Université de Montréal

Bootstrap Methods for Factor Models

par
Antoine A. Djogbenou

Département de sciences économiques
Faculté des arts et sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences économiques

Juillet, 2016

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :
Bootstrap Methods for Factor Models

présentée par :
Antoine A. Djogbenou

a été évaluée par un jury composé des personnes suivantes :

William McCausland,	président-rapporteur
Benoit Perron,	directeur de recherche
Sílvia Gonçalves,	co-directrice de recherche
Marine Carrasco,	membre du jury
Xu Cheng,	examinatrice externe
Jacques Bélair,	vice-doyen, représentant du doyen de la FAS

Thèse acceptée le 28 juillet 2016

Table des matières

Table des figures

Liste des tableaux

à mon feu père, Théophane

Remerciements

J'ai un devoir de reconnaissance à l'endroit de mes directeurs de thèse, Sìlvia Gonçalves et Benoit Perron que j'exprime sincèrement ici. Je leur serai toujours redevable pour leur disponibilité, leur implication et leur soutien, tout au long de ce cheminement de doctorat. Travailler avec eux a été très enrichissant pour moi et j'espère vivement pouvoir continuer à apprendre d'eux dans les années à venir.

Ma reconnaissance infinie va à l'endroit de tout le corps professoral du Département de sciences économiques de l'Université de Montréal pour tout ce que j'ai appris d'eux. Elle va en particulier à Marine Carrasco qui me suit depuis la première présentation orale de cette thèse. Je remercie également le personnel administratif du département pour leur disponibilité sans cesse renouvelée.

J'adresse mes sincères remerciements à Guy Tchuente, Juste Somé, Idrissa Ouili et William Gbohoui pour avoir été présents pour moi dès mes débuts à Montréal et à Mohamed Doukali pour sa grande amitié. Je voudrais dire un grand merci à Désiré Kédagni, pour avoir rêvé ensemble depuis notre premier cycle d'aller aussi loin que possible et relevé ensemble les défis qui se sont présentés à nous. Mes chaleureux remerciements vont à l'endroit de mes nombreux collègues et amis de l'Université de Montréal que je ne pourrai tous citer ici pour les discussions instructives et les bons moments passés ensembles.

À ma mère et mes sœurs pour qui ces cinq années de thèse ont sans doute paru longues, je suis reconnaissant pour le soutien moral continu. A ma merveilleuse épouse Mariam, je témoigne ma profonde gratitude pour sa grande compréhension, sa patience et surtout pour avoir été ce support sur lequel je pouvais m'appuyer quand tout semblait s'écrouler autour de moi.

Ce cheminement n'aurait pas été possible sans le soutien financier du Département de Sciences Économiques de l'Université de Montréal, le Centre Interuniversitaire de Recherche en Économie Quantitative (CIREQ) et la Faculté des Études Supérieures et Postdoctorales que je remercie ici.

Résumé

Cette thèse développe des méthodes bootstrap pour les modèles à facteurs qui sont couramment utilisés pour générer des prévisions depuis l'article pionnier de Stock et Watson (2002) sur les indices de diffusion. Ces modèles tolèrent l'inclusion d'un grand nombre de variables macroéconomiques et financières comme prédicteurs, une caractéristique utile pour inclure diverses informations disponibles aux agents économiques. Ma thèse propose donc des outils économétriques qui améliorent l'inférence dans les modèles à facteurs utilisant des facteurs latents extraits d'un large panel de prédicteurs observés. Il est subdivisé en trois chapitres complémentaires dont les deux premiers en collaboration avec Sílvia Gonçalves et Benoit Perron.

Dans le premier article, nous étudions comment les méthodes bootstrap peuvent être utilisées pour faire de l'inférence dans les modèles de prévision pour un horizon de h périodes dans le futur. Pour ce faire, il examine l'inférence bootstrap dans un contexte de régression augmentée de facteurs où les erreurs pourraient être autocorrélées. Il généralise les résultats de Gonçalves et Perron (2014) et propose puis justifie deux approches basées sur les résidus : le block wild bootstrap et le dependent wild bootstrap. Nos simulations montrent une amélioration des taux de couverture des intervalles de confiance des coefficients estimés en utilisant ces approches comparativement à la théorie asymptotique et au wild bootstrap en présence de corrélation sérielle dans les erreurs de régression.

Le deuxième chapitre propose des méthodes bootstrap pour la construction des intervalles de prévision permettant de relâcher l'hypothèse de normalité des innovations. Nous y proposons des intervalles de prédiction bootstrap pour une observation h périodes dans le futur et sa moyenne conditionnelle. Nous supposons que ces prévisions sont faites en utilisant un ensemble de facteurs extraits d'un large panel de variables. Parce que nous traitons ces facteurs comme latents, nos prévisions dépendent à la fois des facteurs estimés et les coefficients de régression estimés. Sous des conditions de régularité, Bai et Ng (2006) ont proposé la construction d'intervalles asymptotiques sous l'hypothèse de Gaussianité des innovations. Le bootstrap nous permet de relâcher cette hypothèse et de construire des intervalles de prédiction valides sous des hypothèses plus générales. En outre, même en supposant la Gaussianité, le bootstrap conduit à des intervalles plus précis dans les cas où la dimension transversale est relativement faible car il prend en considération le biais de l'estimateur des moindres carrés ordinaires comme le montre une étude récente de Gonçalves et Perron (2014).

Dans le troisième chapitre, nous suggérons des procédures de sélection convergentes pour les regressions augmentées de facteurs en échantillons finis. Nous démontrons premièrement que la méthode de validation croisée usuelle est non-convergente mais que sa généralisation, la validation croisée «leave-d-out» sélectionne le plus petit ensemble de facteurs estimés pour l'espace généré par les vraies facteurs. Le deuxième critère dont nous montrons également la validité généralise l'approximation bootstrap de Shao (1996) pour les regressions augmentées

de facteurs. Les simulations montrent une amélioration de la probabilité de sélectionner parcimonieusement les facteurs estimés comparativement aux méthodes de sélection disponibles. L'application empirique revisite la relation entre les facteurs macroéconomiques et financiers, et l'excès de rendement sur le marché boursier américain. Parmi les facteurs estimés à partir d'un large panel de données macroéconomiques et financières des États Unis, les facteurs fortement correlés aux écarts de taux d'intérêt et les facteurs de Fama-French ont un bon pouvoir prédictif pour les excès de rendement.
Mots-clés : Modèles à facteurs, correlation sérielle, prévision, moyenne conditionnelle, sélection de modèle, validation croisée, bootstrap, inflation, excès de rendement boursier, É.U.

Abstract

This thesis develops bootstrap methods for factor models which are now widely used for generating forecasts since the seminal paper of Stock and Watson (2002) on diffusion indices. These models allow the inclusion of a large set of macroeconomic and financial variables as predictors, useful to span various information related to economic agents. My thesis develops econometric tools that improves inference in factor-augmented regression models driven by few unobservable factors estimated from a large panel of observed predictors. It is subdivided into three complementary chapters. The two first chapters are joint papers with Sílvia Gonçalves and Benoit Perron.

In the first chapter, we study how bootstrap methods can be used to make inference in h-step forecasting models which generally involve serially correlated errors. It thus considers bootstrap inference in a factor-augmented regression context where the errors could potentially be serially correlated. This generalizes results in Gonçalves and Perron (2013) and makes the bootstrap applicable to forecasting contexts where the forecast horizon is greater than one. We propose and justify two residual-based approaches, a block wild bootstrap (BWB) and a dependent wild bootstrap (DWB). Our simulations document improvement in coverage rates of confidence intervals for the coefficients when using BWB or DWB relative to both asymptotic theory and the wild bootstrap when serial correlation is present in the regression errors.

The second chapter provides bootstrap methods for prediction intervals which allow relaxing the normality distribution assumption on innovations. We propose bootstrap prediction intervals for an observation h periods into the future and its conditional mean. We assume that these forecasts are made using a set of factors extracted from a large panel of variables. Because we treat these factors as latent, our forecasts depend both on estimated factors and estimated regression coefficients. Under regularity conditions, Bai and Ng (2006) proposed the construction of asymptotic intervals under Gaussianity of the innovations. The bootstrap allows us to relax this assumption and to construct valid prediction intervals under more general conditions. Moreover, even under Gaussianity, the bootstrap leads to more accurate intervals in cases where the cross-sectional dimension is relatively small as it reduces the bias of the ordinary least squares estimator as shown in a recent paper by Gonçalves and Perron (2014).

The third chapter proposes two consistent model selection procedures for factor-augmented regressions in finite samples. We first demonstrate that the usual cross-validation is inconsistent, but that a generalization, leave-d-out cross-validation, selects the smallest basis of estimated factors for the space spanned by the true factors. The second proposed criterion is a generalization of the bootstrap approximation of the squared error of prediction of Shao (1996) to factor-augmented regressions which we also show is consistent. Simulation evidence documents improvements in the probability of selecting the smallest set of estimated factors than the usually available methods. An illustrative empirical application that analyzes the relationship

between expected stock returns and macroeconomic and financial 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 strong predictive power for excess returns.
Keywords : Factor model, serial correlation, forecast, conditional mean, model selection, cross-validation, bootstrap, excess returns, U.S.

Introduction Générale

Dans les dernières décennies, nous avons noté une disponibilité croissante de données économiques. Leur utilisation pour générer des prévisions a connu un regain d'intérêt depuis le travail pionnier de Stock et Watson (2002) sur les modèles à facteurs augmentés. Ces modèles assument que la variable d'intérêt par exemples l'inflation ou l'excès de rendement boursier dépendent non seulement de variables observées mais aussi de facteurs inobservés résumant l'information d'un grand nombre de variables.

En pratique, parce que les facteurs sont latents, ils sont remplacés par leur version estimée généralement par la méthode des composantes principales. Sous des conditions de régularité, Bai et Ng (2006) montrent que les facteurs extraits peuvent être traités comme si ils éntait observés lorsque la racine carrée de la dimension temporelle sur le nombre de série tend vers zéro. En examinant les propriétés asymptotiques, Gonçalves et Perron (2014) démontrent la présence d'un biais dans la distribution asymptotique de l'estimateur obtenu par la regression augmentée de facteurs. De surcroît, ils suggèrent une méthode de bootstrap en deux étapes permettant de capturer ce biais. Cette méthode du wild bootstrap détruit toute dépendance entre les observations. Ainsi, elle n'est valide que lorsque l'horizon de prévision est 1 car lorsque l'horizon de prévision est supérieure à une période, les innovations sont généralement dépendantes. Ce qui rend invalide le wild bootstrap. Dans le premier chapitre, nous justifions théoriquement la validité du block wild bootstrap et le dépendent wild bootstrap. Ces deux méthodes basées sur les résidus en dimension temporelle se rejoignent sur le fait qu'elles préservent la dépendance dans la variance asymptotique de l'estimateur. Toutefois, la première préserve cette dépendance en considérant k blocs de résidus multipliés chacun par une même variable externe. La seconde approche quant à elle lisse les variables externes au delà des blocs.

Dans le deuxième chapitre, nous justifions la validité du bootstrap pour construire les intervalles de prédiction pour une réalisation future de la variable dépendante ou sa moyenne conditionnelle à l'information disponible. Nos résultats, permettent contrairement à l'approache asymptotique usuelle, de relâcher l'hypothèse de normalité des innovations futures. En appliquant notre demarche à la prévision des changements de l'inflation avec des données trimestrielles de l'économie américaine couvrant la période 1973 à 2014, nos intervalles de prédiction incluent la forte déflation observée pendant la crise financière de 2008 et celle du dernier trimestre de 2011.

Nous complètons notre analyse dans le troixième chapitre par l'examen du choix des facteurs estimés à inclure dans l'équation de prévision. En effet, les facteurs latents F^0 importants pour prédire la variable dépendente ne sont pas nécessairement tous ceux (F) qui résument l'information dans le grand nombre de prédicteurs disponibles X. Nous nous fixons comme objectif de détecter le plus petit ensemble de regresseurs générés capable de recouvrir l'information dans F^0. Bien que beaucoup de travaux se sont intéressé au choix des facteurs estimés reflétant

le mouvement commun dans X, peu se sont penchés sur l'identification de ces derniers. Nous explorons comment la méthode de validation croisée peut être utilisée dans notre contexte de régresseurs générés. Nous montrons que la méthode de sélection de validation croisée usuelle n'est valide que lorsqu'un seul ensemble de facteurs estimés est correct. Pour remédier à cette situation, nous justifions la validité du «leave-d-out cross-validation» avec un d convenablement choisi. Nous proposons également une approche de bootstrap convergente qui contrairement à la méthode de validation croisée leave-d-out évalue l'habileté de prédiction des modèles candidats avec un estimateur se basant sur toutes les observations. En considérant un ensemble de 277 variables macroéconomiques et financières, nous avons étudié les déterminants de l'excès de rendement boursier sur le marché américain en revisitant le travail de Ludvigson et Ng (2007). Les facteurs fortement correlés aux écarts entre les taux d'intérêts et le taux directeur et les facteurs de Fama-French ont un fort pouvoir prédictif pour l'excès de rendement boursier. Nos résultats montrent que les approches suggérées protègent contre la sélection d'un nombre inapproprié de facteurs estimés.

Chapitre 1

Bootstrap Inference in Regressions with Estimated Factors and Serial Correlation

1.1 Introduction

Factor-augmented regressions have become quite popular in research in finance and economics since the seminal paper of Stock and Watson (2002). They are often used in a forecasting context as they allow to summarize a large number of predictors with a small number of indexes.

