{2t} + \varepsilon_t$, with $\varepsilon_t \sim N(0,1)$ 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 conditions for identifying restrictions provided by Bai and Ng (2013) is asymptotically satisfied and make possible the identification of 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 that includes the estimation errors.

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

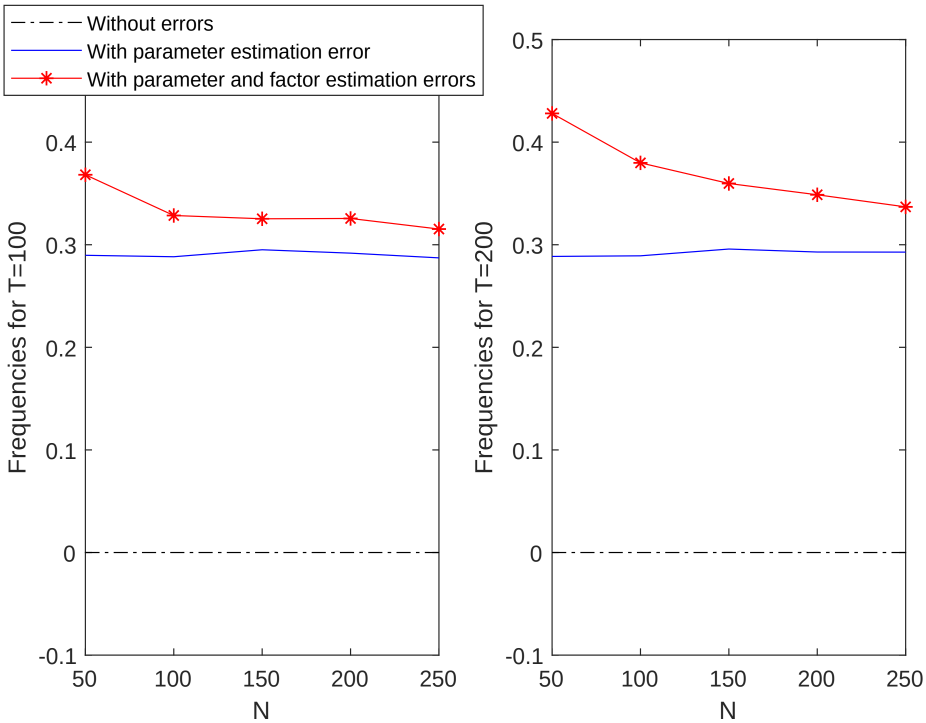
$$CV_{11}(m) = \frac{1}{T}\varepsilon'\varepsilon + \frac{(r(m)+q)}{T}\sigma^2 + \frac{1}{T}\mu'(I - P(m))\mu,$$

where we omit the parameter and the factor estimation errors. The second and third terms are those important for a consistent model selection. The 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

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

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-parameterization.

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 figure reports the frequencies of selecting a larger set of estimated factors than the one that contains the first two estimated factors. The line "without errors" represents the frequencies while minimizing the complexity component and the identifiability one plus $\frac{1}{T}\varepsilon'\varepsilon$. 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.

The sum of the complexity and the identifiability term in the CV_1, helpful to achieve a consistent selection of the estimated factors, corresponds to the conditional mean $\text{E}\left(L_T\left(m\right)|Z, X\right)$ of the infeasible in-sample squared error,

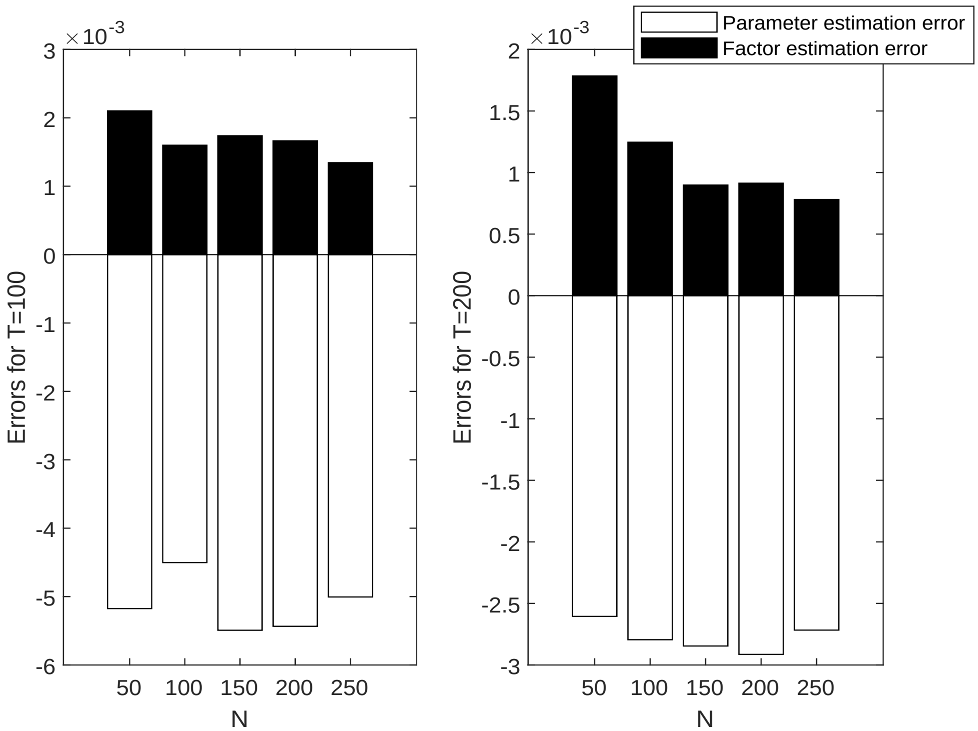
$$L_T\left(m\right) = \frac{1}{T}\left(\hat{\mu}\left(m\right) - \mu\right)'\left(\hat{\mu}\left(m\right) - \mu\right) = \frac{1}{T}\varepsilon' P\left(m\right)\varepsilon + \frac{1}{T}\mu'\left(I - P\left(m\right)\right)\mu,$$

with $\hat{\mu}\left(m\right) = P\left(m\right)y$. To achieve a consistent model selection, we consider alternative approaches.

3.2 Leave-d-out or Delete-d Cross-validation

To avoid the selection of larger models, Shao (1993) suggests a modification of the CV_1 in an observed regressors set-up, using a smaller construction sample to estimate δ by deleting $d \gg 1$ periods for validation. This consists of splitting the T time period observations into $\kappa = T - d$ randomly drawn observations without replacement that are used for parameter estimation and d

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 the note for Figure 1.

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 $\{1, \ldots, T\}$ called validation samples, for each draw $s = \{s(1), \cdots, s(d)\}$, we define

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

The corresponding construction sample is indexed by $s^c = \{1, \ldots, T\} \setminus s$, with $y_{s^c} : \kappa \times 1$ the complement of y_s in y, where $\kappa = T - d$, and $\hat{Z}_{s^c} : \kappa \times 1$ the complement of $\hat{Z}_s$ in $\hat{Z}$. We denote $\tilde{y}_s(m) = \hat{Z}_s(m)\hat{\delta}_{s^c}(m)$, $\hat{\delta}_{s^c}(m) = \left(\hat{Z}_{s^c}(m)' \hat{Z}_{s^c}(m)\right)^{-1} \hat{Z}_{s^c}(m)' y_{s^c} : (r(m) + q) \times 1$. 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 $\{1,\ldots,T\}$. This procedure generalizes the leave-one-out cross-validation because when $d=1$, $\kappa=T-d=T-1$, $s=\{t\}$, $s^c=\{1,\ldots,t-1,\ t+1,\ldots,T\}$ and $\mathcal{R}=\{\{1\},\ldots,\{T\}\}$, with $CV_d(m)=CV_1(m)$. Using a smaller construction sample, the next theorem shows that for correct sets of estimated factors,

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

and for incorrect sets of estimated factors,

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

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

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

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

Assumption 4.

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

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

(d) $\mathrm{E}(\mathrm{e_{it}e_{ju}})=\sigma_{\mathrm{ij,tu}}$ 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. This assumption was also made by Groen and Kapetanios (2013). 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-models, which is typical in the 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) because when $s^c=\{1,\ldots,T\}$, $\frac{1}{\sqrt{T\cdot\kappa}}\sum_{t\,\in\,s^c}\sum_{u=1}^{T}\frac{1}{N}\sum_{i,j}|\sigma_{ij,tu}|=\frac{1}{TN}\sum_{t,u,i,j}^{T}|\sigma_{ij,tu}|\le C$. Assumption 4 (e) and Assumption 4 (d) strengthen Assumption 2 (d) and

Assumption 3 (b), respectively. They are used for proving Lemma 7.2. The 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* $\frac{\kappa}{\min\{N,T\}} \to 0$, $\frac{T^2}{\kappa^2 b} \to 0$, κ, $d \to \infty$, *when* b, T, $N \to \infty$. *Then,*

$$P(\hat{m} = m_0) \to 1,$$

where $\hat{m} = \arg\min_m CV_d(m)$, *if* $\mathcal{M}_2$ *contains at least one correct set of estimated factors.*

The proof of Theorem 1 is given in the Appendix. This result is an extension of Shao (1993) in the case with generated regressors. Given the rate conditions, κ, $d \to \infty$ such that $\frac{\kappa}{T} \to 0$ and $\frac{d}{T} \to 1$. It follows from Theorem 1 that the consistency of the Monte Carlo leave-d-out cross-validation relies on κ being much smaller than d. Developing a general rule for the choice of κ is theoretically appealing, but beyond the scope of this paper. One could consider $\kappa = \min\{T,\ N\}^{3/4}$ and $d = T - \kappa$ because 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 $\kappa = T - 1$ and $d = 1$. The next paragraph studies an alternative selection procedure using a bootstrap method.