Because these indexes are treated as latent factors in an approximate factor model, the estimated regression contains estimated regressors which poses challenges for inference. Under regularity conditions, Bai and Ng (2006) derived the asymptotic distribution of regression estimates. One of the key conditions used in their work is that $\sqrt{T}/N \to 0$. In that case, the error in estimating the factors can be neglected and inference can proceed as if they were observed.

Gonçalves and Perron (2014) (GP (2014) thereafter) showed that the finite sample properties of the asymptotic approach of Bai and Ng (2006) can be poor, especially if N is not sufficiently large relative to T. In particular, estimation of factors leads to an asymptotic bias term in the OLS estimator if $\sqrt{T}/N \to c$ and $c \neq 0$. They provided a set of high level conditions under which any residual-based bootstrap method is valid in this context and showed that a bootstrap algorithm based on the wild bootstrap removes this bias and outperforms the asymptotic approach of Bai and Ng (2006) in simulation experiments. This wild bootstrap algorithm is only valid when the forecasting horizon is one because it does not reproduce serial correlation. In general, when the forecasting horizon is larger than one and the model is correctly specified, the residuals in the factor-augmented regression will follow a moving average process (Diebold (2007), pp. 256-257).

In this paper, we extend the work of Bai and Ng (2006) and GP (2014) by considering errors

This chapter is a joint paper with Sílvia Gonçalves and Benoit Perron. The authors are grateful for comments from seminar participants at the Toulouse School of Economics, Pompeu Fabra University and Duke University, as well as from participants at the Workshop on Bootstrap Methods for Time Series, Copenhagen, Denmark, September 2013, and the MAESG conference in Emory, Atlanta, November 2013. Gonçalves acknowledges financial support from the NSERC and MITACS whereas Perron acknowledges financial support from the SSHRC and MITACS. We acknowledge that this chapter has been published by Journal of Time Series Analysis and agree with the publication disclosure.

that are serially correlated. Bai and Ng effectively ruled out possible serial correlation since their estimator of the asymptotic variance of the scaled average of the scores is only consistent with heteroskedasticity. We begin by providing an asymptotic theory under general assumptions on the serial correlation of the error term (of the strong mixing type) and proposing a consistent estimator of the covariance matrix in that case. As in GP (2014), we allow $\sqrt{T}/N \to c > 0$ so that a bias term appears in the asymptotic distribution. Secondly, we propose two residual-based bootstrap schemes and show that they provide valid inference in this context. The first scheme which we call the block wild bootstrap (BWB) was proposed by Yeh (1998) for a linear regression with fixed scalar regressor and strong mixing errors. It is implemented by separating the residuals into non-overlapping blocks of observations and multiplying the elements of each block by the same realization of an external variable. The fact that each element in a block is multiplied by the same external draw generates correlation among the elements within a block but enforces independence across blocks. The second scheme we consider is the dependent wild bootstrap (DWB) originally proposed by Shao (2010) in the context of the smooth function model with time series observations. The DWB differs from the BWB by smoothing the external draws across blocks. Our main contribution is to show that these two methods are valid in the context of a factor augmented regression model with estimated factors and serially correlated errors, characterized by a strong mixing assumption.

The remainder of the paper is organized as follows. Section 1.2 introduces our assumptions, provides the asymptotic distribution of the OLS estimator, and proposes a consistent estimator of the covariance matrix. Section 1.3 considers bootstrap inference using our two proposed algorithms. Section 1.4 presents our simulation experiments, and Section 1.5 concludes. Mathematical proofs appear in the Appendix 0.1.

1.2 Assumptions and asymptotic results

We consider the following standard factor-augmented regression model,

$$y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h}, \quad t = 1, \dots, T-h,$$

where y_{t+h} denotes the variable of interest, for example GDP growth or inflation, with h the forecast horizon. The $r \times 1$ vector F_t consists of *latent* factors which help forecast y_{t+h}. These are thought as common latent factors in a panel factor model given by

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

where λ_i, $i = 1, \dots, N$, are the $r \times 1$ factor loadings and e_{it} is an idiosyncratic error term, $i = 1, \dots, N$, $t = 1, \dots, T$. The vector W_t contains a smaller set of other observed regressors (including for instance a constant and lags of y_t). We will denote the set of regressors as $z_t = (F_t', W_t')'$, $t = 1, \dots, T$.

We impose the following assumptions. Throughout, $\|M\| = (trace\,(M'M))^{1/2}$ denotes the Euclidean norm, $M > 0$ denotes positive definiteness for a square matrix, and C represents a generic finite constant.

Assumption 1 (factor model)

a) $E\|F_t\|^4 \leq C$ and $\Sigma_F = \lim_{T\to\infty} E\left(T^{-1}F'F\right) = \lim_{T\to\infty} E\left(T^{-1}\sum_{t=1}^{T} F_t F_t'\right) > 0$.

b) $\|\lambda_i\| \leq C$ if λ_i are deterministic, or $E\|\lambda_i\| \leq C$ if not, and $N^{-1}\Lambda'\Lambda = N^{-1}\sum_{i=1}^{N}\lambda_i\lambda_i' \to^P \Sigma_\Lambda > 0$.

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

Assumption 2 (Idiosyncratic errors)

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

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

c) $E\left|N^{-1/2}\sum_{i=1}^{N}(e_{it}e_{is} - E(e_{it}e_{is}))\right|^4 \leq C$ for all (t,s).

Assumption 3 (Moments and weak dependence among $\{z_t\}$, $\{\lambda_i\}$, and $\{e_{it}\}$)

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

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

c) $E\left\|(NT)^{-1/2}\sum_{t=1}^{T}z_t e_t'\Lambda\right\|^2 \leq C$ where $E(z_t\lambda_i' e_{it}) = 0$ for all (i,t).

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

e) As $N,T \to \infty$, $(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j' e_{it}e_{jt} - \Gamma \to^P 0$, where $\Gamma \equiv \lim_{N,\,T\to\infty}T^{-1}\sum_{t=1}^{T}\Gamma_t > 0$, and $\Gamma_t \equiv Var\left(N^{-1/2}\sum_{i=1}^{N}\lambda_i e_{it}\right)$.

Assumption 4 (Weak dependence betwen ε_{t+h} and e_{it})

a) For each t and $h \geq 0$, $E\left|(NT)^{-1/2}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_{s+h}(e_{it}e_{is} - E(e_{it}e_{is}))\right| \leq C$.

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

Assumption 5 (Moments and dependence of the score vector) For some $r > 2$,

a) $E(z_t\varepsilon_{t+h}) = 0$, $E\|z_t\|^{2r} < C$ and $E\left(\varepsilon_{t+h}^{2r}\right) < C$.

b) $\{(z_t', \varepsilon_{t+h})\}$ is a fourth order stationary strong mixing sequence of size $-\frac{2r}{r-2}$.

c) $\Sigma_{zz} = \lim_{T\to\infty}E\left(T^{-1}\sum_{t=1}^{T}z_t z_t'\right) > 0$.

d) $\Omega = \lim_{T\to\infty}Var\left(T^{-1/2}\sum_{t=1}^{T-h}z_t\varepsilon_{t+h}\right) > 0$.

Assumptions 1-4 are identical to those of GP (2014) whereas Assumption 5 contains the fundamental difference. We replace the high level central limit theorem assumption of GP (2014, cf. Assumption 5(c)) by more primitive assumptions that allow us to show consistency of the bootstrap in this context. Specifically, we impose a strong mixing assumption on $(z_t', \varepsilon_{t+h})$ and require the existence of slightly more than four finite moments for these random variables (which is a strengthening of the moment conditions used by GP (2014)). Under these assumptions, we can show that a central limit theorem holds for the regression scores (using the latent factors), thus verifying Assumption 5 of GP (2014). Our strong mixing assumption allows for quite general serial dependence, including the class of stationary ARMA processes. This is the case even when $h = 1$, where the condition $E(z_t\varepsilon_{t+h}) = 0$ imposes further restrictions on the form of serial correlation in ε_t when z_t contains a lagged dependent variable (e.g. it rules out an AR(1) model for ε_t) but does not eliminate it.

To estimate the factor-augmented regression, it is necessary to use an estimator of the latent factors F_t. It is well known that factor models suffer from a lack of identification. As shown by Bai (2003), the principal component $\widetilde{F}_t$ is only consistent for a rotation of F_t, denoted by HF_t, where H denotes the associated rotation matrix. Bai showed that the rotation matrix H is given by

$$H = \tilde{V}^{-1}\frac{\tilde{F}'F}{T}\frac{\Lambda'\Lambda}{N}, \tag{1.1}$$

where $\tilde{V}$ is a $r \times r$ diagonal matrix with the r largest eigenvalues of XX'/NT, in decreasing order on the diagonal.

It is useful to rewrite the model as

$$y_{t+h} = \hat{z}'\delta + \alpha' H^{-1}\left(HF_t - \tilde{F}_t\right) + \varepsilon_{t+h},$$

where $\delta' = (\alpha' H^{-1} \ \beta')$ and $\hat{z}_t' = \left(\tilde{F}_t', W_t'\right)$. The consequence of the lack of identification of the factor model is that the coefficients associated with the estimated factors are rotated versions of those associated with the true latent factors. Bai and Ng (2013) provide three sets of conditions under which $H_0 = p\lim H = diag\,(\pm 1)$. Under those conditions, α will be identified up to sign.

The OLS estimator from regressing y_{t+h} on $\tilde{F}_t$ and W_t is given by

$$\hat{\delta} = \left(\hat{\alpha}', \hat{\beta}'\right)' = \left(\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\right)^{-1}\sum_{t=1}^{T-h}\hat{z}_t y_{t+h},$$

and it will be such that $\hat{\delta} \to^P \delta \equiv \left(\ \alpha' H^{-1} \quad \beta'\ \right)'$ under our assumptions. We denote $\Phi_0 \equiv diag\,(H_0, I)$. The following theorem provides the asymptotic distribution of the OLS estimator. The proof is in the Appendix.

Theorem 1. *Under Assumptions 1-5, if* $\frac{\sqrt{T}}{N} \to c < \infty$, *as* $N, T \to \infty$, *then*

$$\sqrt{T}\left(\hat{\delta} - \delta\right) \to^d N\left(-c\Delta_\delta, \Sigma_\delta\right),$$

with $\Sigma_\delta = \Phi_0'^{-1}\Sigma_{zz}^{-1}\Omega\Sigma_{zz}^{-1}\Phi_0^{-1}$, *and*

$$\Delta_\delta = \left(\Phi_0\Sigma_{zz}\Phi_0'\right)^{-1}\begin{pmatrix} \Sigma_{\tilde{F}} + V\Sigma_{\tilde{F}}V \\ \Sigma_{W\tilde{F}}V\Sigma_{\tilde{F}}V^{-1}\end{pmatrix}\left(H_0^{-1}\right)'\alpha$$

where $\Sigma_{W\tilde{F}} = p\lim\left(\frac{W'\tilde{F}}{T}\right)$, $\Sigma_{\tilde{F}} = V^{-1}Q\Gamma Q'V^{-1}$, $Q = p\lim\frac{\tilde{F}'F}{T}$, *and* $V = p\lim\tilde{V}$.

Theorem 1 follows from Theorem 2.1 of GP (2014), where the asymptotic normality of the OLS estimator was obtained under a high level CLT assumption on the regression scores. Instead, here we allow dependence of unknown form by assuming a mixing condition on the regressors and on the regression errors. This primitive condition will be useful to establish the consistency of the BWB and DWB in Section 1.3, as well as the consistency of a HAC estimator of Ω, as we prove next. Note that under this mixing condition, Ω is not necessarily of the form $\Omega = E\left(z_t z_t'\varepsilon_{t+h}^2\right)$ assumed by Bai and Ng (2006).

To carry out inference or construct prediction intervals, a consistent covariance estimator of Σ_δ is required. As we allow for serial correlation in the score, a HAC estimator of Σ_δ is appropriate,

$$\hat{\Sigma}_\delta = \left(T^{-1}\hat{z}'\hat{z}\right)^{-1}\hat{\Omega}\left(T^{-1}\hat{z}'\hat{z}\right)^{-1}$$

with

$$\hat{\Omega} = \hat{\Xi}_0 + \sum_{j=1}^{T-h-1} k\left(\frac{j}{M_T}\right)\left[\hat{\Xi}_j + \hat{\Xi}_j'\right],$$

where $\hat{\Xi}_j = \frac{1}{T}\sum_{t=1}^{T-h-j}\widehat{z}_t\widehat{z}_{t+j}'\widehat{\varepsilon}_{t+h}\widehat{\varepsilon}_{t+h+j}$ is the autocovariance matrix of the scores, $k(\cdot)$ is a kernel function, and M_T is a bandwidth.

To prove consistency of this estimator, restrictions must be placed on the kernel function $k(\cdot)$ and bandwidth M_T. We will consider kernels in the family $\mathcal{K}_1$ as in Andrews and Monahan (1992) :

$$\mathcal{K}_1 = \left\{\begin{array}{l} k(\cdot):\mathbb{R}\to[-1,1],\ k(0)=1,\ k(x)=k(-x) \text{ for } x\in\mathbb{R},\ \int_{-\infty}^{+\infty}|k(x)|\,dx<\infty, \\ k(\cdot) \text{ is continuous at 0 and at all but a finite number of points} \end{array}\right\}.$$

In addition, we must strengthen Assumptions 3 and 5. Specifically, we require :

Assumption 3'

d) $E\left(T^{-1}\sum_{t=1}^{T}\left\|N^{-1/2}\sum_{i=1}^{N}\lambda_i e_t\right\|^4\right) \le C$ where $E(\lambda_i e_{it}) = 0$ for all $(i,\ t)$.