3.3 Bootstrap Rule for Model Selection

It follows from the previous subsection that the improvement in the Monte Carlo leave-d-out cross-validation relies on its ability to capture the complexity and the identifiability components in the conditional mean of the infeasible in-sample squared error $L_T(m)$. This is obtained by making the complexity component vanish 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 results from Shao (1996) to the FAR context, where we have generated regressors. We define $\hat{\Gamma}_\kappa(m)$, a bootstrap estimator of the prediction error mean, conditionally to Z and W, which is $\sigma^2 + \mathrm{E}(L_T(m)|Z,X)$, based on the two-step residual bootstrap procedure proposed by Gonçalves and Perron (2014) for FAR. In the case with observed regressors, Shao (1996) considers

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

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 that is conditional on the data. In the observed regressors context, they set the bootstrap version of the matrix of regressors to $Z^*(m) = Z(m)$, while 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 $\sqrt{\frac{T}{\kappa}}\frac{1}{\sqrt{1-\frac{r+q}{T}}}$, where $\kappa \to \infty$ such that $\frac{\kappa}{T} \to 0$. When $\kappa = T$, we obtain to the usual residual bootstrap. In fact, the factor $\sqrt{\frac{T}{\kappa}}$ ensures $\hat{\delta}^*_\kappa(m)$ to converge to δ at a slower rate $\sqrt{\kappa}$, that is more useful for consistent model selection than the usual $\sqrt{T}$. As for the leave-d-out cross-validation, $\kappa = o(T)$ such that $\frac{\kappa}{T} \to 0$ and $\frac{d}{T} \to 1$. If κ=$O(T)$, we have, similar to the leave-one-out cross-validation, a naive estimator of $\mathrm{E}\left(\mathrm{L_T}\left(\mathrm{m}\right)|\mathrm{Z},\mathrm{X}\right)$ 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 $\tilde{F} : T \times r$ from X, $\tilde{F}^* : T \times r$ is extracted from the bootstrap sample X^* and $\hat{Z}^* = \left(\tilde{F}^*,\ W\right)$. Subsets of $\tilde{F}^*$ are denoted by $\tilde{F}^*(m) : T \times r(m)$. We also define $\hat{Z}^*(m) = \left(\tilde{F}^*(m),\ W\right)$ and

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

where

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

with $\hat{Z}^*(m)$ and $y^*(m)$ the bootstrap analogs 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 $\hat{Z}(m) = \left(\tilde{F}(m),\ W\right)$.
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 $\{\varepsilon^*_t\}$ are bootstrap-residual based, respectively on $\{\hat{e}_{it}\}$, and $\{\hat{\varepsilon}_t\}$, with $\hat{\varepsilon}_t = \hat{\varepsilon}_t(M)$, 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 $i = 1,\dots,N$ and $t = 1,\dots,T$.
   (b) $\{\varepsilon^*_t\}_{t=1,\dots,T}$ are i.i.d. draws of $\left\{\sqrt{\frac{T}{\kappa}}\frac{1}{\sqrt{1-\frac{r+q}{T}}}\left(\hat{\varepsilon}_t(M) - \overline{\hat{\varepsilon}(M)}\right)\right\}_{t=1,\dots,T}$.
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 $y^*(m)$.

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

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

By multiplying the second step i.i.d. bootstrap residuals by $\frac{\sqrt{T}}{\sqrt{\kappa}}$, we obtain $\hat{\Gamma}_{\kappa}(m) = \frac{\varepsilon'\varepsilon}{T} + \frac{(r(m)+q)}{\kappa}\sigma^2 + o_P\left(\frac{1}{\kappa}\right)$ for m in $\mathcal{M}_2$ and $\hat{\Gamma}_{\kappa}(m) = \sigma^2 + \frac{1}{T}\mu'(I - P(m))\mu + o_P(1)$ 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* $\frac{\kappa}{\min\{N,T\}} \longrightarrow 0$, *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 the Appendix shows that our bootstrap scheme satisfies the high-level conditions provided by Gonçalves and Perron (2014). This result allows us to show that the bootstrap model selection procedure is consistent.

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* $\frac{\kappa}{\min\{N,T\}} \to 0$ *as* $T,\ N \to \infty$ *and* $E^*|\eta_{it}|^4 \leq C < \infty$. *Then,*

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

where $\hat{m} = \arg\min_m \hat{\Gamma}_{\kappa}(m)$, *if* $\mathcal{M}_2$ *contains at least one correct set of estimated factors.*

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

$$y_t = \alpha' F^0_t + \alpha_0 + \varepsilon_t,$$

where $\alpha_0 = 1$, $F^0_t \subset F_t \sim i.i.d.\text{N}(0, \text{I}_4)$ and $\varepsilon_t \sim i.i.d.\text{N}(0,1)$. Three DGPs are used first.

- For DGP 1, $r_0 = 2$, $F^0_t = (F_{t,1},\ F_{t,2})'$ and $\alpha = (1,\ 1/2)'$.
- For DGP 2, $r_0 = 3$, $F^0_t = (F_{t,1},\ F_{t,2},\ F_{t,3})'$ and $\alpha = (1,\ 1/2,\ -1)'$.
- For DGP 3, $r_0 = 4$, $F^0_t = (F_{t,1},\ F_{t,2},\ F_{t,3},\ F_{t,4})'$ and $\alpha = (1,\ 1/2,\ -1,\ 2)'$.

There are four factors, but only DGP 3 uses them all. DGPs 1 and 2 only use a subset of them to generate the dependent variable y_t. The panel factor model is a matrix of dimension $N \times T$, with elements

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

where $\lambda_{1i} \sim 12\mathrm{N}(0,1)$, $\lambda_{2i} \sim 8\mathrm{N}(0,1)$, $\lambda_{3i} \sim 4\mathrm{N}(0,1)$ and $\lambda_{4i} \sim \mathrm{N}(0,1)$. 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 asymptotically satisfies the 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 it possible to identify estimated factors, when N, $T \to \infty$ go to infinity. As in Djogbenou, Gonçalves, and Perron (2015), $e_{it} \sim \mathrm{N}(0, \sigma_i^2)$ with $\sigma_i \sim \mathrm{U}[.5,\ 1.5]$. We consider 1,000 simulations, 399 replications for the bootstrap model selection and the leave-d-out cross-validation approach, and for sample sizes, $T \in \{100,\ 200\}$, $N \in \{100,\ 200\}$. For the leave-d-out cross-validation and for the bootstrap model selection procedures, $d = T - \kappa$ and $\kappa = (\min\{T, N\})^{3/4}$. The first step bootstrap residuals 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 consistently select the true model to the leave-one-out cross-validation

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

and the BIC modified (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 $\hat{\sigma}^2(m) = \frac{1}{T - r(m) - 1}\sum_{t=1}^{T}\left(y_t - \tilde{F}_t(m)\hat{\delta}(m) - \hat{\alpha}_0\right)^2$, is made by considering subsets of the first four principal component estimated factors.

Table 1 presents 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 factors close to the true number.

The suggested procedures show a higher frequency of selecting factor estimates that span the true models 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 68.70 using the modified BIC and 63.80 using the leave-one-out cross-validation. The bootstrap selection method increases the frequency of the CV_1

by 23.2 percentage points and the CV_d increases it by 24.6 percentage points. 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 subsets of the consistent model. As DGP 3 corresponds to the largest model, CV_1 unsurprisingly performs well.

It is also possible that all the estimated factors estimating the true latent factors summarizing the information in the large panel X_{it} are needed. This corresponds to cases where the identification conditions are not satisfied. To illustrate that we consider a DGP 4, where we modify DGP 1 by changing

$$X_{it} = \lambda_i' F_t^0 + e_{it},$$

with $\lambda_{1i} \sim \mathrm{N}(0,1)$, $\lambda_{2i} \sim \mathrm{N}(0,1)$, and $F_t^0 \sim \mathrm{N}(0, \mathrm{I}_2)$. In this case, we search among all subsets of the first four principal components factor estimates. The simulation results presented in Table 5 show that the proposed model selection procedures have higher probabilities of selecting the first two estimated factors, which consistently estimate F_t^0.

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 the Jurado, Ludvigson, and Ng (2015) financial dataset and variables from the Kenneth R. French website, as described in the Supplemental Appendix.[2] 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 the McCracken and Ng (2015) data, as explained in the Supplemental Appendix. We examine how economic information summarized through a few numbers of estimated factors from real economic activity data and those related to financial markets can explain next quarter excess returns using various selection procedures. Recently, Gonçalves, McCracken, and Perron (2017) 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 market excess returns after controlling for the usual observed factors.

[2] We gratefully thank Sydney C. Ludvigson who provided us their dataset.

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

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

The variables F_t and G_t are latent and represent, respectively, the macroeconomic and 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 and included. The factors are estimated by $\tilde{F}_t$ and $\tilde{G}_t$ using principal components, 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 (see Table 6). Despite the imperfection of naming an estimated factor, it helps us understand the economic message revealed by the data. Figures in the Supplemental Appendix represent the marginal R^2 obtained by regressing each of the variables on the estimated factors.

In the panel, similar to McCracken and Ng (2015), $\tilde{F}_1$ is revealed as a real factor because variables related to production and the labor market are highly correlated to it. The third factor $\tilde{F}_3$, 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 are highly explained by $\tilde{G}_2$, whereas the small minus big Fama−French factor and Cochrane−Piazzesi factor are highly explained by $\tilde{G}_3$. The estimated factor $\tilde{G}_4$ is correlated with oil industry portfolio returns, and $\tilde{G}_6$ is mostly related to utility industry portfolio returns.

The estimated regression takes the form

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

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 7.

The Monte Carlo cross-validation and the bootstrap selection procedures select smaller sets 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 two factors: the third macro factor $\left(\tilde{F}_{3t}\right)$ and the third financial factor $\left(\tilde{G}_{3t}\right)$. Investors care about the spread between interest rates and effective federal funds rates, motivating interventions by the Federal Reserve to support economic expansion. Estimated risk factors also play an important role in predicting the equity premium associated with the U.S. stock market, as in Ludvigson and Ng (2007). We can deduce that the important estimated factors that investors in the U.S. financial market should care about are the interest rate spread factor $\left(\tilde{F}_{3t}\right)$ and market risk factors $\left(\tilde{G}_{3t}\right)$. These factors are simultaneously selected by the leave-d-out cross-validation and the bootstrap model selection approaches. The fact that $\tilde{F}_{3t}$ is always significant, confirms the importance of the spread between interest rates and effective federal funds in policies designed by central banks. We also study the joint significativity of the estimated factors using the F-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 estimated model for the different procedures except the BICM, where no factor is selected.

6 Conclusion

This paper suggests and provides conditions for the validity of two consistent model selection procedures for FAR models. It is the Monte Carlo leave-d-out cross-validation and the bootstrap selection approaches. In finite samples, the simulations document improvement in the probability of selecting the smallest set of estimated factors that span the true model in comparison to other existing consistent model selection methods. The procedures in this paper have been used to select estimated factors for in-sample predictions of one-quarter-ahead excess stock returns in the U. S. market. The in-sample analysis reveals that the estimated factor highly correlated with interest rate spreads, while the generated regressor highly correlated with the Fama−French factors, are driving the underlying unobserved factors and are predicting the excess returns in the U.S. market.

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

7.1 Proofs of results in Section 3

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

Lemma 7.1. *Under Assumptions 1–3,*

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

and

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

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

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

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

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

and

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

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

$O_P\left(\frac{1}{C_{NT}^2}\right)$ and $\frac{1}{T}W'\left[\tilde{F}-FH'\right]=O_P\left(\frac{1}{C_{NT}^2}\right)$, and it follows that $\frac{1}{T}Z'\left[\tilde{F}(m)-F(m)\right]=O_P\left(\frac{1}{C_{NT}^2}\right)$. Indeed, from their proof of Lemma A.2 (b),

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

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, $\frac{1}{T}F'\left[\tilde{F}-FH'\right]=O_P\left(\frac{1}{C_{NT}^2}\right)$, similarly $\frac{1}{T}W'\left[\tilde{F}-FH'\right]=O_P\left(\frac{1}{C_{NT}^2}\right)$, thus $L_{1T}(m)=O_P\left(\frac{1}{C_{NT}^4}\right)$ for any m. Since $\frac{1}{T}Z^{0\prime}Z(m)=O_P(1)$, 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

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

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

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

which is $O_P\left(\frac{1}{C_{NT}^2}\right)$, since $A_{01}(m)=\frac{1}{T}\left(Z(m)-\hat{Z}(m)\right)'\hat{Z}(m)=O_P\left(\frac{1}{C_{NT}^2}\right)$ and $A_{02}(m)=\frac{1}{T}Z(m)'\left(Z(m)-\hat{Z}(m)\right)=O_P\left(\frac{1}{C_{NT}^2}\right)$, 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 $\frac{1}{T}\mu'\tilde{P}(m)\mu=\frac{1}{T}\mu'P(m)\mu+O_P\left(\frac{1}{C_{NT}^2}\right)$. 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 $\gamma_{ut}=\mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N}e_{iu}e_{it}\right)$, $\zeta_{ut}=\frac{1}{N}\sum_{i=1}^{N}\left(e_{iu}e_{it}-\mathrm{E}\left(\frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}}\mathrm{e_{iu}e_{it}}\right)\right)$, $\eta_{ut}=\frac{1}{N}\sum_{i=1}^{N}\lambda_i'F_ue_{it}$, and $\xi_{ut}=\frac{1}{N}\sum_{i=1}^{N}\lambda_i'F_te_{iu}$. By the c-r inequality,

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

Using the Cauchy-Schwartz inequality, we have

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

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

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

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

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

We have the following decomposition

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

Hence, we can write

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

From that bound, it follows that

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

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

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

and

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

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

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

which is lower than $\frac{1}{T} \sum_{t=1}^{T} \left\| \hat{Z}_t(M) - Z_t(M) \right\|^2 = O_P\left(\frac{1}{C_{NT}^2} \right)$ (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), $\max_{1 \le t \le T} \left\| \hat{Z}_t(m) - Z_t(m) \right\| = O_P\left(\frac{1}{T^{1/2}} \right) + O_P\left(\frac{T^{1/2}}{N^{1/2}} \right)$, it follows that $A_3(m) = O_P\left(\frac{1}{T} \right) + O_P\left(\frac{1}{N^{1/2}} \right)$ as $\max_{1 \le t \le T} \left\| \hat{Z}_t(m) \right\| = O_P\left(T^{1/2} \right)$ (since $\frac{1}{T} \max_{1 \le t \le T} \left\| \hat{Z}_t(m) \right\|^2 = O_P(1)$). From the bounds of $A_1(m)$, $A_2(m)$ and $A_3(m)$, we deduce

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

This implies that

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

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

The remaining part of the proof goes similarly as the proof of Shao (1993, Equation 3.4). Noting that, $CV_1(m) = \frac{1}{T} \sum_{t=1}^{T} \left(1 - \hat{Z}_t'(m) \left(\hat{Z}'(m) \hat{Z}(m) \right)^{-1} \hat{Z}_t(m) \right)^{-2} \hat{\varepsilon}_t^2$ (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

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

since $\max_{1 \le t \le T} \left\| \hat{Z}_t'(m) \left(\hat{Z}'(m) \hat{Z}(m) \right)^{-1} \hat{Z}_t(m) \right\| = o_P(1)$. 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 $A_4(m) = \frac{1}{T}\varepsilon'\varepsilon + \tilde{L}_T(m) - 2r_{1T}(m)$, with

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

and

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

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

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

Under our Assumptions 1–3, for any m,

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

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

and

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

given Lemma 7.1. Hence, it follows that

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

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 $m \in \mathcal{M}_1$,

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

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

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

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 $m \in \mathcal{M}_1$,

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

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

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

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

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

and $\frac{1}{T}\hat{Z}'(m)\hat{Z}'(m) = \Sigma_Z(m) + o_P(1)$, it holds under Assumptions 1–3 that

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

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

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

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

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

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

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

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,

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

and

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

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

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

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

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

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

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

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

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

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

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

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 $m \in \mathcal{M}_1$,

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

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

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

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

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

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

$$B^*(m) = \frac{\sqrt{\kappa}}{T}\sum_{t=1}^T\left(\tilde{F}_t^*(m) - H_0^*(m)\,\tilde{F}_t(m)\right)\varepsilon_t^* = \frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\tilde{F}_t^*(m) - H_0^*(m)\,\tilde{F}_t(m)\right)\xi_t^*, \quad (28)$$

with $\xi_t^* = \frac{1}{\sqrt{1-\frac{r+q}{T}}}\left(\hat{\varepsilon}_t - \overline{\hat{\varepsilon}}\right)$. Given Gonçalves and Perron (2014, Lemma B2), $B^*(m) = O_P\left(\frac{1}{C_{NT}}\right)$ as long as $B^* = \frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\tilde{F}_t^* - H_0^*\tilde{F}_t\right)\xi_t^* = O_P\left(\frac{1}{C_{NT}}\right)$ if their Conditions $A^* - D^*$ are verified with ξ_t^* replacing ε_t^*. 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^* are independent and e_{it}^* is independent across (i,t),

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

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 $\mathrm{E}^*\left(\varepsilon_s^{*2}\right)=\frac{1}{1-\frac{r+q}{T}}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2-\bar{\hat{\varepsilon}}^2\right)\le\frac{1}{T-r-q}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$ and that $\mathrm{Var}^*\left(\eta_{it}\eta_{is}\right)$ 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 $\frac{1}{T-r-q}\sum_{t=1}^{T}\hat{\varepsilon}_t^2=O_P(1)$ under Assumptions 1–3, see Gonçalves and Perron (2014), the result follows. We now verify Condition $C^*(b)$. We have

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

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 $\mathrm{E}^*\left(\varepsilon_t^{*2}\right)\le\frac{1}{T-r-q}\sum_{s=1}^{T}\hat{\varepsilon}_s^2$ and $\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{e}_{it}^2\right)^2\le\frac{1}{T}\sum_{t=1}^{T}\tilde{e}_{it}^4$. 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 $\frac{1}{T}\sum_{t=1}^{T}\xi_t^{*4}=O_{p^*}(1)$ in probability. Using the definition of $\mathrm{E}^*\left(\xi_t^{*4}\right)$ and the c-r inequality,

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

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

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

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

$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Psi_t^*$, where Ψ_t^* are conditionally independent for $t=1,\ldots,T$, with $\mathrm{E}^*\left(\Psi_t^*\right)=\Omega^{*-\frac{1}{2}}\hat{Z}_t\mathrm{E}^*\left(\xi_t^*\right)=0$ and

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

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

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

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

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

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

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

where $\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^{2s}\leq\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^4\right)^{s/2}=O_P(1)$ 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 $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{Z}_t\xi_t^*\xrightarrow{d^*}\mathrm{N}\left(0,\ \Omega^*\right)$.

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

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{Z}^*(m)\left(\tilde{F}_t^*(m)-H_0^*(m)\tilde{F}_t(m)\right)'\left(H_0^*(m)\right)^{-1\prime}\hat{\alpha}(m)=O_P\left(\frac{\sqrt{T}}{N}\right)$$

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

$$C^*(m)=\sqrt{\kappa}\frac{1}{T}\sum_{t=1}^{T}\hat{Z}^*(m)\left(\tilde{F}_t^*(m)-H_0^*(m)\tilde{F}_t(m)\right)'\left(H_0^*(m)\right)^{-1\prime}\hat{\alpha}(m)=O_P\left(\frac{\sqrt{\kappa}}{N}\right)=o_P(1).$$

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

$$\Omega^*(m)^{-\frac{1}{2}}A^*(m)=\Omega^*(m)^{-\frac{1}{2}}\Phi_0^*(m)\sqrt{\kappa}\frac{1}{T}\sum_{t=1}^{T}\hat{Z}_t(m)\varepsilon_t^*=\Omega^*(m)^{-\frac{1}{2}}\Phi_0^*(m)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{Z}_t(m)\xi_t^*, \tag{30}$$

with $\Omega^*(m)=Var^*\left(\Phi_0^*(m)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{Z}_t(m)\xi_t^*\right)$ and $\Psi_t^*(m)\equiv\Omega^{*-\frac{1}{2}}(m)\hat{Z}_t(m)\xi_t^*$. Finally,

$$A^*(m)\xrightarrow{P}N\left(0,\ \sigma^2\Phi_0^*(m)\left(\underset{N,T\to\infty}{\text{plim}}\frac{1}{T}Z(m)'Z(m)\right)\Phi_0^*(m)'\right).$$

Indeed, $\Omega^*(m) = \Phi_0^*(m)\left(\frac{1}{T-(r+q)}\sum_{t=1}^{T}\hat{Z}_t(m)\hat{Z}_t(m)'\left(\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}\right)^2\right)\Phi_0^*(m)'$, with

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

and $\frac{1}{T-(r+q)}\sum_{t=1}^{T}\hat{Z}_t(m)\hat{Z}_t(m)'\hat{\varepsilon}_t^2$ an estimate of $\sigma^2 \underset{N,T\to\infty}{\text{plim}} \frac{1}{T}Z(m)'Z(m)$ given Assumption 3. Hence, we have that $\Omega^*(m) \xrightarrow{P} \sigma^2\Phi_0^*(m)\left(\underset{N,T\to\infty}{\text{plim}}\frac{1}{T}Z(m)'Z(m)\right)\Phi_0^*(m)'$.