Assumption 5' For some $r > 2$,

a) $E(z_t\varepsilon_{t+h}) = 0$, $E\|z_t\|^{4r} < C$ and $E\left(\varepsilon_{t+h}^{4r}\right) < C$.

b) $\{(z_t',\ \varepsilon_{t+h})\}$ is a fourth order stationary strong mixing sequence of size $-\frac{3r}{r-2}$.

The other parts of these two assumptions remain as before. By strengthening Assumption 5.a) by Assumption 5'.a) we have that $E\|z_t\varepsilon_{t+h}\|^{2r} < C$, which is sufficient for the proof of our next result. Assumption 5' is analogous to the assumptions made in Andrews (1991, Lemma 1) to prove consistency of the HAC estimator.

Lemma 2. *Suppose that Assumptions 1-5, with Assumptions 3 and 5 strengthened by Assumptions 3' and 5' respectively, hold. Suppose further that* $k(\cdot)$ *belongs to the set* $\mathcal{K}_1$ *and that* $M_T \to \infty$ *as* $T \to \infty$ *such that* $\frac{M_T}{T}^2 \to 0$*. If* $\frac{\sqrt{T}}{N} \to c < \infty$ *as* $N, T \to \infty$*, then* $\widehat{\Sigma}_\delta \to^P \Sigma_\delta$.

This lemma shows that a HAC covariance estimator is consistent for Σ_δ despite the presence of estimated regressors. This implies that, as in Bai and Ng (2006), asymptotic inference can be carried out as if the factors were observed if $\sqrt{T}/N \to 0$ since in that case, the asymptotic distribution of $\sqrt{T}\left(\hat{\delta}-\delta\right)$ is centered at 0. If $\sqrt{T}/N \to c > 0$, Lemma 2.1 shows that HAC estimation is still possible, but inference is complicated by the need to account for the bias term in the asymptotic distribution. As in GP (2014), we consider the bootstrap to accomplish this in the next section.

1.3 Bootstrap inference

1.3.1 General residual-based bootstrap : review

In this section, we consider bootstrap inference on the coefficients of the factor-augmented regression. The proposed bootstrap scheme resamples the idiosyncratic and regression residuals separately and is similar to the one in GP (2014) with the difference that in the second step, residuals $\{\hat{\varepsilon}_{t+h}\}$ are resampled by either the block wild bootstrap or the dependent wild bootstrap. As usual, we will denote with asterisks quantities in the bootstrap world. We will also denote by E^* (and Var^*) the expectation (and variance) under the bootstrap measure P^*.

Bootstrap algorithm

1. For $t = 1, \ldots, T$, generate $X_t^* = \tilde{\Lambda}\tilde{F}_t + e_t^*$, where $\{e_{it}^*\}$ is a resampled version of $\left\{\tilde{e}_{it} = X_{it} - \tilde{\lambda}_i'\tilde{F}_t\right\}$. In this step, we use the wild bootstrap and set

$$e_{it}^* = \tilde{e}_{it} \cdot \eta_{it}, i = 1, \ldots, N, t = 1, \ldots, T$$

   where η_{it} is an i.i.d. draw (over i and t) from an external random variable with mean 0 and variance 1.

2. Estimate the bootstrap factors $\left\{\tilde{F}_t^* : t = 1, \ldots, T\right\}$ by principal components using X^*.

3. For $t = 1, \ldots, T - h$, generate $y_{t+h}^* = \hat{\alpha}'\tilde{F}_t + \hat{\beta}'W_t + \varepsilon_{t+h}^*$, where the error term ε_{t+h}^* is a resampled version of $\hat{\varepsilon}_{t+h}$. In this step, we will use either the block wild bootstrap or the dependent wild bootstrap as detailed below to accommodate serial correlation in ε_{t+h}. [1]

4. Regress y_{t+h}^* generated in step 3 on the bootstrap estimated factors $\tilde{F}_t^*$ obtained in step 2 and on the observed regressors W_t and obtain the OLS estimator $\hat{\delta}^*$,

$$\hat{\delta}^* = \left(\sum_{t=1}^{T-h} \hat{z}_t^* \hat{z}_t^{*\prime}\right)^{-1} \sum_{t=1}^{T-h} \hat{z}_t^* y_{t+h}^*, \quad \text{where} \quad \hat{z}_t^* = \left(\tilde{F}_t^{*\prime}, W_t'\right)'.$$

5. Repeat steps 1-4 B times.

As in the sample, the principal component estimator in the bootstrap consistently estimates the space of factors only. The specific rotation that is estimated is given by the bootstrap analogue of the H matrix,

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

where $\tilde{V}^*$ is the $r \times r$ diagonal matrix containing on the main diagonal the r largest eigenvalues of $X^*X^{*\prime}/NT$, in decreasing order. Note that contrary to H, which depends on unknown population parameters, H^* is fully observed. Using the results in Bai and Ng (2013) , H^* converges asymptotically to a diagonal matrix with $+1$ or -1 on the main diagonal, see GP (2014) for more details.

[1]When W_t includes lagged values of the dependent variable, it is also possible to generate y_{t+h}^* recursively as in $y_{t+h}^* = \widehat{\alpha}'\widetilde{F}_t + \widehat{\beta}y_t^* + \varepsilon_{t+h}^*$, $t = 1, \ldots, T - h$. Simulation results did not show any noticeable improvements from doing this, so we concentrate on a fixed-design scheme which allows for a unified treatment of W_t.

The consequence of this lack of identification is that the bootstrap OLS estimator estimates $\delta^* = \left(\hat{\alpha}' H^{*-1} \; \hat{\beta}'\right)' = \left(\Phi^{*-1}\right)' \hat{\delta}$ which is different from $\hat{\delta}$. GP (2014) suggested using a rotated version of this estimator, $\tilde{\delta}^* = \Phi^{*\prime}\hat{\delta}^*$ for bootstrap inference, and we will do the same here.

The next assumption is a modified version of Assumptions 6-8 in GP (2014) applied to our context.

Assumption 6

a) λ_i are either deterministic such that $\|\lambda_i\| \leq C < \infty$, or stochastic such that $E\|\lambda_i\|^{12} \leq C < \infty$ for all i, and $E\|F_t\|^{12} \leq C < \infty$.

b) $E|e_{it}|^{12} \leq C < \infty$, for all (i,t) and $E(e_{it}e_{js}) = 0$, if $i \neq j$.

c) z_t and ε_{t+h} are independent of e_{is} for all (i, t, s).

Assumption 6.b) excludes cross-sectional dependence among idiosyncratic errors as in Assumption 8 of GP (2014). This is required because we use the wild bootstrap in step 1 of the bootstrap algorithm which destroys such dependence. We could relax this assumption if we were willing to assume that $\sqrt{T}/N \to 0$ as in Bai and Ng (2006). In that case, the bias term of the OLS estimator is 0, and this is the only quantity that depends on the properties of the idiosyncratic errors asymptotically. In that situation, factor estimation error does not matter asymptotically, and the key condition for bootstrap validity is to replicate the properties of the regression errors ε_{t+h}, as we are doing here with our two proposed blocking methods.

We now consider the two bootstrap schemes to generate ε^*_{t+h} in step 3 of this algorithm.

1.3.2 Block wild bootstrap

The first scheme we consider is the block wild bootstrap (BWB) first proposed by Yeh (1998) and analyzed in other contexts by Shao (2011) and Urbain and Smeekes (2013).

First, we form non-overlapping blocks of size b_T of consecutive residuals. For simplicity, we assume that $(T-h)/b_T = k_T$, where k_T is an integer and denotes the number of blocks of size b_T. For $l = 1, \ldots, b_T$ and $j = 1, \ldots, k_T$, we let

$$y^*_{(j-1)b_T+l+h} = \hat{\alpha}'\tilde{F}_{(j-1)b_T+l} + \hat{\beta}'W_{(j-1)b_T+l} + \varepsilon^*_{(j-1)b_T+l+h}, \tag{1.2}$$

where

$$\varepsilon^*_{(j-1)b_T+l+h} = \hat{\varepsilon}_{(j-1)b_T+l+h} \cdot \nu_j$$

and ν_j is an external random variable with mean 0, variance 1, and independent and identically distributed across blocks. In other words, the bootstrap data is obtained by multiplying each residual by an external variable that is the same for all observations within a block. The next theorem shows the consistency of the bootstrap based on the rotated version of the OLS estimator, $\Phi^{*\prime}\hat{\delta}^*$.

Theorem 3. *Under the same assumptions as in Lemma 2, assuming* $E^*|\eta_{it}|^4 \leq C < \infty$, *for all* (i,t), *and* $E^*|\nu_j|^{4q} \leq C < \infty$, $j = 1, \ldots, k_T$, *for some* $q > 1$, *if* $\frac{\sqrt{T}}{N} \to c < \infty$ *and* $b_T \to \infty$ *such that* $\frac{b_T^2}{T} \to 0$, *as* $N, T \to \infty$, *then* $\sup_{x \in \mathbb{R}^{\dim(\delta)}} \left| P^*\left(\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right) \leq x\right) - P\left(\sqrt{T}\left(\hat{\delta} - \delta\right) \leq x\right)\right| \to^P 0$.

1.3.3 Dependent wild bootstrap

In this section, we consider the dependent wild bootstrap as an alternative to the block wild bootstrap. The dependent wild bootstrap was proposed by Shao (2010) and differs from the BWB by the fact that the draws of the external variable are smoothed across observations. The DWB is implemented by multiplying each residual by a variable which is a local weighted average of external draws. The local weighting makes neighboring observations dependent, and this explains why it is valid under serial correlation. More formally, the DWB observations are obtained as

$$\varepsilon_{t+h}^{*} = \hat{\varepsilon}_{t+h} \cdot w_{t+h}^{*},$$

where w_{t+h}^{*} is the typical element of a vector w^{*} of length $T-h$ of random draws with mean 0 and covariance matrix K, with typical element $K_{ij} = E^{*}\left(w_{i}^{*} \cdot w_{j}^{*}\right) = k_{dwb}\left(\frac{j-i}{l_T}\right)$, with $k_{dwb}\left(\cdot\right)$ a kernel function and l_T a bandwidth parameter. Following Shao (2010, Assumption 2.1), we assume that w^{*} is l_T-dependent. In our simulations, we set $w^{*} = K^{1/2}w$, where $w \sim N\left(0, I_{T-h}\right)$. Because the choices of kernel and bandwidth used to construct the DWB observations do not need to coincide with the choices of kernel and bandwidth used to construct the HAC estimator, we use different notations here.

We make the same assumptions as for the BWB with the addition of the following restriction on the class of kernels.

Assumption 7 $k_{dwb}: \mathbb{R} \rightarrow [0,1]$ is symmetric with compact support on $[-1,1]$, $k_{dwb}(0) = 1$, $\lim_{x\rightarrow 0}\left\{1 - k_{dwb}(x)\right\}/|x|^{q} \neq 0$ for some $q \in (0,\ 2]$ such that $\psi(\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} k_{dwb}(x)\, e^{i\xi x}dx \geq 0$ for all $\xi \in \mathbb{R}$.

The condition $\psi(\xi) \geq 0$ ensures that the matrix K is positive definite (see Shao (2010)). These assumptions are satisfied by the Bartlett and Parzen kernels but not for the truncated, quadratic spectral and the Tukey-Hanning kernels (see Andrews (1991), Davidson and De Jong (2000) and Shao (2010)).

The following theorem justifies the dependent wild bootstrap for inference on δ.

Theorem 4. *Under the same assumptions as in Lemma 2 and Assumption 7, and assuming* $E^{*}\left|\eta_{it}\right|^{4} \leq C < \infty$, $E^{*}\left|w_{t}^{*}\right|^{2r} \leq C < \infty$, *for some* $r > 2$, *if* $\frac{\sqrt{T}}{N} \rightarrow c < \infty$ *and* $l_T \rightarrow \infty$ *such that* $T^{-1}l_{T}^{2(r+1)/r} \rightarrow 0$, *as* $N, T \rightarrow \infty$, *then*

$$\sup_{x\in\mathbb{R}^{\dim(\delta)}}\left|P^{*}\left(\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^{*} - \hat{\delta}\right) \leq x\right) - P\left(\sqrt{T}\left(\hat{\delta} - \delta\right) \leq x\right)\right| \rightarrow^{P} 0.$$

This result is the DWB analog of Theorem 3 for the BWB. Both theorems allow us to use these two methods for constructing percentile confidence intervals using the bootstrap. In order to construct percentile-t intervals (Hall, 1992), we need a consistent estimator of the variance of $\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^{*} - \hat{\delta}\right)$ to define studentized statistics. This estimator is given by $\Phi^{*\prime}\hat{\Sigma}_{\delta}^{*}\Phi^{*}$, where

$$\hat{\Sigma}_{\delta}^{*} = \left(T^{-1}\hat{z}^{*\prime}\hat{z}^{*}\right)^{-1}\hat{\Omega}^{*}\left(T^{-1}\hat{z}^{*\prime}\hat{z}^{*}\right)^{-1},$$

with $\hat{\Omega}^{*}$ being a HAC estimator

$$\hat{\Omega}^{*} = \hat{\Xi}_{0}^{*} + \sum_{j=1}^{T-h} k^{*}\left(\frac{j}{M_{T}^{*}}\right)\left[\hat{\Xi}_{j}^{*} + \hat{\Xi}_{j}^{*\prime}\right]$$

where $k^*(\cdot)$ and M_T^* denote the kernel function and the bandwidth parameter used in the bootstrap HAC estimator and $\hat{\Xi}_j^* = \frac{1}{T}\sum_{t=1}^{T-h-j} \hat{z}_t^* \hat{z}_{t+j}^{*\prime} \hat{\varepsilon}_{t+h}^* \hat{\varepsilon}_{t+h+j}^*$.