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

$$\frac{1}{T}\sum_{t=1}^{T}\hat{Z}_t^*(m)\hat{Z}_t^*(m) = \Phi_0^*(m)\left[\text{plim}\frac{1}{T}Z(m)'Z(m)\right]\Phi_0^*(m)' + o_{p^*}(1)$$

and

$$\sqrt{\kappa}\left(\hat{\delta}_\kappa^*(m) - \Phi_0^*(m)^{-1}\hat{\delta}(m)\right) = \left(\frac{1}{T}\hat{Z}^{*\prime}(m)\hat{Z}^*(m)\right)^{-1}\left[A^*(m) + o_{p^*}(1)\right],$$

by the asymptotic equivalence Lemma,

$$\sqrt{\kappa}\left(\hat{\delta}_\kappa^*(m) - \Phi_0^*(m)^{-1\prime}\hat{\delta}(m)\right) \xrightarrow{d^*} \text{N}\left(0, \Phi_0^*(m)^{-1\prime}\left[\underset{N,T\to\infty}{\text{plim}}\frac{1}{T}Z(m)'Z(m)\right]^{-1}\Phi_0^*(m)^{-1}\right).$$

□

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 $\check{m}$ such that an $r_0 \times r(\check{m})$ matrix $Q(\check{m})$ satisfies $F_t^0 = Q(\check{m})F_t(\check{m})$, 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 factors belongs to $\mathcal{M}_2$. However, in the second situation, our bootstrap selection rule picks the smaller set of estimated factors in $\mathcal{M}_2$.

Part 1: We observe that for any m,

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

where

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

and

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

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 (Gonçalves and Perron, 2014), we can write that $D_2(m) = D_{21}(m) + D_{22}(m) + 2D_{23}(m)$, with

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

and

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

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

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

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

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

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

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

For D_{22}, we use the fact that

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

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

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

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

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

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 $\gamma_{st}^* = \mathrm{E}^*\left(\frac{1}{N}\sum_{i=1}^{N} e_{is}^* e_{it}^*\right)$, $\zeta_{st}^* = \frac{1}{N}\sum_{i=1}^{N}\left(e_{is}^* e_{it}^* - \mathrm{E}^*\left(\frac{1}{N}\sum_{i=1}^{N} e_{is}^* e_{it}^*\right)\right)$, $\eta_{st}^* = \frac{1}{N}\sum_{i=1}^{N}\tilde{\lambda}_i'\tilde{F}_s e_{it}^*$, 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 $\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}^*\left\|A_{1t}^*\right\|^2$, $\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}^*\left\|A_{2t}^*\right\|^2$, $\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}^*\left\|A_{3t}^*\right\|^2$ and $\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}^*\left\|A_{4t}^*\right\|^2$.

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

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

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

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

given condition $A^*\left(c\right)$ 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 obtain that $D_{22}(m) = O_P\left(C_{NT}^{-2}\right)$. Finally, we can write by an application of the Cauchy-Schwarz 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,

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

Moreover, we have that

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

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

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

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

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

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

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

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

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

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

Given the assumptions that there exists matrix $Q(m)$ such that $F_t^0 = Q(m)\,F_t(m)$ 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,

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

and

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

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

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

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

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

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,

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

Hence,

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

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

7.2 Simulation Results

Table 1: Average number of estimated factors that are selected

DGP	$T=$	CV_1 100	CV_1 200	BICM 100	BICM 200	CV_d 100	CV_d 200	$\hat{\Gamma}_{\kappa}$ 100	$\hat{\Gamma}_{\kappa}$ 200
DGP 1	$N=100$	2.36	2.39	1.56	1.69	2.04	2.00	2.15	2.03
	$N=200$	2.32	2.40	1.72	1.87	2.05	2.10	2.14	2.16
DGP 2	$N=100$	3.10	3.17	2.54	2.64	2.92	2.94	3.00	2.96
	$N=200$	3.10	3.16	2.67	2.81	2.95	3.01	3.02	3.03
DGP 3	$N=100$	3.89	3.95	3.45	3.61	3.82	3.88	3.86	3.91
	$N=200$	3.90	3.96	3.58	3.72	3.83	3.93	3.87	3.94
DGP 4	$N=100$	2.16	2.22	1.42	1.51	1.81	1.79	1.83	1.83
	$N=200$	2.18	2.18	1.54	1.59	1.86	1.87	1.87	1.86

Note: This table reports the average number of estimated factors that are selected over 1,000 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
Estimated factors	$N = 100$		$N = 100$		$N = 100$		$N = 100$	
$\tilde{F}_{t,1}$	01.80	00.40	43.20	31.20	06.00	03.50	03.20	02.70
$\tilde{F}_{t,2}$	00.00	00.00	00.80	00.10	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3}\right)$	00.70	00.20	00.00	00.00	00.20	00.10	00.60	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,4}\right)$	00.40	00.10	00.00	00.00	00.40	00.00	00.30	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2}\right)^\bigstar$	63.80	63.80	56.00	68.70	83.30	93.30	78.60	91.90
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	00.10	00.00	00.00	00.00	00.10	00.00	00.10	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)$	17.20	18.90	00.00	00.00	05.70	02.40	09.30	04.50
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	11.80	12.60	00.00	00.00	03.90	00.70	06.80	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
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	04.10	04.00	00.00	00.00	00.30	00.00	01.00	00.00
Estimated factors	$N = 200$		$N = 200$		$N = 200$		$N = 200$	
$\tilde{F}_{t,1}$	01.70	00.20	27.80	12.70	04.70	00.30	02.80	00.30
$\tilde{F}_{t,2}$	00.00	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	00.10	00.00	00.30	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
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2}\right)^\bigstar$	67.10	63.30	72.00	87.30	85.70	89.60	81.10	84.50
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)$	16.00	20.30	00.00	00.00	05.10	07.00	08.70	10.20
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	11.90	12.90	00.10	00.00	04.10	03.00	06.30	04.40
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.00	00.00	00.00	00.00	00.00	00.00	00.10	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	03.00	03.30	00.00	00.00	00.30	00.10	00.80	00.60

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
Estimated factors	$N = 100$		$N = 100$		$N = 100$		$N = 100$	
$\tilde{F}_{t,3}$	00.00	00.00	00.20	00.00	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t1},\ \tilde{F}_{t,3}\right)$	06.00	02.40	44.10	35.90	12.20	07.30	08.40	06.10
$\left(\tilde{F}_{t2},\ \tilde{F}_{t,3}\right)$	00.10	00.00	01.10	00.20	00.40	00.00	00.20	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)^\star$	76.80	77.80	54.60	63.90	82.60	91.80	81.90	92.30
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	01.20	00.90	00.00	00.00	00.30	00.10	00.80	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
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	15.80	18.90	00.00	00.00	04.40	00.80	08.60	01.60
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	32.70	19.10	09.00	02.80	05.50	02.40
$\left(\tilde{F}_{t2},\ \tilde{F}_{t,3}\right)$	00.00	00.00	00.50	00.10	00.00	00.00	00.00	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)^\star$	81.70	81.30	66.70	80.90	86.20	93.50	86.70	92.20
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	01.00	00.30	00.00	00.00	00.40	00.20	00.50	00.20
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	13.80	17.20	00.00	00.00	04.40	03.50	07.30	05.20

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
Estimated factors	$N = 100$		$N = 100$		$N = 100$		$N = 100$	
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,4}\right)$	00.00	00.00	01.50	00.20	00.10	00.00	00.10	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
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	11.10	04.60	45.40	38.40	17.40	11.70	13.40	09.40
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.20	00.00	02.40	00.60	00.60	00.10	00.40	00.10
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)^\star$	88.60	95.40	48.10	60.70	81.80	88.20	86.00	90.50
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
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	09.70	04.00	37.70	27.60	15.70	06.90	12.20	05.60
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.30	00.00	01.90	00.00	00.60	00.00	00.60	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)^\star$	89.90	96.00	58.70	72.40	83.40	93.10	87.10	94.40

Note: See note for Table 2.

Table 5: Frequencies for DGP 4 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
Estimated factors	$N = 100$		$N = 100$		$N = 100$		$N = 100$	
$\tilde{F}_{t,1}$	06.70	03.50	29.70	24.80	13.20	11.20	11.00	08.90
$\tilde{F}_{t,2}$	06.50	02.90	28.60	24.30	12.00	10.60	09.90	08.50
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3}\right)$	01.60	00.90	00.00	00.00	00.60	00.20	00.20	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,4}\right)$	01.70	00.70	00.00	00.00	00.70	00.00	00.10	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2}\right)^\star$	53.80	64.00	41.70	50.90	66.50	77.10	76.10	82.60
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)$	01.20	01.10	00.00	00.00	00.30	0.10	00.10	0.00
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	01.60	00.50	00.00	00.00	00.60	0.00	00.00	0.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)$	11.10	12.10	00.00	00.00	02.70	00.20	00.80	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	12.70	11.50	00.00	00.00	03.30	00.60	00.80	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.50	00.20	00.00	00.00	00.00	00.00	00.10	00.00
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.40	00.20	00.00	00.00	00.10	00.00	00.10	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	02.20	02.40	00.00	00.00	00.00	00.00	00.80	00.00
Estimated factors	$N = 200$		$N = 200$		$N = 200$		$N = 200$	
$\tilde{F}_{t,1}$	05.50	04.20	23.90	21.00	11.70	09.30	09.30	07.40
$\tilde{F}_{t,2}$	05.40	04.50	22.20	20.00	09.60	08.40	08.20	07.40
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3}\right)$	01.10	00.50	00.00	00.00	00.40	00.20	00.20	00.10
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,4}\right)$	01.10	00.90	00.00	00.00	00.70	00.40	00.70	00.10
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2}\right)^\star$	57.90	63.40	53.80	59.00	69.20	76.00	78.60	83.90
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)$	00.90	00.80	00.00	00.00	00.10	00.60	00.10	00.10
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	01.20	00.70	00.00	00.00	00.70	00.50	00.20	00.00
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3}\right)$	11.60	10.90	00.00	00.00	04.50	02.00	00.10	00.50
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,4}\right)$	12.70	12.00	00.10	00.00	02.80	02.40	00.90	00.30
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.10	00.20	00.00	00.00	00.00	00.00	00.10	00.00
$\left(\tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	00.10	00.00	00.00	00.00	00.00	00.00	00.10	00.10
$\left(\tilde{F}_{t,1},\ \tilde{F}_{t,2},\ \tilde{F}_{t,3},\ \tilde{F}_{t,4}\right)$	02.40	01.90	00.00	00.00	00.10	00.00	01.40	00.10

Note: See note for Table 2.