The consistency of $\hat{\Sigma}_\delta^*$ is formalized in the next lemma.

Lemma 5. *Suppose the assumptions of Theorems 3 and 4 hold for the DWB and the BWB, respectively. Let $k^*(\cdot)$ belong to the set $\mathcal{K}_1$ and $M_T^* \to \infty$ as $T \to \infty$ such that $\frac{M_T^{*2}}{T} \to 0$. If $\frac{\sqrt{T}}{N} \to c < \infty$ as $N, T \to \infty$, then $\widehat{\Sigma}_\delta^* \to^{P^*} \Sigma_\delta^* \equiv \left(\Phi_0^{*\prime}\right)^{-1} \Sigma_\delta \left(\Phi_0^*\right)^{-1}$, in probability.*

This result implies the consistency of the bootstrap distribution of the studentized statistic for any given coefficient and justifes the construction of symmetric or equal-tailed percentile-t confidence intervals.

1.4 Simulation results

In this section, we report results of a simulation experiment to document the properties of the bootstrap inference procedures above. Our design follows Gonçalves, Perron, and Djogbenou (2015) closely. We consider a single factor model,

$$y_{t+h} = \alpha F_t + \varepsilon_{t+h},$$

where $\alpha = 1$ and F_t is an AR(1) process, $F_t = 0.8F_{t-1} + u_t$, with u_t drawn for a normal distribution with mean 0 and variance $1 - (0.8)^2$ independently over time.

We consider three possibilities for the error term ε_{t+h}. In the first two designs, we set $h = 1$ or 12, and let the error term follow an MA$(h-1)$ as is appropriate if the forecasting model is correctly specified. In each case, following Cheng and Hansen (2013), the MA process is $\varepsilon_{t+h} = \sum_{j=0}^{h-1} (0.8)^j v_{t+h-j}$, and $v_t \sim N\left(0, \left(\sum_{j=0}^{h-1} (0.8)^{2j}\right)^{-1}\right)$ so that ε_{t+h} has variance 1. Finally, in the last design, we set $h = 1$ and generate ε_{t+h} from an AR(1) process, $\varepsilon_{t+h} = .8\varepsilon_{t+h-1} + v_{t+h}$, with v_{t+h} drawn for a normal with expectation 0 and variance $(1 - .8^2)$. This design is plausible for cases where the forecasting model is dynamically misspecified.

As in Gonçalves, Perron, and Djogbenou (2015), the $(T \times N)$ matrix of panel variables is generated as,

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

where λ_i is drawn from a $U[0,1]$ distribution (independent across i) and e_{it} is heteroskedastic but independent over i and t. The variance of e_{it} is drawn from $U[.5, 1.5]$ for each i.

We consider asymptotic and bootstrap confidence intervals at a nominal level of 95% for the regression coefficient. Asymptotic inference is conducted using a HAC estimator with a quadratic spectral kernel and with bandwidth selected by the data-based rule from Andrews (1991), both in the original sample and in the bootstrap samples. We consider three bootstrap schemes for generating ε_{t+h}^* in step 3 of our algorithm : the wild bootstrap, the block wild bootstrap with block size equal to the integer part of the bandwidth choice in the sample, and the dependent wild bootstrap with Bartlett kernel and bandwidth equal to the one selected in the sample.

We consider two values for each of N and T, 50 and 100, so that we have a total of four sample sizes. For all our bootstrap schemes, we let $\eta_{it} \sim N(0,1)$. Moreover, for the BWB, we

Table 1. Simulation results

			MA(h-1) errors								AR(1) errors			
			h = 1				h = 12				h = 1			
		N =	50		100		50		100		50		100	
		T =	50	100	50	100	50	100	50	100	50	100	50	100
Coverage rates for coefficient		OLS	56,9	54,3	75,8	77,8	68,7	71,2	75,3	81,0	61,5	68,5	72,1	79,5
		True Ft	92,0	93,8	91,7	93,5	80,5	86,0	80,6	85,7	77,7	84,8	78,7	86,0
	Symmetric t	WB	87,0	89,3	91,1	92,5	82,9	88,8	84,7	90,1	84,3	90,3	86,3	91,0
		BWB	86,9	89,4	90,9	92,4	84,3	88,9	86,0	90,5	85,1	90,0	87,3	91,2
		DWB	86,9	89,5	90,8	92,7	84,5	89,2	86,3	90,5	85,1	89,8	87,3	91,5
	Equal-tailed t	WB	89,1	91,0	90,7	92,7	74,1	80,6	74,5	81,2	75,5	80,6	77,0	82,7
		BWB	89,0	91,3	91,1	92,5	77,2	84,5	78,4	85,4	78,7	85,3	80,9	86,7
		DWB	89,0	90,9	90,5	92,4	77,9	85,0	79,0	85,9	78,9	85,7	81,3	87,1
Length of intervals		OLS	0,55	0,40	0,54	0,39	0,92	0,69	0,92	0,70	0,75	0,62	0,76	0,63
		True Ft	0,57	0,40	0,57	0,40	0,98	0,72	0,97	0,73	0,81	0,66	0,81	0,66
	Symmetric t	WB	0,99	0,72	0,82	0,57	1,45	1,16	1,23	0,96	1,38	1,15	1,17	0,93
		BWB	1,00	0,72	0,83	0,57	1,56	1,17	1,36	0,99	1,45	1,14	1,25	0,95
		DWB	1,00	0,73	0,83	0,57	1,56	1,17	1,35	1,00	1,45	1,14	1,25	0,95
	Equal-tailed t	WB	0,70	0,48	0,65	0,44	1,05	0,78	0,98	0,75	0,96	0,75	0,90	0,71
		BWB	0,71	0,48	0,65	0,44	1,20	0,87	1,15	0,85	1,09	0,84	1,04	0,80
		DWB	0,70	0,48	0,65	0,44	1,20	0,88	1,15	0,85	1,09	0,84	1,04	0,80
Bias		OLS	-0,21	-0,16	-0,14	-0,10	-0,20	-0,17	-0,13	-0,10	-0,21	-0,16	-0,14	-0,10
		True Ft	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
		WB	-0,13	-0,12	-0,10	-0,08	-0,14	-0,12	-0,10	-0,08	-0,14	-0,12	-0,10	-0,08
		BWB	-0,13	-0,12	-0,10	-0,08	-0,14	-0,12	-0,10	-0,08	-0,14	-0,12	-0,10	-0,08
		DWB	-0,13	-0,12	-0,10	-0,08	-0,14	-0,12	-0,10	-0,08	-0,14	-0,12	-0,10	-0,08
Bandwidth choices		OLS	1,59	1,64	1,59	1,64	4,09	5,70	4,51	6,23	4,56	6,01	4,95	6,46
		WB	1,50	1,55	1,51	1,56	1,56	1,65	1,57	1,68	1,59	1,67	1,60	1,70
		BWB	1,56	1,62	1,57	1,63	2,64	3,75	2,99	4,29	2,85	3,90	3,32	4,47
		DWB	1,56	1,62	1,56	1,64	2,47	3,67	2,75	4,16	2,74	3,85	3,10	4,37

let $\nu_j \sim N(0,1)$ whereas we let $w^* = K^{1/2}w$, with $w \sim N(0, I_{T-h})$ for the DWB. We set the number of replications to 5,000 and the number of bootstrap to 399.

Table 1 reports our simulation results. We report coverage rates of confidence intervals, the bias of the estimators, the length of the confidence intervals, and the bandwidth choices made in the sample and in the bootstrap.

The first set of results are coverage rates of the confidence intervals. We report results for the OLS estimator, the OLS estimator if we did not have to estimate the factors, and six bootstrap intervals. We report coverage rates of symmetric-t and equal-tailed-t intervals for the wild bootstrap (WB), the block wild bootstrap (BWB) and dependent wild bootstrap (DWB). Remember that the wild bootstrap is not valid with serial correlation.

The results for the first DGP are similar to those of GP (2014). The OLS estimator suffers from severe undercoverage. These distortions come from the presence of a bias associated with the estimation of the factor. This is illustrated in two ways : first, the OLS estimator with the true factor has coverage much closer to the nominal level, and second, the bias results show that the OLS estimator is biased (downward) when the factor must be estimated (and this bias goes down with N and T), while the estimator is essentially unbiased when we use the true factor.

The bootstrap is successful in removing this bias and providing more reliable inference. Whereas coverage is only 57% with $N = T = 50$ for asymptotic theory, symmetric bootstrap intervals have a coverage rate of about 87% and equal-tailed intervals about 89%. As N and T increase, coverage rates approach their nominal levels. With this design, all three bootstrap methods are asymptotically valid, and we see only small differences among them.

It is interesting to note that the equal-tailed intervals are much shorter than the symmetric intervals. This is because the sampling distribution of the OLS estimator is shifted to the left, and imposing symmetry around 0 is inappropriate in this case and entails a cost. We also see that the equal-tailed intervals provide slightly better coverage than the symmetric ones.

Many of the same features are reproduced in the other two designs. The OLS estimator is still biased due to the estimation of the factor, but the effect on coverage is not as dramatic as the bias of the estimator is unaffected but its variance increases. Thus, the t-statistic is less shifted to the left than in the first design, and the overall effect is that coverage improves. We do see the effect of serial correlation on the deterioration of inference for the OLS estimator with the true factor.

In the last two designs, we see differences among bootstrap methods. The wild bootstrap does not reproduce serial correlation and leads to intervals with lower coverage rates with equal-tailed intervals. On the other hand, we see little difference with the symmetric-t intervals. The fact that the wild bootstrap does not reproduce serial correlation is highlighted by the selected bandwidths. The selected bandwidth in the wild bootstrap is similar to the selected bandwidth when the data was i.i.d in the first design. The selected bandwidth in the BWB and DWB are lower than in the sample but large enough to capture some of the serial correlation in the bootstrap errors. Moreover, the dependent wild bootstrap provides slightly better coverage than the BWB. However, contrary to the first design, the symmetric intervals provide much better coverage than the equal-tailed intervals. This is due to the fact that the bias is less important in these designs than in the first one relative to the variance. Nevertheless, the equal-tailed intervals are much shorter than the symmetric ones.

Conclusion

In this paper, we theoretically justify two bootstrap methods for inference on the coefficients in factor-augmented regressions with serial correlation. Serial correlation naturally arises in a multi-step forecasting context or in a forecasting model that is dynamically misspecified. Our proposed bootstrap algorithm resamples the idiosyncratic errors with the wild bootstrap and the regression errors with either the block wild bootstrap or dependent wild bootstrap. Both methods are proved to provide valid inference under strong mixing dependence despite factor estimation error.

The results in this paper can be used to construct valid prediction intervals for the conditional mean or the realization of the variable of interest h periods into the future. This extension of the current results is explored in a recent paper by Gonçalves, Perron, and Djogbenou (2015).

Chapitre 2

Bootstrap Prediction Intervals for Factor Models

2.1 Introduction

Forecasting using factor-augmented regression models has become increasingly popular since the seminal paper of Stock and Watson (2002). The main idea underlying the so-called diffusion index forecasts is that when forecasting a given variable of interest, a large number of predictors can be summarized by a small number of indexes when the data follows an approximate factor model. The indexes are the latent factors driving the panel factor model and can be estimated by principal components. Point forecasts can be obtained by running a standard OLS regression augmented with the estimated factors.

In this paper, we consider the construction of prediction intervals in factor-augmented regression models using the bootstrap. In particular, our main contribution is to show the consistency of bootstrap intervals for a future target variable and its conditional mean. Our results allow for the construction of bootstrap prediction intervals without assuming Gaussianity and with better finite-sample properties than those based on asymptotic theory.

To be more specific, suppose that y_{t+h} denotes the variable to be forecast (where h is the forecast horizon) and let X_t be a N-dimensional vector of candidate predictors. We assume that y_{t+h} follows a factor-augmented regression model,

$$y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h}, \; t = 1, \ldots, T-h, \tag{2.1}$$

where W_t is a vector of observed regressors (including for instance lags of y_t) which jointly with F_t help forecast y_{t+h}. The r-dimensional vector F_t describes the common latent factors in the

This chapter is a joint paper with Sílvia Gonçalves and Benoit Perron. We are grateful for comments from Lutz Kilian and seminar participants at the University of California, San Diego, the University of Southern California, Queen's, Pompeu Fabra, the Toulouse School of Economics, Duke, Western, Sherbrooke, and Michigan as well as from participants at the 23rd meeting of the (EC)^2, the 2013 North American Winter Meeting of the Econometric Society, the 2013 meeting of the Canadian Economics Association, the 2013 Joint Statistical Meetings, the conference "Bootstrap Methods for Time Series" in Copenhagen, the MAESG conference in Atlanta, and the 2014 Canadian Econometric Study Group. We also thank Tatevik Sekhposyan for providing the data for the empirical application. Gonçalves acknowledges financial support from the NSERC and MITACS whereas Perron acknowledges financial support from the SSHRC and MITACS. We acknowledge that this chapter has been published by Journal of Business and Economics Statistics and agree with the publication disclosure.

panel factor model,

$$X_{it} = \lambda_i' F_t + e_{it}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{2.2}$$

where the $r \times 1$ vector λ_i contains the factor loadings and e_{it} is an idiosyncratic error term.

The goal is to forecast y_{T+h} or its conditional mean $y_{T+h|T} = \alpha' F_T + \beta' W_T$ using $\{(y_t, X_t, W_t) : t = 1, \ldots, T\}$, the available data at time T. Since factors are not observed, the diffusion index forecast approach typically involves a two-step procedure : in the first step we estimate F_t by principal components (yielding $\tilde{F}_t$) and in the second step we regress y_{t+h} on W_t and $\tilde{F}_t$ to obtain the regression coefficients. The point forecast is then constructed as $\hat{y}_{T+h|T} = \hat{\alpha}' \tilde{F}_T + \hat{\beta}' W_T$. Because we treat factors as latent, point forecasts depend both on estimated factors and regression coefficients. These two sources of parameter uncertainty must be accounted for when constructing prediction intervals and confidence intervals, as shown by Bai and Ng (2006).

Under regularity conditions, Bai and Ng (2006) derived the asymptotic distribution of regression estimates and the corresponding forecast errors and proposed the construction of asymptotic intervals. Our motivation for using the bootstrap as an alternative method of inference is twofold. First, the finite sample properties of the asymptotic approach of Bai and Ng (2006) can be poor, especially if N is not sufficiently large relative to T. This was recently shown by Gonçalves and Perron (2014) in the context of confidence intervals for the regression coefficients, and as we will show below, the same is true in the context of prediction intervals. In particular, estimation of factors leads to an asymptotic bias term in the OLS estimator if $\sqrt{T}/N \to c$ and $c \neq 0$. Gonçalves and Perron (2014) proposed a bootstrap method that removes this bias and outperforms the asymptotic approach of Bai and Ng (2006). Second, the bootstrap allows for the construction of prediction intervals for y_{T+h} that are consistent under more general assumptions than the asymptotic approach of Bai and Ng (2006). In particular, the bootstrap does not require the Gaussianity assumption on the regression errors that justifies the asymptotic prediction intervals of Bai and Ng (2006). As our simulations show, prediction intervals based on the Gaussianity assumption perform poorly when the regression error is asymmetrically distributed whereas the bootstrap prediction intervals do not suffer significant size distortions.

We apply our procedure to forecasting inflation changes using quarterly observations on the US GDP deflator for the period 1973-2014. The resulting bootstrap intervals differ in interesting ways from the asymptotic ones in specific periods. In particular, the 95% equal-tailed percentile-t bootstrap intervals are shifted downwards and lie entirely below 0 following the financial crisis of 2008 and during the last quarter of 2011. These periods were marked by a significant concern of deflation. Our intervals are more consistent with such concerns than the asymptotic ones which include some probability of increasing inflation.

The remainder of the paper is organized as follows. Section 2.2 introduces our forecasting model and considers asymptotic prediction intervals. Section 2.3 describes two bootstrap prediction algorithms. Section 2.4 presents a set of high level assumptions on the bootstrap idiosyncratic errors under which the bootstrap distribution of the estimated factors at a given time period is consistent for the distribution of the sample estimated factors. These results together with the results of Gonçalves and Perron (2014) and Djogbenou, Gonçalves, and Perron (2014) regarding inference on the coefficients are used in Section 2.5 to show the asymptotic validity of wild bootstrap prediction intervals. Section 2.6 presents our simulation experiments, while Section 2.7 presents an empirical illustration of our methods. Finally, Section 2.8 concludes. Mathematical proofs appear in the Appendix 0.2.

2.2 Prediction intervals based on asymptotic theory

This section introduces our assumptions and reviews the asymptotic theory-based prediction intervals proposed by Bai and Ng (2006).

2.2.1 Assumptions

Let $z_t = \begin{pmatrix} F_t' & W_t' \end{pmatrix}'$, where z_t is $p \times 1$, with $p = r + q$. Following Bai and Ng (2006), we make the following assumptions.

Assumption 1

(a) $E\left\|F_t\right\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \rightarrow^P \Sigma_F > 0$, where Σ_F is a non-random $r \times r$ matrix.

(b) The loadings λ_i are either deterministic such that $\|\lambda_i\| \leq M$, or stochastic such that $E\left\|\lambda_i\right\|^4 \leq M$. In either case, $\Lambda'\Lambda/N \rightarrow^P \Sigma_\Lambda > 0$, where Σ_Λ is a non-random matrix.

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

Assumption 2

(a) $E(e_{it}) = 0$, $E\left|e_{it}\right|^4 \leq M$.

(b) $E(e_{it}e_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ for all (t,s), $|\sigma_{ij,ts}| \leq \tau_{ts}$ for all (i,j). Furthermore, $\sum_{s=1}^{T} \tau_{ts} \leq M$, for each t, and $\frac{1}{NT}\sum_{t,s,i,j} |\sigma_{ij,ts}| \leq M$.

(c) For every (t,s), $E\left|N^{-1/2}\sum_{i=1}^{N}\left(e_{it}e_{is} - E\left(e_{it}e_{is}\right)\right)\right|^4 \leq M$.

(d) $\frac{1}{NT^2}\sum_{t,s,l,u}\sum_{i,j}\left|Cov\left(e_{it}e_{is}, e_{jl}e_{ju}\right)\right| < M < \infty$.

(e) For each t, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} \rightarrow^d N(0, \Gamma_t)$, where $\Gamma_t \equiv \lim_{N\rightarrow\infty} Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right) > 0$.

Assumption 3 The variables $\{\lambda_i\}$, $\{F_t\}$ and $\{e_{it}\}$ are three mutually independent groups. Dependence within each group is allowed.

Assumption 4

(a) $E(\varepsilon_{t+h}) = 0$ and $E\left|\varepsilon_{t+h}\right|^4 < M$.

(b) $E(\varepsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, \ldots) = 0$ for any $h > 0$, and (z_t', ε_t) are independent of the idiosyncratic errors e_{is} for all (i,s,t).

(c) $E\left\|z_t\right\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} z_t z_t' \rightarrow^P \Sigma_{zz} > 0$.

(d) As $T \rightarrow \infty$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} z_t\varepsilon_{t+h} \rightarrow^d N(0, \Omega)$, where $E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}\right\|^2 < M$, and $\Omega \equiv \lim_{T\rightarrow\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}\right) > 0$.

Assumptions 1 and 2 are standard in the approximate factors literature, allowing in particular for weak cross sectional and serial dependence in e_{it} of unknown form. Assumption 3 assumes independence among the factors, the factor loadings and the idiosyncratic error terms. We could allow for weak dependence among these three groups of variables at the cost of introducing restrictions on this dependence. Assumption 4 imposes moment conditions on $\{\varepsilon_{t+h}\}$, on $\{z_t\}$ and on the score vector $\{z_t\varepsilon_{t+h}\}$. Part c) requires $\{z_t z_t'\}$ to satisfy a law of large numbers. Part d) requires the score to satisfy a central limit theorem, where Ω denotes the limiting variance of the scaled average of the scores. We generalize the form of the covariance matrix assumed in Bai and Ng (2006) to allow for serial correlation as this will generally be the case when the forecast horizon is greater than 1.

2.2.2 Normal-theory intervals

As described in Section 1, the diffusion index forecasts are based on a two step estimation procedure. The first step consists of extracting the common factors $\tilde{F}_t$ from the N-dimensional panel X_t. In particular, given X, we estimate F and Λ with the method of principal components. F is estimated with the $T \times r$ matrix $\tilde{F} = \begin{pmatrix} \tilde{F}_1 & \dots & \tilde{F}_T \end{pmatrix}'$ composed of $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of XX'/TN (arranged in decreasing order), where the normalization $\frac{\tilde{F}'\tilde{F}}{T} = I_r$ is used. The matrix containing the estimated loadings is then $\tilde{\Lambda} = \left(\tilde{\lambda}_1, \dots, \tilde{\lambda}_N\right)' = X'\tilde{F}\left(\tilde{F}'\tilde{F}\right)^{-1} = X'\tilde{F}/T$.

In the second step, we run an OLS regression of y_{t+h} on $\hat{z}_t = \begin{pmatrix} \tilde{F}_t' & W_t' \end{pmatrix}'$, i.e. we compute

$$\hat{\delta} \equiv \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \left(\sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t'\right)^{-1} \sum_{t=1}^{T-h} \hat{z}_t y_{t+h}, \tag{2.3}$$

where $\hat{\delta}$ is $p \times 1$ with $p = r + q$.

Suppose the object of interest is $y_{T+h|T}$, the conditional mean of $y_{T+h} = \alpha' F_T + \beta' W_T + \varepsilon_{T+h}$ at time T. The point forecast is $\hat{y}_{T+h|T} = \hat{\alpha}'\tilde{F}_T + \hat{\beta}' W_T$ and the forecast error is given by

$$\hat{y}_{T+h|T} - y_{T+h|T} = \frac{1}{\sqrt{T}} \hat{z}_T' \sqrt{T}\left(\hat{\delta} - \delta\right) + \frac{1}{\sqrt{N}} \alpha' H^{-1} \sqrt{N}\left(\tilde{F}_T - HF_T\right), \tag{2.4}$$

where $\delta \equiv \begin{pmatrix} \alpha' H^{-1} & \beta' \end{pmatrix}'$ is the probability limit of $\hat{\delta}$. The matrix H is defined as

$$H = \tilde{V}^{-1} \frac{\tilde{F}'F}{T} \frac{\Lambda'\Lambda}{N}, \tag{2.5}$$

where $\tilde{V}$ is the $r \times r$ diagonal matrix containing on the main diagonal the r largest eigenvalues of XX'/NT, in decreasing order (cf. Bai (2003)). It arises because factor models are only identified up to rotation, implying that the principal component estimator $\tilde{F}_t$ converges to HF_t, and the OLS estimator $\hat{\alpha}$ converges to $H^{-1\prime}\alpha$. It must be noted that forecasts do not depend on this rotation since the product is uniquely identified.

The above decomposition shows that the asymptotic distribution of the forecast error depends on two sources of uncertainty : the first is the usual parameter estimation uncertainty associated with estimation of α and β, and the second is the factors estimation uncertainty. Under Assumptions 1-4, and assuming that $\sqrt{T}/N \to 0$ and $\sqrt{N}/T \to 0$ as $N, T \to \infty$, Bai and Ng (2006) show that the studentized forecast error

$$\frac{\hat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{\hat{B}_T}} \to^d N(0,1), \tag{2.6}$$

where $\hat{B}_T$ is a consistent estimator of the asymptotic variance of $\hat{y}_{T+h|T}$ given by

$$\hat{B}_T = \widehat{Var}\left(\hat{y}_{T+h|T}\right) = \frac{1}{T} \hat{z}_T' \hat{\Sigma}_\delta \hat{z}_T + \frac{1}{N} \hat{\alpha}' \hat{\Sigma}_{\tilde{F}_T} \hat{\alpha}. \tag{2.7}$$

Here, $\hat{\Sigma}_\delta$ consistently estimates $\Sigma_\delta = Var\left(\sqrt{T}\left(\hat{\delta}-\delta\right)\right)$ and $\hat{\Sigma}_{\tilde{F}_T}$ consistently estimates $\Sigma_{\tilde{F}_T} = Var\left(\sqrt{N}\left(\tilde{F}_T - HF_T\right)\right)$. In particular, under Assumptions 1-4,

$$\hat{\Sigma}_\delta = \left(T^{-1}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\right)^{-1}\hat{\Omega}_T\left(T^{-1}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\right)^{-1}, \tag{2.8}$$

where $\hat{\Omega}_T$ is a heteroskedasticity and autocorrelation consistent (HAC) estimator of $\Omega = \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}\right)$, and

$$\hat{\Sigma}_{\tilde{F}_T} = \tilde{V}^{-1}\tilde{\Gamma}_T\tilde{V}^{-1}, \tag{2.9}$$

where $\tilde{\Gamma}_T$ is an estimator of $\Gamma_T = \lim_{N\to\infty} Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{iT}\right)$ which depends on the cross sectional dependence and heterogeneity properties of e_{iT}. Bai and Ng (2006) provide three different estimators of Γ_T. Section 2.5 below considers such an estimator.

The central limit theorem result in (2.6) justifies the construction of an asymptotic $100(1-\alpha)\%$ level confidence interval for $y_{T+h|T}$ given by

$$\left(\hat{y}_{T+h|T} - z_{1-\alpha/2}\sqrt{\hat{B}_T}, \hat{y}_{T+h|T} + z_{1-\alpha/2}\sqrt{\hat{B}_T}\right), \tag{2.10}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution.

When the object of interest is a prediction interval for y_{T+h}, Bai and Ng (2006) propose

$$\left(\hat{y}_{T+h|T} - z_{1-\alpha/2}\sqrt{\hat{C}_T}, \hat{y}_{T+h|T} + z_{1-\alpha/2}\sqrt{\hat{C}_T}\right), \tag{2.11}$$

where

$$\hat{C}_T = \hat{B}_T + \hat{\sigma}_\varepsilon^2,$$

with $\hat{B}_T$ as above and $\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$. The validity of (2.11) depends on the additional assumption that ε_t is i.i.d. $N\left(0,\sigma_\varepsilon^2\right)$.