7.3 Empirical Application Details

We present here the empirical results.

Table 6: 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 7: Estimation results for $R_{t+1} = \alpha_1' (m) \tilde{F}_t (m) + \alpha_2' (m) \tilde{G}_t (m) + \beta W_t + u_{t+1} (m)$

Regressors	CV_1	BICM	CV_d	$\hat{\Gamma}_\kappa$
$constant$	10.85★★	6.56	11.27★★	10.22★★
$(t-stat)$	(2.63)	(1.48)	(2.51)	(2.26)
CAY_t	20.23	28.13★★	21.30★	22.42★
$(t-stat)$	(1.64)	(2.37)	(1.77)	(1.74)
$RREL_t$	0.50	−0.33★	0.06	−0.16
$(t-stat)$	(1.59)	(−1.75)	(0.26)	(−0.59)
$d-p_t$	1.84★★	1.03	1.89★★	1.75★★
$(t-stat)$	(2.67)	(1.40)	(2.52)	(2.31)
VOL_t	0.15★	0.05	0.12	0.15★★
$(t-stat)$	(1.81)	(0.46)	(1.28)	(1.97)
$\tilde{F}_{1t}$	−0.71★★			
$(t-stat)$	(−2.05)			
$\tilde{F}_{3t}$	1.35★★		0.99★★	0.98★★
$(t-stat)$	(3.67)		(2.86)	(2.58)
$\tilde{F}_{4t}$	−0.65★★			
$(t-stat)$	(−2.35)			
$\tilde{G}_{2t}$	0.59★★			0.63★★
$(t-stat)$	(2.46)			(2.46)
$\tilde{G}_{3t}$	0.49★★		0.58★	0.63★★
$(t-stat)$	(2.04)		(1.77)	(2.51)
$\tilde{G}_{4t}$	−0.71★★			−0.70★★
$(t-stat)$	(−2.42)			(−2.37)
$\tilde{G}_{6t}$	0.55★★			0.55★★
$(t-stat)$	(2.12)			(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 t-test statistics are presented into parenthesis. ★★ indicates the significant coefficients at 5% whereas those significant at 10% are indicated by ★. We control for some usual observed 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 columns show coefficient estimates when additional generated regressors $\hat{m}_j$, $j = 1, 2, 3$ are selected by the different procedures. We tested whether the additional estimated factors are jointly significant. The F-test statistic corresponds to the difference between the sum of squared residuals of the estimated model without estimated factors and when $\hat{m}_j$, $j = 1, 2, 3$ and 4, are included divided by the sum of squared residuals of the estimated model without estimated factors, corrected by the degrees of freedom. The critical values are based on the 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 degrees of freedom.

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.

Ludvigson, S. C. and S. Ng (2009). Macro Factors in Bond Risk Premia. *Review of Financial Studies 22*(12), 5027–5067.

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.

Supplemental Appendix for "Model Selection in Factor-Augmented Regressions with Estimated Factors"

Antoine A. Djogbenou*
Yorku University
daa@yorku.ca

February 25, 2019

Abstract

This is the supplemental appendix for Djogbenou (2019). Two sections are included. The first provides details on the data used. It consists of a large scale of 277 macroeconomic and financial variables in the U.S. economy. These datasets are constructed following Jurado, Ludvigson, and Ng (2015) and McCracken and Ng (2015). The second section presents the plotted R^2 while regressing the variables on each of the estimated factors. These plots are useful for understanding the economic information revealed by the estimated macroeconomic and financial factors.

Keywords: Factor models, big economic data, macroeconomic and financial series, macroeconomic and financial factors.

1 Data

The macroeconomic data are formed following McCracken and Ng (2015). Four series are dropped to obtain a balanced dataset indexed from 1 to 130, as listed below. This macro data contains eight groups of variables related to output and income (Group 1); labor market (Group 2); housing (Group 3); consumption, orders and inventories (Group 4), money and credit (Group 5); interest rates and exchange rates (Group 6); prices (Group 7) and stock market (Group 8). The quarterly version of McCracken and Ng (2015) are downloaded from the St. Louis Federal Reserve Economic Data (FRED). Since not all the data are available on the FRED website or some have missing values, we complete the dataset by aggregating the appropriate monthly data from McCracken and Ng (2015). These variables are listed with a star. Afterwards, the data are transformed to ensure stationarity. In the Tcode column, 1, 2, 3, 4, 5, 6, and 7 correspond to level, first difference, second difference, log transformation, first difference of the log, second difference of the log and growth rate, respectively.

*We gratefully thank Sydney C. Ludvigson who provided us with the dataset used in Jurado, Ludvigson, and Ng (2015). We would also like to thank Michael McCracken and Serena Ng for making public the data that are used in McCracken and Ng (2015).

The financial data series are indexed from 1 to 147 and correspond to the Jurado, Ludvigson, and Ng (2015) database. This dataset includes the group representing dividends and yields, the group of risk factors, the group of industry portfolios and the portfolios sorted on size and book-to-market ratio group. Because the data from Jurado, Ludvigson, and Ng (2015) are monthly, we downloaded quarterly available from the Kenneth R. French database and constructed the remaining one using similar steps as Jurado, Ludvigson, and Ng (2015). The quarterly returns of portfolios are obtained by computing the three-month returns from monthly versions in the Kenneth R. French database. We also applied $Log\,(1 + x/100)$ times 400 instead of 1,200 used by Jurado, Ludvigson, and Ng (2015) to have the corresponding annual version. Except for the logged dividend price ratio which corresponds in our database to the end of the corresponding quarter in Jurado, Ludvigson, and Ng (2015), the variables in Group 1 are summed over the quarter using monthly data from Jurado, Ludvigson, and Ng (2015). As in Ludvigson and Ng (2007), the quarterly CP factor of Cochrane and Piazzesi (2005) is its average over the quarter.

1.1 Macroeconomic Series

Group 1 : Output and Income

No.	Code	Description	Tcode
1	RPI	Real Personal Income	5
2	W875RX1	RPI ex.Transfers	5
3	INDPRO	Industrial Production Index	5
4	PFPNSS	IP Final Products and Supplies	5
5	IPFINAL	IP Final Products	5
6	IPCONGD	IP Consumer Goods	5
7	IPDCONGD	IP Durable Consumer Goods	5
8	IPNCONGD	IP Nondurable Consumer Goods	5
9	IPBUSEQ	IP Business Equipment	5
10	IPMAT	IP Materials	5
11	IPDMAT	IP Durable Materials	5
12	IPNMAT	IP Nondurable Materials	5
13	IPMANSICS	IP Manufacturing	5
14	IPB51222S	IP Residential Utilities	5
15	IPFUELS	IP Fuels	5
16	NAPMPI	ISM Manufacturing: Production Index	1
17	CUMFNS	Capacity Utilization: Manufacturing	2

Group 2 : Labor Market

No.	Code	Description	Tcode
18★	HWI	Help-Wanted Index for US	2
19★	HWIURATIO	Ratio of Help Wanted to Number of.Unemployed	2
20	CLF16OV	Civilian Labor Force	5
21	CE16OV	Civilian Employment	5
22	UNRATE	Civilian Unemployment Rate	2
23	UEMPMEAN	Average Duration of Unemployment	2
24	UEMPLT5	Civilians Unemployed less than 5 Weeks	5
25	UEMP5TO14	Civilians Unemployed 5-14 Weeks	5
26	UEMP15OV	Civilians Unemployed greater than 15 Weeks	5
27	UEMP15T26	Civilians Unemployed 15-26 Weeks	5
28	UEMP27OV	Civilians Unemployed greater than 27 Weeks	5
29★	CLAIMSx	Initial Claims	5
30	PAYEMS	All Employees: Total non farm	5
31	USGOOD	All Employees: Goods-Producing	5
32	CES1021000001	All Employees: Mining and Logging	5
33	USCONS	All Employees: Construction	5
34	MANEMP	All Employees: Manufacturing	5
35	DMANEMP	All Employees: Durable goods	5
36	NDMANEMP	All Employees: Nondurable goods	5
37	SRVPRD	All Employees: Service Industries	5
38	USTPU	All Employees: TT&U	5
39	USWTRADE	All Employees: Wholesale Trade	5
40	USTRADE	All Employees: Retail Trade	5
41	USFIRE	All Employees: Financial Activities	5
42	USGOVT	All Employees: Government	5
43	CES0600000007	Hours: Goods-Producing	1
44	AWOTMAN	Overtime Hours: Manufacturing	2
45	AWHMAN	Hours: Manufacturing	1
46	NAPMEI	ISM Manufacturing: Employment	1
47	CES0600000008	Ave. Hourly Earnings: Goods	6
48	CES2000000008	Ave. Hourly Earnings: Construction	6
49	CES3000000008	Ave. Hourly Earnings: Manufacturing	6

Group 3 : Housing

No.	Code	Description	Tcode
50	HOUST	Starts:Total	4
51	HOUSTNE	Starts:Northeast	4
52	HOUSTMW	Starts:Midwest	4
53	HOUSTS	Starts:South	4
54	HOUSTW	Starts:West	4
55	PERMIT	Permits	4
56	PERMITNE	Permits: Northeast	4
57	PERMITMW	Permits: Midwest	4
58	PERMITS	Permits: South	4
59	PERMITW	Permits: West	4

Group 4 : Consumption, Orders and Inventories

No	Code	Description	Tcode
60	DPCERA3M086SBEA	Real PCE	5
61★	CMRMTSPLx	Real M&T Sales	5
62★	RETAILx	Retail and Food Services Sales	5
63	NAPM	ISM: PMI Composite Index	1
64	NAPMNOI	ISM: New Orders Index	1
65	NAPMSDI	ISM: Supplier Deliveries Index	1
66	NAPMII	ISM: Inventories Index	1
67★	AMDMNOx	Orders: Durable Goods	5
68★	AMDMUOx	Unfilled Orders: Durable Goods	5
69★	BUSINVx	Total Business Inventories	5
70★	ISRATIOx	Inventories to Sales Ratio	2

Group 5 : Money and Credit

No.	Code	Description	Tcode
71	M1SL	M1 Money Stock	6
72	M2SL	M2 Money Stock	6
73	M2REAL	Real M2 Money Stock	5
74	AMBSL	St.Louis Adjusted Monetary Base	6
75	TOTRESNS	Total Reserves	6
76	NONBORRES	Non borrowed Reserves	6
77	BUSLOANS	Commercial and Industrial Loans	6
78	REALLN	Real Estate Loans	1
79	NONREVSL	Total Non revolving Credit	6
80★	CONSPI	Credit to PI ratio	2
81	MZMSL	MZM Money Stock	6
82	DTCOLNVHFNM	Consumer Motor Vehicle Loans	6
83	DTCTHFNM	Total Consumer Loans and Leases	6
84	INVEST	Securities in Bank Credit	6

Group 6 : Interest Rate and Exchange Rates

No.	Code	Description	Tcode
85	FEDFUNDS	Effective Federal Funds Rate	2
86★	CP3M	3-Month AA Financial Commercial Paper Rate	2
87	TB3MS	3-Month T-bill	2
88	TB6MS	6-Month T-bill	2
89	GS1	1-Year T-bond	2
90★	GS5	5-Year T-bond	2
91	GS10	10-Year T-bond	2
92	AAA	Moody's Seasoned Aaa Corporate Bond Yield	2
93	BAA	Moody's Seasoned Baa Corporate Bond Yield	2
94★	COMPAPFF	3-Month Commercial Paper Minus FEDFUNDS	1
95	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	1
96	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS	1
97	T1YFFM	1-Year Treasury C Minus FEDFUNDS	1

Group 6 : Interest Rate and Exchange Rates (cont.)

No.	Code	Description	Tcode
98	T5YFFM	5-Year Treasury C Minus FEDFUNDS	1
99	T10YFFM	10-Year Treasury C Minus FEDFUNDS	1
100	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS	1
101	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS	1
102★	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	5
103★	EXJPUSx	Japan / U.S. FX Rate	5
104★	EXUSUKx	U.S. / U.K. FX Rate	5
105★	EXCAUSx	Canada / U.S. FX Rate	5

Group 7 : Prices

No.	Code	Description	Tcode
106	PPIFGS	Producer Price Index: Finished Goods	6
107	PPIFCG	PPI: Finished Consumer Goods	6
108	PPIITM	PPI: Intermediate Materials	6
109	PPICRM	PPI: CrudeMaterials	6
110★	OILPRICEx	Crude Oil Prices: WTI	6
111	PPICMM	PPI: Commodities	6
112	NAPMPRI	ISM Manufacturing: Prices	1
113	CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items	6
114	CPIAPPSL	CPI for All Urban Consumers: Apparel	6
115	CPITRNSL	CPI for All Urban Consumers: Transportation	6
116	CPIMEDSL	CPI for All Urban Consumers: Medical Care	6
117	CUSR0000SAC	CPI for All Urban Consumers: Commodities	6
118	CUUR0000SAD	CPI for All Urban Consumers: Durables	6
119	CUSR0000SAS	CPI for All Urban Consumers: Services	6
120	CPIULFSL	CPI for All Urban Consumers: All Items Less Food	6
121	CUUR0000SA0L2	CPI for All Urban Consumers: All items less shelter	6
122	CUSR0000SA0L5	CPI for All Urban Consumers: All items less medical care	6
123	PCEPI	Personal Consumption Expenditures: Chain-type Price Index	6
124	DDURRG3M086SBEA	Personal Consumption Expenditures: Durable goods	6
125	DNDGRG3M086SBEA	Personal Consumption Expenditures: Nondurable goods	6
126	DSERRG3M086SBEA	Personal Consumption Expenditures: Services	6

Group 8 : Stock Market

No.	Code	Description	Tcode
127★	S&P 500	S&P's Common Stock Price Index: Composite	5
128★	S&P: indust	S&P's Common Stock Price Index: Industrials	5
129★	S&P div yield	S&P's Composite Common Stock: Dividend Yield	2
130★	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio	5

1.2 Financial Data Set

Group 1 : Yield and Dividends

No.	Code	Description	Tcode
1	D_log(DIV)	Log difference of sum of the dividends in the last 4 quarters	1
2	D_log(P)	Log difference of CRSP portfolio price when dividends are not reinvested	1
3	D_DIVreinvested	Log difference of sum of the dividends in the last 4 quarters (reinvested)	1
4	D_Preinvested	Log difference of CRSP portfolio price when dividends are reinvested	1
5	d-p	DIVreinveste - Preinveste = log(DIV) - log(P)	1

Group 2 : Risk Factors

No.	Code	Description	Tcode
6	R15-R11	Small stock value spread: (Small, High) minus (Small, Low) sorted on (size, B/M)	1
7	factor	Piazzesi-Cochrane risk factor, quarterly average	1
8	Mkt-RF	Fama-French market risk factor: Market excess return	1
9	SMB	Fama-French market risk factor: Small Minus Big, sorted on size	1
10	HML	Fama-French market risk factor: High Minus Low, sorted on book-to-market	1
11	UMD	Momentum risk factor: Up Minus Down, sorted on momentum	1

Group 3 : Industries Portfolio

No.	Code	Description	Tcode
12	Agric	Agric industry portfolio	1
13	Food	Food industry portfolio	1
14	Beer	Beer industry portfolio	1
15	Smoke	Smoke industry portfolio	1
16	Toys	Toys industry portfolio	1
17	Fun	Fun industry portfolio	1
18	Books	Books industry portfolio	1
19	Hshld	Hshld industry portfolio	1
20	Clths	Clths industry portfolio	1
21	MedEq	MedEq industry portfolio	1
22	Drugs	Drugs industry portfolio	1
23	Chems	Chems industry portfolio	1
24	Rubbr	Rubbr industry portfolio	1
25	Txtls	Txtls industry portfolio	1
26	BldMt	BldMt industry portfolio	1
27	Cnstr	Cnstr industry portfolio	1
28	Steel	Steel industry portfolio	1
29	Mach	Mach industry portfolio	1
30	ElcEq	ElcEq industry portfolio	1
31	Autos	Autos industry portfolio	1
32	Aero	Aero industry portfolio	1
33	Ships	Ships industry portfolio	1
34	Mines	Mines industry portfolio	1

Group 3 : Industries Portfolio (cont.)

No.	Code	Description	Tcode
35	Coal	Coal industry portfolio	1
36	Oil	Oil industry portfolio	1
37	Util	Util industry portfolio	1
38	Telcm	Telcm industry portfolio	1
39	PerSv	PerSv industry portfolio	1
40	BusSv	BusSv industry portfolio	1
41	Comps	Comps industry portfolio	1
42	Chips	Chips industry portfolio	1
43	LabEq	LabEq industry portfolio	1
44	Paper	Paper industry portfolio	1
45	Boxes	Boxes industry portfolio	1
46	Trans	Trans industry portfolio	1
47	Whisl	Whisl industry portfolio	1
48	Rtail	Rtail industry portfolio	1
49	Meals	Meals industry portfolio	1
50	Banks	Banks industry portfolio	1
51	Insur	Insur industry portfolio	1
52	RIEst	RIEst industry portfolio	1
53	Fin	Fin industry portfolio	1
54	Other	Other industry portfolio	1

Group 4 : Size/Book-to-Market

No.	Code	Description	Tcode
55	ports_2	(small, 2) Portfolio sorted on (size, book-to-market)	1
56	ports_4	(small, 4) Portfolio sorted on (size, book-to-market)	1
57	ports_5	(small, 5) Portfolio sorted on (size, book-to-market)	1
58	ports_6	(small, 6) Portfolio sorted on (size, book-to-market)	1
59	ports_7	(small, 7) Portfolio sorted on (size, book-to-market)	1
60	ports_8	(small, 8) Portfolio sorted on (size, book-to-market)	1
61	ports_9	(small, 9) Portfolio sorted on (size, book-to-market)	1
62	ports_high	(small, high) Portfolio sorted on (size, book-to-market)	1
63	ports_low	(small, low) Portfolio sorted on (size, book-to-market)	1
64	port2_2	(2, 2) Portfolio sorted on (size, book-to-market)	1
65	port2_3	(2, 3) Portfolio sorted on (size, book-to-market)	1
66	port2_4	(2, 4) Portfolio sorted on (size, book-to-market)	1
67	port2_5	(2, 5) Portfolio sorted on (size, book-to-market)	1
68	port2_6	(2, 6) Portfolio sorted on (size, book-to-market)	1
69	port2_7	(2, 7) Portfolio sorted on (size, book-to-market)	1
70	port2_8	(2, 8) Portfolio sorted on (size, book-to-market)	1
71	port2_9	(2, 9) Portfolio sorted on (size, book-to-market)	1
72	port2_high	(2, high) Portfolio sorted on (size, book-to-market)	1
73	port2_low	(2, low) Portfolio sorted on (size, book-to-market)	1
74	port3_2	(3, 2) Portfolio sorted on (size, book-to-market)	1
75	port3_3	(3, 3) Portfolio sorted on (size, book-to-market)	1
76	port3_4	(3, 4) Portfolio sorted on (size, book-to-market)	1
77	port3_5	(3, 5) Portfolio sorted on (size, book-to-market)	1
78	port3_6	(3, 6) Portfolio sorted on (size, book-to-market)	1
79	port3_7	(3, 7) Portfolio sorted on (size, book-to-market)	1
80	port3_8	(3, 8) Portfolio sorted on (size, book-to-market)	1
81	port3_9	(3, 9) Portfolio sorted on (size, book-to-market)	1
82	port3_high	(3, high) Portfolio sorted on (size, book-to-market)	1
83	port3_low	(3, low) Portfolio sorted on (size, book-to-market)	1
84	port4_2	(4, 2) Portfolio sorted on (size, book-to-market)	1
85	port4_3	(4, 3) Portfolio sorted on (size, book-to-market)	1
86	port4_4	(4, 4) Portfolio sorted on (size, book-to-market)	1
87	port4_5	(4, 5) Portfolio sorted on (size, book-to-market)	1
88	port4_6	(4, 6) Portfolio sorted on (size, book-to-market)	1
89	port4_7	(4, 7) Portfolio sorted on (size, book-to-market)	1
90	port4_8	(4, 8) Portfolio sorted on (size, book-to-market)	1
91	port4_9	(4, 2) Portfolio sorted on (size, book-to-market)	1
92	port4_high	(4, high) Portfolio sorted on (size, book-to-market)	1
93	port4_low	(4, low) Portfolio sorted on (size, book-to-market)	1
94	port5_2	(5, 2) Portfolio sorted on (size, book-to-market)	1
95	port5_3	(5, 3) Portfolio sorted on (size, book-to-market)	1
96	port5_4	(5, 4) Portfolio sorted on (size, book-to-market)	1
97	port5_5	(5, 5) Portfolio sorted on (size, book-to-market)	1
98	port5_6	(5, 6) Portfolio sorted on (size, book-to-market)	1
99	port5_7	(5, 7) Portfolio sorted on (size, book-to-market)	1

Group 4 : Size/Book-to-Market (cont.)