An important condition that justifies (2.10) and (2.11) is that $\sqrt{T}/N \to 0$. This condition ensures that the term reflecting the parameter estimation uncertainty in the forecast error decomposition (2.4), $\sqrt{T}\left(\hat{\delta}-\delta\right)$, is asymptotically normal with a mean of zero and a variance-covariance matrix that does not depend on the factors estimation uncertainty. As was recently shown by Gonçalves and Perron (2014), when $\sqrt{T}/N \to c \neq 0$,

$$\sqrt{T}\left(\hat{\delta}-\delta\right) \to^d N\left(-c\Delta_\delta, \Sigma_\delta\right),$$

where Δ_δ is a bias term that reflects the contribution of the factors estimation error to the asymptotic distribution of the regression estimates $\hat{\delta}$. In this case, the two terms in (2.4) will depend on the factors estimation uncertainty and a natural question is whether this will have an effect on the prediction intervals (2.10) and (2.11) derived by Bai and Ng (2006) under the assumption that $c = 0$. As we argue next, these intervals remain valid even when $c \neq 0$. The main reason is that when $\sqrt{T}/N \to c \neq 0$, the ratio $N/T \to 0$, which implies that the parameter estimation uncertainty associated with δ is dominated asymptotically by the uncertainty from having to estimate F_T.

More formally, when $\sqrt{T}/N \to c \neq 0$, $N/T \to 0$ and the convergence rate of $\hat{y}_{T+h|T}$ is $\sqrt{N}$, implying that

$$\begin{aligned}\sqrt{N}\left(\hat{y}_{T+h|T} - y_{T+h|T}\right) &= \sqrt{N/T}\sqrt{T}\left(\hat{\delta} - \delta\right)' \hat{z}_T + \alpha' H^{-1}\sqrt{N}\left(\tilde{F}_T - HF_T\right) \\ &= \alpha' H^{-1}\sqrt{N}\left(\tilde{F}_T - HF_T\right) + o_P(1).\end{aligned}$$

Thus, the forecast error is asymptotically $N\left(0, \alpha' H^{-1}\Sigma_{\tilde{F}_T} H^{-1\prime}\alpha\right)$. Since $N\hat{B}_T = (N/T)\,\hat{z}_T'\hat{\Sigma}_\delta \hat{z}_T + \hat{\alpha}'\hat{\Sigma}_{\tilde{F}_T}\hat{\alpha} = \alpha' H^{-1}\Sigma_{\tilde{F}_T} H^{-1\prime}\alpha + o_P(1)$, the studentized forecast error given in (2.6) is still $N(0,1)$ as $N, T \to \infty$. For the studentized forecast error associated with forecasting y_{T+h}, the variance of $\hat{y}_{T+h}$ is asymptotically (as $N, T \to \infty$) dominated by the variance of the error term σ_ε^2, implying that neither the parameter estimation uncertainty nor the factors estimation uncertainty contribute to the asymptotic variance.

2.3 Description of bootstrap intervals

Following Gonçalves and Perron (2014), we consider the following bootstrap data-generating process :

$$\begin{aligned} X_t^* &= \tilde{\Lambda}\tilde{F}_t + e_t^*, & (2.12)\\ y_{t+h}^* &= \hat{\alpha}'\tilde{F}_t + \hat{\beta}' W_t + \varepsilon_{t+h}^*, & (2.13)\end{aligned}$$

where $\left\{e_t^* = \left(e_{1t}^*, \ldots, e_{Nt}^*\right)'\right\}$ denotes a bootstrap sample from $\left\{\tilde{e}_t = X_t - \tilde{\Lambda}\tilde{F}_t\right\}$ and $\left\{\varepsilon_{t+h}^*\right\}$ is a resampled version of $\left\{\hat{\varepsilon}_{t+h} = y_{t+h} - \hat{\alpha}'\tilde{F}_t - \hat{\beta}' W_t\right\}$.

Our goal in this section is to describe two general bootstrap algorithms that can be used to compute intervals for $y_{T+h|T}$ and y_{T+h} for *any* choice of $\{e_t^*\}$ and $\left\{\varepsilon_{t+h}^*\right\}$. The specific method of generating $\{e_t^*\}$ and $\left\{\varepsilon_{t+h}^*\right\}$ will depend on the assumptions we make on $\{e_{it}\}$ and $\{\varepsilon_{t+h}\}$, respectively. In Section 2.5 we describe several methods. For example, we rely on the wild bootstrap to generate both $\{e_t^*\}$ and $\left\{\varepsilon_{t+1}^*\right\}$ when constructing confidence intervals for $y_{T+1|T}$. The wild bootstrap is justified in this setting since we assume away cross sectional dependence in e_{it} and we assume that ε_{t+1} is a m.d.s. when $h = 1$. For one-step ahead prediction intervals we strengthen the m.d.s. assumption to an i.i.d. assumption on ε_{t+1}, and therefore we generate ε_{t+1}^* using the i.i.d. bootstrap. For multi-step prediction intervals, we generate ε_{t+h}^* with either the block wild bootstrap or the dependent wild bootstrap of Djogbenou et al. (2014) to account for possible serial correlation.

We estimate the factors by the method of principal components using the bootstrap panel data set $\{X_t^* : t = 1, \ldots, T\}$. We let $\tilde{F}^* = \left(\tilde{F}_1^*, \ldots, \tilde{F}_T^*\right)'$ denote the $T \times r$ matrix of bootstrap estimated factors which equal the r eigenvectors of $X^* X^{*\prime}/NT$ (multiplied by $\sqrt{T}$) corresponding to the r largest eigenvalues. The $N \times r$ matrix of estimated bootstrap loadings is given by $\tilde{\Lambda}^* = \left(\tilde{\lambda}_1^*, \ldots, \tilde{\lambda}_N^*\right)' = X^{*\prime}\tilde{F}^*/T$. We then run a regression of y_{t+h}^* on $\tilde{F}_t^*$ and W_t using observations $t = 1, \ldots, T-h$. We let $\hat{\delta}^*$ denote the corresponding OLS estimator

$$\hat{\delta}^* = \left(\sum_{t=1}^{T-h} \hat{z}_t^* \hat{z}_t^{*\prime} \right)^{-1} \sum_{t=1}^{T-h} \hat{z}_t^* y_{t+h}^*,$$

where $\hat{z}_t^* = \left(\tilde{F}_t^{*\prime}, W_t' \right)'$.

The steps for obtaining a bootstrap confidence interval for $y_{T+h|T}$ are as follows.

Algorithm 1 (Bootstrap confidence interval for $y_{T+h|T}$)

1. For $t = 1, \ldots, T$, generate
$$X_t^* = \tilde{\Lambda} \tilde{F}_t + e_t^*,$$
where $\{e_{it}^*\}$ is a resampled version of $\left\{ \tilde{e}_{it} = X_{it} - \tilde{\lambda}_i' \tilde{F}_t \right\}$.

2. Estimate the bootstrap factors $\left\{ \tilde{F}_t^* : t = 1, \ldots, T \right\}$ using X^*.

3. For $t = 1, \ldots, T-h$, generate
$$y_{t+h}^* = \hat{\alpha}' \tilde{F}_t + \hat{\beta}' W_t + \varepsilon_{t+h}^*,$$
where the error term ε_{t+h}^* is a resampled version of $\hat{\varepsilon}_{t+h}$.

4. Regress y_{t+h}^* generated in step 3 on the bootstrap estimated factors $\tilde{F}_t^*$ obtained in step 2 and on the fixed regressors W_t and obtain the OLS estimator $\hat{\delta}^*$.

5. Obtain bootstrap forecasts
$$\hat{y}_{T+h|T}^* = \hat{\alpha}^{*\prime} \tilde{F}_T^* + \hat{\beta}^{*\prime} W_T \equiv \hat{\delta}^{*\prime} \hat{z}_T^*,$$
and bootstrap variance
$$\hat{B}_T^* = \frac{1}{T} \hat{z}_T^{*\prime} \hat{\Sigma}_\delta^* \hat{z}_T^* + \frac{1}{N} \hat{\alpha}^{*\prime} \hat{\Sigma}_{\tilde{F}_T}^* \hat{\alpha}^*, \tag{2.14}$$
where the choice of $\hat{\Sigma}_\delta^*$ and $\hat{\Sigma}_{\tilde{F}_T}^*$ depends on the properties of ε_{t+h}^* and e_{it}^*.

6. Let $y_{T+h|T}^* = \hat{\alpha}' \tilde{F}_T + \hat{\beta}' W_T$ and compute bootstrap prediction errors :
   (a) For equal-tailed percentile-t bootstrap intervals, compute studentized bootstrap prediction errors as
   $$s_{T+h}^* = \frac{\hat{y}_{T+h|T}^* - y_{T+h|T}^*}{\sqrt{\hat{B}_T^*}}.$$
   (b) For symmetric percentile-t bootstrap intervals, compute $\left| s_{T+h}^* \right|$.

7. Repeat this process B times, resulting in statistics $\left\{ s_{T+h,1}^*, \ldots, s_{T+h,B}^* \right\}$ and $\left\{ \left| s_{T+h,1}^* \right|, \ldots, \left| s_{T+h,B}^* \right| \right\}$.

8. Compute the corresponding empirical quantiles :
   (a) For equal-tailed percentile-t bootstrap intervals, $q_{1-\alpha}^*$ is the empirical $1-\alpha$ quantile of $\left\{ s_{T+h,1}^*, \ldots, s_{T+h,B}^* \right\}$.

(b) For symmetric percentile-t bootstrap intervals, $q^*_{|\cdot|,1-\alpha}$ is the empirical $1-\alpha$ quantile of $\left\{\left|s^*_{T+h,1}\right|,\ldots,\left|s^*_{T+h,B}\right|\right\}$.

A $100(1-\alpha)\%$ equal-tailed percentile-t bootstrap interval for $y_{T+h|T}$ is given by

$$EQ^{1-\alpha}_{y_{T+h|T}} \equiv \left(\hat{y}_{T+h|T} - q^*_{1-\alpha/2}\sqrt{\hat{B}_T}, \hat{y}_{T+h|T} - q^*_{\alpha/2}\sqrt{\hat{B}_T}\right), \tag{2.15}$$

whereas a $100(1-\alpha)\%$ symmetric percentile-t bootstrap interval for $y_{T+h|T}$ is given by

$$SY^{1-\alpha}_{y_{T+h|T}} \equiv \left(\hat{y}_{T+h|T} - q^*_{|\cdot|,1-\alpha}\sqrt{\hat{B}_T}, \hat{y}_{T+h|T} + q^*_{|\cdot|,1-\alpha}\sqrt{\hat{B}_T}\right), \tag{2.16}$$

When prediction intervals for a new observation y_{T+h} are the object of interest, the algorithm reads as follows.

Algorithm 2 (Bootstrap prediction interval for y_{T+h})

1. Identical to Algorithm 1.
2. Identical to Algorithm 1.
3. Generate $\left\{y^*_{1+h},\ldots,y^*_T,y^*_{T+1},\ldots,y^*_{T+h}\right\}$ using

$$y^*_{t+h} = \hat{\alpha}'\tilde{F}_t + \hat{\beta}'W_t + \varepsilon^*_{t+h},$$

where $\left\{\varepsilon^*_{1+h},\ldots,\varepsilon^*_T,\varepsilon^*_{T+1},\ldots,\varepsilon^*_{T+h}\right\}$ is a bootstrap sample obtained from $\left\{\hat{\varepsilon}_{1+h},\ldots,\hat{\varepsilon}_T\right\}$.

4. Not making use of the stretch $\left\{y^*_{T+1},\ldots,y^*_{T+h}\right\}$, compute $\hat{\delta}^*$ as in Algorithm 1.
5. Obtain the bootstrap point forecast $\hat{y}^*_{T+h|T}$ as in Algorithm 1 but compute its variance as

$$\hat{C}^*_T = \hat{B}^*_T + \hat{\sigma}^{*2}_\varepsilon,$$

where $\hat{\sigma}^{*2}_\varepsilon$ is a consistent estimator of $\sigma^2_\varepsilon = Var\left(\varepsilon_{T+h}\right)$ and $\hat{B}^*_T$ is as in Algorithm 1.

6. Let $y^*_{T+h} = \hat{\alpha}'\tilde{F}_T + \hat{\beta}'W_T + \varepsilon^*_{T+h}$ and compute bootstrap prediction errors :

   (a) For equal-tailed percentile-t bootstrap intervals, compute studentized bootstrap prediction errors as

$$s^*_{T+h} = \frac{\hat{y}^*_{T+h|T} - y^*_{T+h}}{\sqrt{\hat{C}^*_T}}.$$

   (b) For symmetric percentile-t bootstrap intervals, compute $\left|s^*_{T+h}\right|$.

7. Identical to Algorithm 1.
8. Identical to Algorithm 1.

A $100(1-\alpha)\,\%$ equal-tailed percentile-t bootstrap interval for y_{T+h} is given by

$$EQ^{1-\alpha}_{y_{T+h}} \equiv \left(\hat{y}_{T+h|T} - q^*_{1-\alpha/2}\sqrt{\hat{C}_T}, \hat{y}_{T+h|T} - q^*_{\alpha/2}\sqrt{\hat{C}_T}\right), \tag{2.17}$$

whereas a $100(1-\alpha)\,\%$ symmetric percentile-t bootstrap interval for y_{T+h} is given by

$$SY^{1-\alpha}_{y_{T+h}} \equiv \left(\hat{y}_{T+h|T} - q^*_{|\cdot|,1-\alpha}\sqrt{\hat{C}_T}, \hat{y}_{T+h|T} + q^*_{|\cdot|,1-\alpha}\sqrt{\hat{C}_T}\right). \tag{2.18}$$

The main differences between the two algorithms is that in step 3 of Algorithm 2 we generate observations for y^*_{t+h} for $t = 1, \ldots, T$ instead of stopping at $t = T - h$. This allows us to obtain a bootstrap observation for y^*_{T+h}, the bootstrap analogue of y_{T+h}, which we will use in constructing the studentized statistic s^*_{T+h} in step 6 of Algorithm 2. The point forecast is identical to Algorithm 1 and relies only on observations for $t = 1, \ldots, T-h$, but the bootstrap variance $\hat{C}^*_T$ contains an extra term $\hat{\sigma}^{*2}_{\varepsilon}$ that reflects the uncertainty associated with the error of the new observation ε_{T+h}.