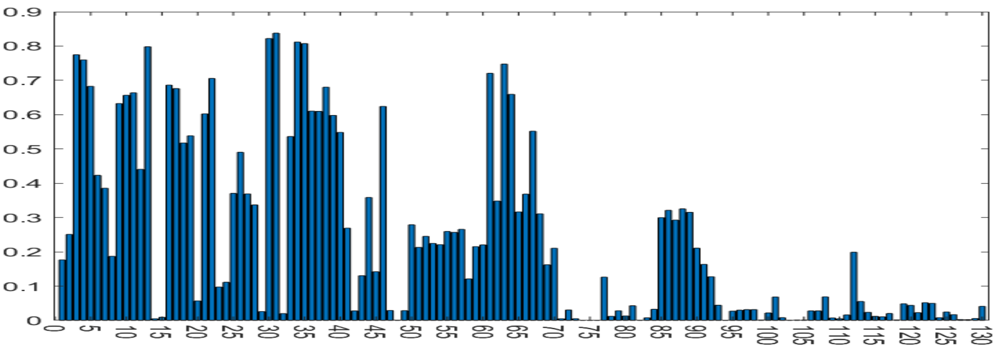
No.	Code	Description	Tcode
100	port5_8	(5, 8) Portfolio sorted on (size, book-to-market)	1
101	port5_9	(5, 9) Portfolio sorted on (size, book-to-market)	1
102	port5_high	(5, high) Portfolio sorted on (size, book-to-market)	1
103	port5_low	(5, low) Portfolio sorted on (size, book-to-market)	1
104	port6_2	(6, 2) Portfolio sorted on (size, book-to-market)	1
105	port6_3	(6, 3) Portfolio sorted on (size, book-to-market)	1
106	port6_4	(6, 4) Portfolio sorted on (size, book-to-market)	1
107	port6_5	(6, 5) Portfolio sorted on (size, book-to-market)	1
108	port6_6	(6, 6) Portfolio sorted on (size, book-to-market)	1
109	port6_7	(6, 7) Portfolio sorted on (size, book-to-market)	1
110	port6_8	(6, 8) Portfolio sorted on (size, book-to-market)	1
111	port6_9	(6, 9) Portfolio sorted on (size, book-to-market)	1
112	port6_high	(6, high) Portfolio sorted on (size, book-to-market)	1
113	port6_low	(6, low) Portfolio sorted on (size, book-to-market)	1
114	port7_2	(7, 2) Portfolio sorted on (size, book-to-market)	1
115	port7_3	(7, 3) Portfolio sorted on (size, book-to-market)	1
116	port7_4	(7, 4) Portfolio sorted on (size, book-to-market)	1
117	port7_5	(7, 5) Portfolio sorted on (size, book-to-market)	1
118	port7_6	(7, 6) Portfolio sorted on (size, book-to-market)	1
119	port7_7	(7, 7) Portfolio sorted on (size, book-to-market)	1
120	port7_8	(7, 8) Portfolio sorted on (size, book-to-market)	1
121	port7_9	(7, 9) Portfolio sorted on (size, book-to-market)	1
122	port7_low	(7, low) Portfolio sorted on (size, book-to-market)	1
123	port8_2	(8, 2) Portfolio sorted on (size, book-to-market)	1
124	port8_3	(8, 3) Portfolio sorted on (size, book-to-market)	1
125	port8_4	(8, 4) Portfolio sorted on (size, book-to-market)	1
126	port8_5	(8, 5) Portfolio sorted on (size, book-to-market)	1
127	port8_6	(8, 6) Portfolio sorted on (size, book-to-market)	1
128	port8_7	(8, 7) Portfolio sorted on (size, book-to-market)	1
129	port8_8	(8, 7) Portfolio sorted on (size, book-to-market)	1
130	port8_9	(8, 9) Portfolio sorted on (size, book-to-market)	1
131	port8_high	(8, high) Portfolio sorted on (size, book-to-market)	1
132	port8_low	(8, low) Portfolio sorted on (size, book-to-market)	1
133	port9_2	(9, 2) Portfolio sorted on (size, book-to-market)	1
134	port9_3	(9, 3) Portfolio sorted on (size, book-to-market)	1
135	port9_4	(9, 4) Portfolio sorted on (size, book-to-market)	1
136	port9_5	(9, 5) Portfolio sorted on (size, book-to-market)	1
137	port9_6	(9, 6) Portfolio sorted on (size, book-to-market)	1
138	port9_7	(9, 7) Portfolio sorted on (size, book-to-market)	1
139	port9_8	(9, 8) Portfolio sorted on (size, book-to-market)	1
140	port9_high	(9, high) Portfolio sorted on (size, book-to-market)	1
141	port9_low	(9, low) Portfolio sorted on (size, book-to-market)	1
142	port10_2	(10, 2) Portfolio sorted on (size, book-to-market)	1
143	port10_3	(10, 3) Portfolio sorted on (size, book-to-market)	1
144	port10_4	(10, 4) Portfolio sorted on (size, book-to-market)	1
145	port10_5	(10, 5) Portfolio sorted on (size, book-to-market)	1
146	port10_6	(10, 6) Portfolio sorted on (size, book-to-market)	1
147	port10_7	(10, 7) Portfolio sorted on (size, book-to-market)	1

2 Plots of the R-squared Regressing the Variables on each Estimated Factors

The R^2 that are related to macroeconomic and financial estimated factors are plotted here.

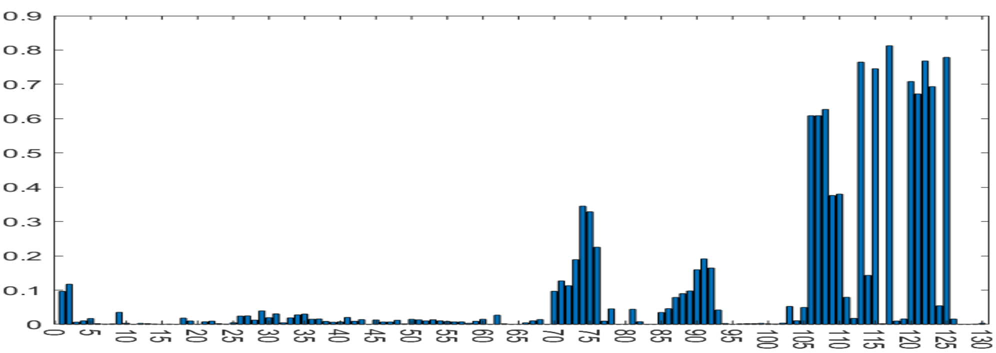
2.1 Macroeconomic Factors

Figure 1: Marginal R^2 for Estimated Macro Factor 1



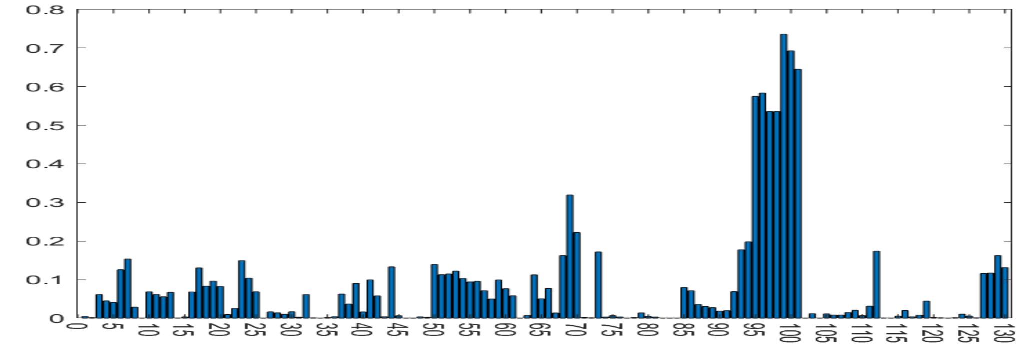
Note: This figure plots the different R^2 obtained after regressing the variables in X_1 on in each estimated factor.

Figure 2: Marginal R^2 for Estimated Macro Factor 2



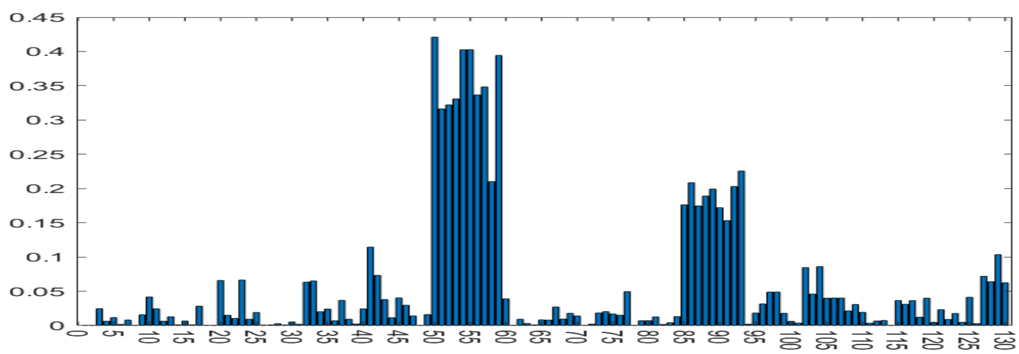
Note: See note for Figure 1.