Note that Algorithm 2 generates bootstrap point forecasts $\hat{y}^*_{T+h|T}$ and bootstrap future observations y^*_{T+h} that are conditional on W_T. This is important because the point forecast $\hat{y}_{T+h|T}$ depends on W_T. When W_t contains lagged dependent variables (e.g. $W_t = y_t$ and $h = 1$), steps 5 and 6 of Algorithm 2 set $W_T = y_T$ when computing $\hat{y}^*_{T+1|T}$ and y^*_{T+1}. This is effectively equivalent to setting $y^*_T = y_T$ for the purposes of computing these quantities. However, Step 3 of Algorithm 2 generates observations on $\left\{y^*_{t+1} : t = 1, \ldots, T\right\}$ that do not necessarily satisfy the requirement that $y^*_T = y_T$. As recently discussed by Pan and Politis (2014), we can account for parameter estimation uncertainty in predictions generated by autoregressive models by relying on a forward bootstrap method that contains two steps : one step generates the bootstrap data by relying on the forward representation of the model. This step accounts for parameter estimation uncertainty even if $y^*_T \neq y_T$. In a second step, we evaluate the bootstrap prediction and future observation conditional on the last value(s) of the observed variable. Our Algorithm 2 can be viewed as a version of the forward bootstrap method of Pan and Politis (2014) when some of the regressors are latent factors that need to be estimated.

2.4 Bootstrap distribution of estimated factors

The asymptotic validity of the bootstrap intervals for y_{T+h} and $y_{T+h|T}$ described in the previous section depends on the ability of the bootstrap to capture two sources of estimation error : the parameter estimation error and the factors estimation error. In particular, the bootstrap estimation error for the conditional mean is given by

$$\hat{y}^*_{T+h|T} - y^*_{T+h|T} = \frac{1}{\sqrt{T}}\hat{z}^{*\prime}_T\sqrt{T}\left(\hat{\delta}^* - \delta^*\right) + \frac{1}{\sqrt{N}}\hat{\alpha}' H^{*-1}\sqrt{N}\left(\tilde{F}^*_T - H^* F_T\right),$$

where $\delta^* = \Phi^{*\prime-1}\hat{\delta}$ and $\Phi^* = diag\left(H^*, I_q\right)$. Here, H^* is the bootstrap analogue of the rotation matrix H defined in (2.5), i.e.

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

where $\tilde{V}^*$ is the $r \times r$ diagonal matrix containing on the main diagonal the r largest eigenvalues of $X^*X^{*\prime}/NT$, in decreasing order. Note that contrary to H, which depends on unknown population parameters, H^* is fully observed. Using the results in Bai and Ng (2013) , H^* converges asymptotically to a diagonal matrix with $+1$ or -1 on the main diagonal, see Gonçalves and Perron (2014) for more details.

Adding and subtracting appropriately, we can write

$$\hat{y}^*_{T+h|T} - y^*_{T+h|T} = \frac{1}{\sqrt{T}} \hat{z}'_T \sqrt{T} \left(\Phi^{*\prime} \hat{\delta}^* - \hat{\delta} \right) + \frac{1}{\sqrt{N}} \hat{\alpha}' \sqrt{N} \left(H^{*-1} \tilde{F}^*_T - \tilde{F}_T \right) + o_{P^*}(1). \quad (2.19)$$

As usual in the bootstrap literature, we use P^* to denote the bootstrap probability measure, conditional on a given sample; E^* and Var^* denote the corresponding bootstrap expected value and variance operators. For any bootstrap statistic T^*_{NT}, we write $T^*_{NT} = o_{P^*}(1)$, in probability, or $T^*_{NT} \to^{P^*} 0$, in probability, when for any $\delta > 0$, $P^*(|T^*_{NT}| > \delta) = o_P(1)$. We write $T^*_{NT} = O_{P^*}(1)$, in probability, when for all $\delta > 0$ there exists $M_\delta < \infty$ such that $\lim_{N,T\to\infty} P[P^*(|T^*_{NT}| > M_\delta) > \delta] = 0$. Finally, we write $T^*_{NT} \to^{d^*} D$, in probability, if conditional on a sample with probability that converges to one, T^*_{NT} weakly converges to the distribution D under P^*, i.e. $E^*(f(T^*_{NT})) \to^P E(f(D))$ for all bounded and uniformly continuous functions f. See Chang and Park (2003) for similar notation and for several useful bootstrap asymptotic properties.

The stochastic expansion (2.19) shows that the bootstrap estimation error captures the two forms of estimation uncertainty in (2.4) provided: (1) the bootstrap distribution of $\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right)$ is a consistent estimator of the distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$, and (2) the bootstrap distribution of $\sqrt{N}\left(H^{*-1}\tilde{F}^*_T - \tilde{F}_T\right)$ is a consistent estimator of the distribution of $\sqrt{N}\left(\tilde{F}_T - HF_T\right)$. Gonçalves and Perron (2014) discussed conditions for the consistency of the bootstrap distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$. Here we propose a set of conditions that justifies using the bootstrap to consistently estimate the distribution of the estimated factors $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$ at each point t.

Condition $\mathcal{A}$.

$\mathcal{A}$.1. For each t, $\sum_{s=1}^T |\gamma^*_{st}|^2 = O_P(1)$, where $\gamma^*_{st} = E^*\left(\frac{1}{N}\sum_{i=1}^N e^*_{it} e^*_{is}\right)$.

$\mathcal{A}$.2. For each t, $\frac{1}{T}\sum_{s=1}^T E^*\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(e^*_{it}e^*_{is} - E^*\left(e^*_{it}e^*_{is}\right)\right)\right|^2 = O_P(1)$.

$\mathcal{A}$.3. For each t, $E^*\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^T\sum_{i=1}^N \tilde{F}_s\left(e^*_{it}e^*_{is} - E^*\left(e^*_{it}e^*_{is}\right)\right)\right\|^2 = O_P(1)$.

$\mathcal{A}$.4. $E^*\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^T\sum_{i=1}^N \tilde{F}_t\tilde{\lambda}'_i e^*_{it}\right\|^2 = O_P(1)$.

$\mathcal{A}$.5. $\frac{1}{T}\sum_{t=1}^T E^*\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\lambda}_i e^*_{it}\right\|^2 = O_P(1)$.

$\mathcal{A}$.6. For each t, $\Gamma_t^{*-1/2}\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\lambda}_i e^*_{it} \to^{d^*} N(0, I_r)$, in probability, where $\Gamma^*_t = Var^*\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\lambda}_i e^*_{it}\right)$ is uniformly positive definite.

Condition $\mathcal{A}$ is the bootstrap analogue of Bai's (2003) assumptions used to derive the limiting distribution of $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$. Gonçalves and Perron (2014) also relied on similar high level assumptions to study the bootstrap distribution of $\sqrt{T}\left(\hat{\delta}^* - \delta^*\right)$. In particular, Conditions $\mathcal{A}$.4 and $\mathcal{A}$.5 correspond to their Conditions B*(c) and B*(d), respectively. Since our goal here is to characterize the limiting distribution of the bootstrap estimated factors at each point t, we need to complement some of their other conditions by requiring boundedness in probability of some bootstrap moments at each point in time t (in addition to boundedness in probability of the time average of these bootstrap moments; e.g. Conditions $\mathcal{A}$.1 and $\mathcal{A}$.2 expand Conditions A*(b) and A*(c) in Gonçalves and Perron (2014) in this manner). We also require that a central limit theorem applies to the scaled cross sectional average of $\tilde{\lambda}_i e_{it}^*$, at each time t (Condition $\mathcal{A}$.6). This high level condition ensures asymptotic normality for the bootstrap estimated factors. It was not required by Gonçalves and Perron (2014) because their goal was only to consistently estimate the distribution of the regression estimates, not of the estimated factors.

Theorem 6. *Suppose Assumptions 1 and 2 hold. Under Condition $\mathcal{A}$, as $N, T \to \infty$ such that $\sqrt{N}/T^{3/4} \to 0$, we have that for each t,*

$$\sqrt{N}\left(\tilde{F}_t^* - H^* \tilde{F}_t\right) = H^* \tilde{V}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{\lambda}_i e_{it}^* + o_{P^*}(1),$$

in probability, which implies that

$$\Pi_t^{*-1/2} \sqrt{N}\left(H^{*-1} \tilde{F}_t^* - \tilde{F}_t\right) \to^{d^*} N(0, I_r),$$

in probability, where $\Pi_t^* = \tilde{V}^{-1} \Gamma_t^* \tilde{V}^{-1}$.

Theorem 1.(i) of Bai (2003) shows that under regularity conditions weaker than Assumptions 1 and 2 and provided $\sqrt{N}/T \to 0$, $\sqrt{N}\left(\tilde{F}_t - HF_t\right) \to^d N(0, \Pi_t)$, where $\Pi_t = V^{-1} Q \Gamma_t Q' V^{-1}$, $Q = p\lim\left(\frac{\tilde{F}'F}{T}\right)$. Theorem 6 is its bootstrap analogue. A stronger rate condition ($\sqrt{N}/T^{3/4} \to 0$ instead of $\sqrt{N}/T \to 0$) is used to show that the remainder terms in the stochastic expansion of $\sqrt{N}\left(\tilde{F}_t^* - H^* \tilde{F}_t\right)$ are asymptotically negligible. This rate condition is a function of the number of finite moments for F_s we assume. In particular, if we replace Assumption 1(a) with $E\left\|F_t\right\|^q \leq M$ for all t, then the required rate restriction is $\sqrt{N}/T^{1-1/q} \to 0$. See Remarks 1 and 3 below.

To prove the consistency of Π_t^* for Π_t we impose the following additional condition.

Condition $\mathcal{B}$. For each t, $p\lim \Gamma_t^* = Q \Gamma_t Q'$.

Condition $\mathcal{B}$ requires that Γ_t^*, the bootstrap variance of the scaled cross sectional average of the scores $\tilde{\lambda}_i e_{it}^*$, be consistent for $Q \Gamma_t Q'$. This in turn requires that we resample $\tilde{e}_{it}$ in a way that preserves the cross sectional dependence and heterogeneity properties of e_{it}.

Corollary 7. *Under Assumptions 1 and 2 and Conditions $\mathcal{A}$ and $\mathcal{B}$, we have that for each t, as $N, T \to \infty$ such that $\sqrt{N}/T^{3/4} \to 0$, $\sqrt{N}\left(H^{*-1} \tilde{F}_t^* - \tilde{F}_t\right) \to^{d^*} N(0, \Pi_t)$, in probability, where $\Pi_t = V^{-1} Q \Gamma_t Q' V^{-1}$ is the asymptotic covariance matrix of $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$.*

Corollary 7 justifies using the bootstrap to construct confidence intervals for the rotated factors HF_t provided Conditions $\mathcal{A}$ and $\mathcal{B}$ hold. These conditions are high level conditions that can be checked for any particular bootstrap scheme used to generate e_{it}^*. We verify them for a wild bootstrap in Section 2.5 when proving the consistency of bootstrap confidence intervals for the conditional mean.

The fact that factors and factor loadings are not separately identified implies the need to rotate the bootstrap estimated factors in order to consistently estimate the distribution of the sample factor estimates, i.e. we use $\sqrt{N}\left(H^{*-1}\tilde{F}_t^* - \tilde{F}_t\right)$ to approximate the distribution of $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$. A similar rotation was discussed in Gonçalves and Perron (2014) in the context of bootstrapping the regression coefficients $\hat{\delta}$.

2.5 Validity of bootstrap intervals

2.5.1 Confidence intervals for $y_{T+1|T}$

We begin by considering intervals for next period's conditional mean. For this purpose, we use a two-step wild bootstrap scheme, as in Gonçalves and Perron (2014). Specifically, we rely on Algorithm 1 and we let

$$\varepsilon_{t+1}^* = \hat{\varepsilon}_{t+1} \cdot v_{t+1}, \quad t = 1, \ldots, T-1, \tag{2.20}$$

with v_{t+1} i.i.d.$(0,1)$, and

$$e_{it}^* = \tilde{e}_{it} \cdot \eta_{it}, \quad t = 1, \ldots, T, \ i = 1, \ldots, N, \tag{2.21}$$

where η_{it} is i.i.d.$(0,1)$ across (i,t), independently of v_{t+1}.

To prove the asymptotic validity of this method we strengthen Assumptions 1-4 as follows.

Assumption 5. λ_i are either deterministic such that $\|\lambda_i\| \leq M < \infty$, or stochastic such that $E\|\lambda_i\|^{12} \leq M < \infty$ for all i; $E\|F_t\|^{12} \leq M < \infty$; $E|e_{it}|^{12} \leq M < \infty$, for all (i,t); and for some $q > 1$, $E|\varepsilon_{t+1}|^{4q} \leq M < \infty$, for all t.

Assumption 6. $E(e_{it}e_{js}) = 0$ if $i \neq j$.

With $h = 1$, our Assumption 4(b) on ε_{t+h} becomes a martingale difference sequence assumption, and the wild bootstrap in (2.20) is natural. This assumption rules out serial correlation in ε_{t+1} but allows for conditional heteroskedasticity. Below, we consider the case where $h > 1$.