Figure 3: Marginal R^2 for Estimated Macro Factor 3



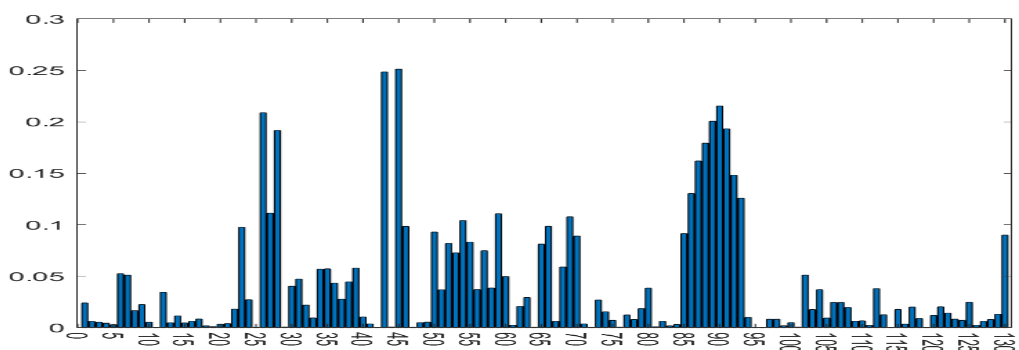
Note: See note for Figure 1.

Figure 4: Marginal R^2 for Estimated Macro Factor 4



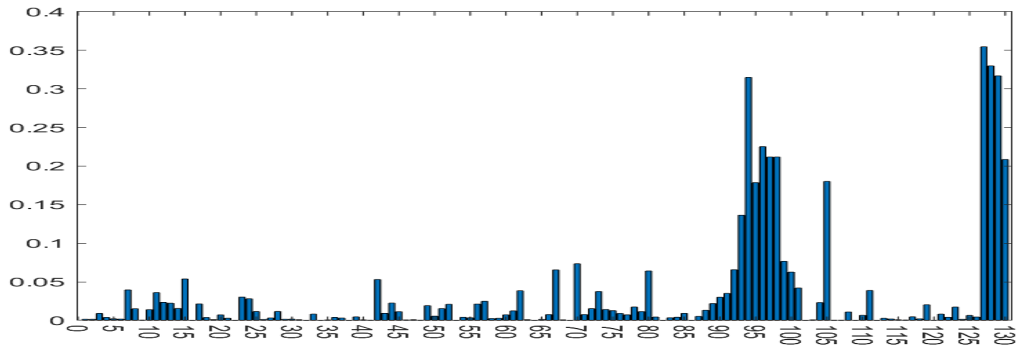
Note: See note for Figure 1.

Figure 5: Marginal R^2 for Estimated Macro Factor 5



Note: See note for Figure 1.

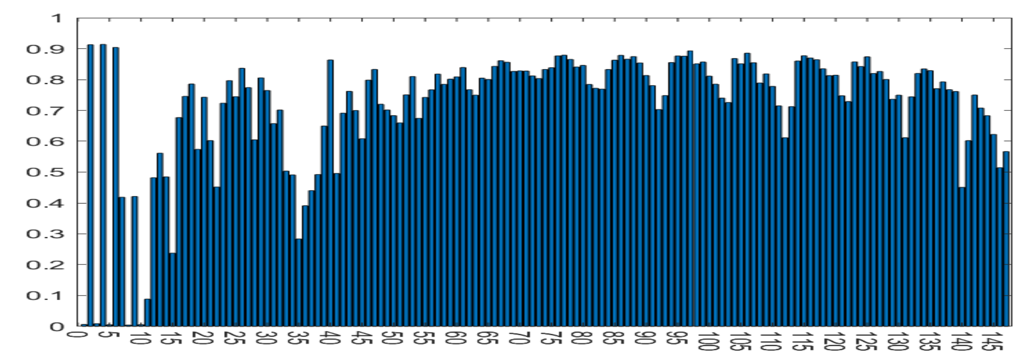
Figure 6: Marginal R^2 for Estimated Macro Factor 6



Note: See note for Figure 1.

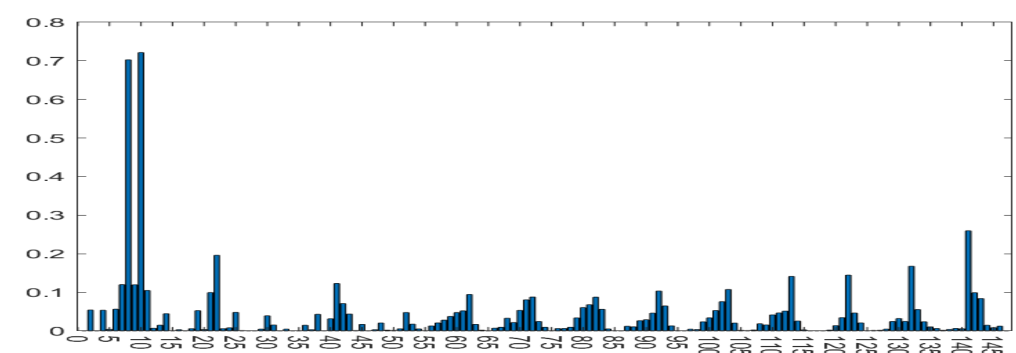
2.2 Financial Factors

Figure 7: Marginal R^2 for Estimated Financial Factor 1



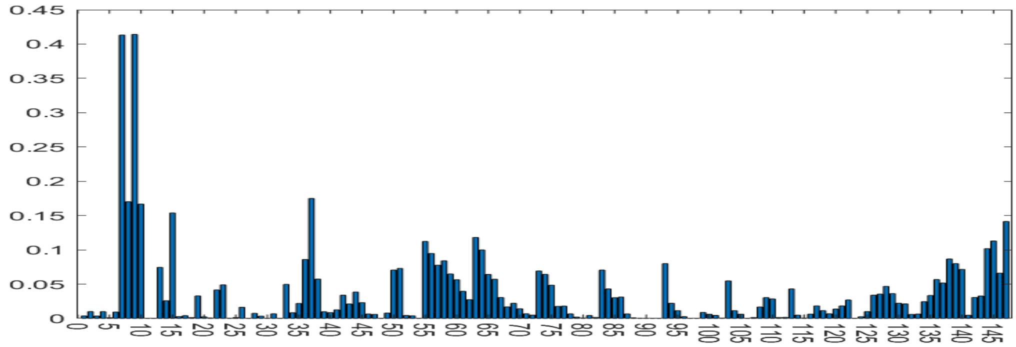
Note: This figure plots the different R^2 obtained after regressing the factor on variables in X_2.

Figure 8: Marginal R^2 for Estimated Financial Factor 2



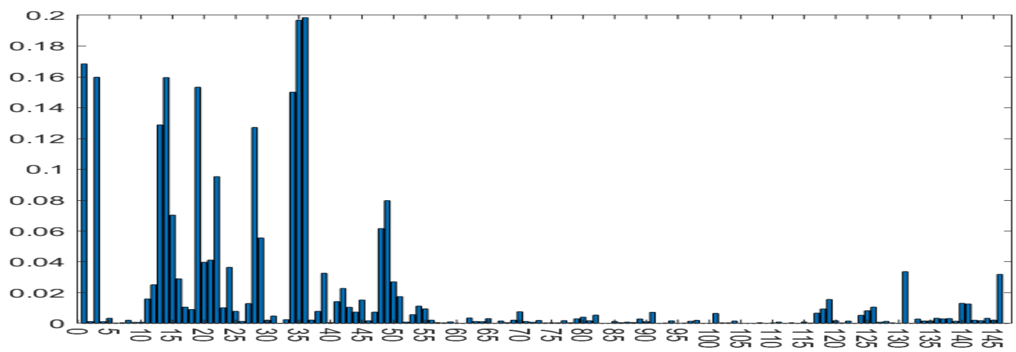
Note: See note for Figure 7.

Figure 9: Marginal R^2 for Estimated Financial Factor 3



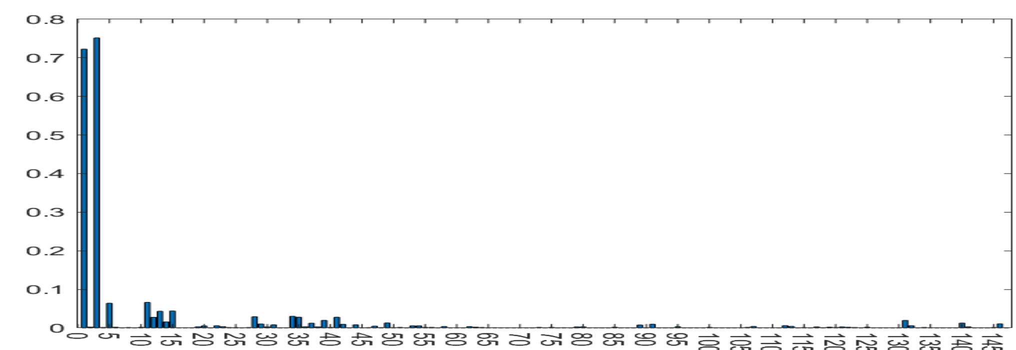
Note: See note for Figure 10.

Figure 10: Marginal R^2 for Estimated Financial Factor 4



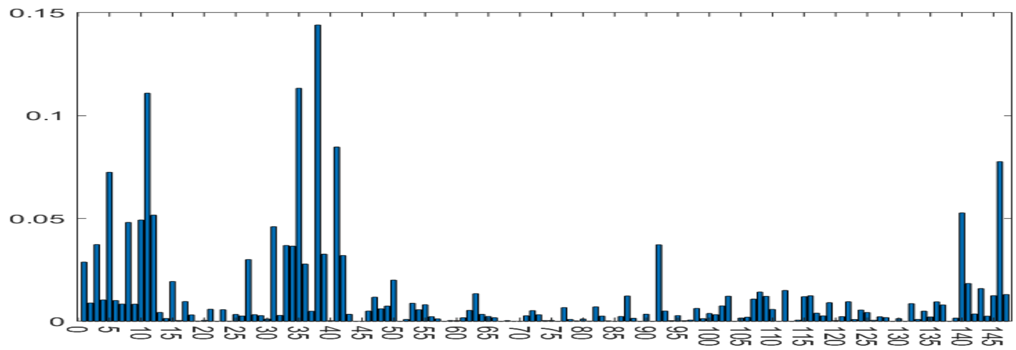
Note: See note for Figure 7.

Figure 11: Marginal R^2 for Estimated Financial Factor 5



Note: See note for Figure 7.

Figure 12: Marginal R^2 for Estimated Financial Factor 6



Note: See note for Figure 7.

References

Cochrane, J. H. and M. Piazzesi (2005). Bond Risk Premia. *American Economic Review 95*(1), 138–160.

Djogbenou, A. A. (2019). Model Selection in Factor-Augmented Regressions with Estimated Factors. Working papers, York University, Department of Economics.

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

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.