Assumption 6 assumes the absence of cross sectional correlation in the idiosyncratic errors and motivates the use of the wild bootstrap in (2.21). As the results in the previous sections show, prediction intervals for y_{T+h} or $y_{T+h|T}$ are a function of the factors estimation uncertainty even when this source of uncertainty is asymptotically negligible for the estimation of the distribution of the regression coefficients (i.e. even when $\sqrt{T}/N \to c = 0$). Since factors estimation uncertainty depends on the cross sectional correlation of the idiosyncratic errors e_{it} (via $\Gamma_T = \lim_{N\to\infty} Var\left(1/\sqrt{N}\sum_{i=1}^N \lambda_i e_{iT}\right)$), bootstrap prediction intervals need to mimic this form of correlation to be asymptotically valid. Contrary to the pure time series context, a natural ordering does not exist in the cross sectional dimension, which implies that proposing a nonparametric bootstrap method (e.g. a block bootstrap) that replicates the cross sectional dependence is challenging if a parametric model is not assumed. Therefore, we follow Gonçalves and Perron (2014) and use a wild bootstrap to generate e_{it}^* under Assumption 6.

The bootstrap percentile-t method, as described in Algorithm 1 and equations (2.15) and (2.16), requires the choice of two variances, $\hat{B}_T$ and its bootstrap analogue $\hat{B}_T^*$. To compute $\hat{B}_T$ we use (2.7), where $\hat{\Sigma}_\delta$ is given in (2.8). $\hat{\Sigma}_{\tilde{F}_T}$ is given in (2.9), where

$$\tilde{\Gamma}_T = \frac{1}{N}\sum_{i=1}^N \tilde{\lambda}_i \tilde{\lambda}_i' \tilde{e}_{iT}^2$$

is estimator 5(a) in Bai and Ng (2006), and it is a consistent estimator of (a rotated version of) $\Gamma_T = \lim_{N\to\infty} Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i e_{iT}\right)$ under Assumption 6. We compute $\hat{B}_T^*$ using (2.14) and relying on the heteroskedasticity-robust bootstrap analogues of $\hat{\Sigma}_\delta$ and $\hat{\Sigma}_{\tilde{F}_T}$.

Theorem 8. *Suppose Assumptions 1-6 hold and we use Algorithm 1 with* $\varepsilon_{t+1}^* = \hat{\varepsilon}_{t+1}\cdot v_{t+1}$ *and* $e_{it}^* = \tilde{e}_{it}\cdot\eta_{it}$*, where* $v_{t+1} \sim i.i.d.(0,1)$ *for all* $t = 1,\ldots,T-1$ *and* $\eta_{it} \sim i.i.d.(0,1)$ *for all* $i = 1,\ldots,N$ *;* $t = 1,\ldots,T$*, and* v_{t+1} *and* η_{it} *are mutually independent. Moreover, assume that* $E^*|\eta_{it}|^4 < C$ *for all* (i,t) *and* $E^*|v_{t+1}|^4 < C$ *for all* t*. If* $\sqrt{T}/N \to c$*, where* $0 \le c < \infty$*, and* $\sqrt{N}/T^{11/12} \to 0$*, then conditional on* $\{y_t, X_t, W_t : t = 1,\ldots,T\}$*,*

$$\frac{\hat{y}_{T+1|T}^* - y_{T+1|T}^*}{\sqrt{\hat{B}_T^*}} \to^{d^*} N(0,1),$$

in probability.

Remark 1. *The rate restriction* $\sqrt{N}/T^{11/12} \to 0$ *is slightly stronger than the rate used by Bai (2003) (cf.* $\sqrt{N}/T \to 0$*). It is weaker than the restriction* $\sqrt{N}/T^{3/4} \to 0$ *used in Theorem 6 and Corollary 7 because we have strengthened the number of factor moments that exist from 4 to 12 (compare Assumption 5 with Assumption 1(a)). See Remark 3 in the Appendix.*

Remark 2. *Since* $\frac{\hat{y}_{T+1|T} - y_{T+1|T}}{\sqrt{\hat{B}_T}} \to^d N(0,1)$*, as shown by Bai and Ng (2006), Theorem 8 implies that bootstrap confidence intervals for* $y_{T+1|T}$ *obtained with Algorithm 1 have the correct coverage probability asymptotically.*

2.5.2 Prediction intervals for y_{T+1}

In this section we provide a theoretical justification for bootstrap prediction intervals for y_{T+1} as described in Algorithm 2. In particular, our goal is to prove that a bootstrap prediction interval contains the future observation y_{T+1} with unconditional probability that converges to the nominal level as $N, T \to \infty$.

We add the following assumption.

Assumption 7. ε_{t+1} is i.i.d.$(0, \sigma_\varepsilon^2)$ with a continuous distribution function $F_\varepsilon(x) = P(\varepsilon_{t+1} \le x)$.

Assumption 7 strengthens the m.d.s. Assumption 4.(b) by requiring the regression errors to be i.i.d. However, and contrary to Bai and Ng (2006), F_ε does not need to be Gaussian. The continuity assumption on F_ε is used below to prove that the Kolmogorov distance between the bootstrap distribution of the studentized forecast error and the distribution of its sample analogue converges in probability to zero.

Let the studentized forecast error be defined as

$$s_{T+1} \equiv \frac{\hat{y}_{T+1|T} - y_{T+1}}{\sqrt{\hat{B}_T + \hat{\sigma}_\varepsilon^2}},$$

where $\hat{\sigma}_\varepsilon^2$ is a consistent estimate of $\sigma_\varepsilon^2 = Var(\varepsilon_{T+1})$ and $\hat{B}_T = \widehat{Var}(\hat{y}_{T+1|T}) = \frac{1}{T}\hat{z}_T'\hat{\Sigma}_\delta\hat{z}_T + \frac{1}{N}\hat{\alpha}'\hat{\Sigma}_{\tilde{F}_T}\hat{\alpha}$. Given Assumption 7, we can use

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^2 \quad \text{and} \quad \hat{\Sigma}_\delta = \hat{\sigma}_\varepsilon^2\left(\frac{1}{T}\sum_{t=1}^{T-1}\hat{z}_t\hat{z}_t'\right)^{-1}. \tag{2.22}$$

Our goal is to show that the bootstrap can be used to estimate consistently $F_{T,s}(x) = P(s_{T+1} \leq x)$, the distribution function of s_{T+1}. Note that we can write

$$\begin{aligned} \hat{y}_{T+1|T} - y_{T+1} &= \left(\hat{y}_{T+1|T} - y_{T+1|T}\right) + \left(y_{T+1|T} - y_{T+1}\right) \\ &= -\varepsilon_{T+1} + O_P\left(1/\delta_{NT}\right), \end{aligned}$$

given that $\hat{y}_{T+1|T} - y_{T+1|T} = O_P\left(\frac{1}{\delta_{NT}}\right)$, where $\delta_{NT} = \min\left(\sqrt{N}, \sqrt{T}\right)$ (this follows under the assumptions of Theorem 8). Since $\hat{\sigma}_\varepsilon^2 \to^P \sigma_\varepsilon^2$ and $\hat{B}_T = O_P\left(1/\delta_{NT}^2\right) = o_P(1)$, it follows that

$$s_{T+1} = -\frac{\varepsilon_{T+1}}{\sigma_\varepsilon} + o_P(1). \tag{2.23}$$

Thus, as $N, T \to \infty$, s_{T+1} converges in distribution to the random variable $-\frac{\varepsilon_{T+1}}{\sigma_\varepsilon}$, i.e.

$$F_{T,s}(x) \equiv P(s_{T+1} \leq x) \to P\left(-\frac{\varepsilon_{T+1}}{\sigma_\varepsilon} \leq x\right) = 1 - F_\varepsilon(-x\sigma_\varepsilon) \equiv F_{\infty,s}(x),$$

for all $x \in \mathbb{R}$. If we assume that ε_{t+1} is i.i.d. $N(0, \sigma_\varepsilon^2)$, as in Bai and Ng (2006), then $F_\varepsilon(-x\sigma_\varepsilon) = \Phi(-x) = 1 - \Phi(x)$, implying that $F_{T,s}(x) \to \Phi(x)$, i.e. $s_{T+1} \to^d N(0,1)$. Nevertheless, this is not generally true unless we make the Gaussianity assumption. We note that although asymptotically the variance of the prediction error $\hat{y}_{T+1|T} - y_{T+1}$ does not depend on any parameter nor factors estimation uncertainty (as it is dominated by σ_ε^2 for large N and T), we still suggest using $\hat{C}_T = \hat{B}_T + \hat{\sigma}_\varepsilon^2$ to studentize $\hat{y}_{T+1|T} - y_{T+1}$ since $\hat{\sigma}_\varepsilon^2$ will underestimate the true forecast variance for finite T and N. Politis (2013) and Pan and Politis (2014) discuss notions of asymptotic validity that require taking into account the estimation of the condition mean. More specifically, in addition to requiring that the interval contains the true observation with the desired nominal coverage probability asymptotically, they require the bootstrap to capture parameter estimation uncertainty. To the extent that their definitions can be extended to the case of generated regressors, we expect our bootstrap intervals to satisfy these stricter notions of validity.

Next we show that the bootstrap yields a consistent estimate of the distribution of s_{T+1} without assuming that ε_{t+1} is Gaussian. Our proposal is based on a two-step residual based bootstrap scheme, as described in Algorithm 2 and equations (2.17) and (2.18), where in step 3 we generate $\left\{\varepsilon_2^*, \ldots, \varepsilon_T^*, \varepsilon_{T+1}^*\right\}$ as a random sample obtained from the centered residuals $\left\{\hat{\varepsilon}_2 - \bar{\hat{\varepsilon}}, \ldots, \hat{\varepsilon}_T - \bar{\hat{\varepsilon}}\right\}$. Resampling in an i.i.d. fashion is justified under Assumption 7. We recenter the residuals because $\bar{\hat{\varepsilon}}$ is not necessarily zero unless W_t contains a constant regressor. Nevertheless, since $\bar{\hat{\varepsilon}} = o_P(1)$, resampling the uncentered residuals is also asymptotically valid in our context. We compute $\hat{B}_T^*$ and $\hat{\sigma}_\varepsilon^{*2}$ using the bootstrap analogues of $\hat{\Sigma}_\delta$ and $\hat{\sigma}_\varepsilon^2$ introduced in (2.22). Note that $\hat{\sigma}_\varepsilon^{*2}$ is a consistent estimator of σ_ε^2 and $\hat{B}_T^* = o_{P^*}(1)$, in probability.

As above, we can write

$$\begin{aligned} \hat{y}_{T+1|T}^* - y_{T+1}^* &= \left(\hat{y}_{T+1|T}^* - y_{T+1|T}^*\right) + \left(y_{T+1|T}^* - y_{T+1}^*\right) \\ &= -\varepsilon_{T+1}^* + O_{P^*}\left(1/\delta_{NT}\right), \end{aligned}$$

in probability, which in turn implies

$$s_{T+1}^* \equiv \frac{\hat{y}_{T+1|T}^* - y_{T+1}^*}{\sqrt{\hat{B}_T^* + \hat{\sigma}_\varepsilon^{*2}}} = -\frac{\varepsilon_{T+1}^*}{\sigma_\varepsilon} + o_{P^*}(1). \tag{2.24}$$

Thus, $F^*_{T,s}(x) = P^*\left(s^*_{T+1} \leq x\right)$, the bootstrap distribution of s^*_{T+1} (conditional on the sample) is asymptotically the same as the bootstrap distribution of $-\frac{\varepsilon^*_{T+1}}{\sigma_\varepsilon}$.

Let $F^*_{T,\varepsilon}$ denote the bootstrap distribution function of ε^*_t. It is clear from the stochastic expansions (2.23) and (2.24) that the crucial step is to show that ε^*_{T+1} converges weakly in probability to ε_{T+1}, i.e. $d\left(F^*_{T,\varepsilon}, F_\varepsilon\right) \to^P 0$ for any metric that metrizes weak convergence. In the following we use Mallows metric which is defined as $d_2\left(F_X, F_Y\right) = \left(\inf\left(E\left|X-Y\right|^2\right)\right)^{1/2}$ over all joint distributions for the random variables X and Y having marginal distributions F_X and F_Y, respectively.

Lemma 9. *Under Assumptions 1-7, and as $T, N \to \infty$ such that $\sqrt{T}/N \to c$, $0 \leq c < \infty$, $d_2\left(F^*_{T,\varepsilon}, F_\varepsilon\right) \to^P 0$.*

Corollary 10. *Under the same assumptions as Theorem 8 strengthened by Assumption 7, we have that*

$$\sup_{x\in\mathbb{R}} \left|F^*_{T,s}(x) - F_{\infty,s}(x)\right| \to 0,$$

in probability.

Corollary 10 implies the asymptotic validity of the bootstrap prediction intervals given in (2.17) and (2.18), where asymptotic validity means that the interval contains y_{T+1} with unconditional probability converging to the nominal level asymptotically. Specifically, we can show that $P\left(y_{T+1} \in EQ^{1-\alpha}_{y_{T+1}}\right) \to 1-\alpha$ and $P\left(y_{T+1} \in SY^{1-\alpha}_{y_{T+1}}\right) \to 1-\alpha$ as $N, T \to \infty$. See e.g. Beran (1987) and Wolf and Wunderli (2015, Proposition 1). For instance,

$$\begin{aligned} P\left(y_{T+1} \in EQ^{1-\alpha}_{y_{T+1}}\right) &= P\left(s_{T+1} \leq q^*_{1-\alpha/2}\right) - P\left(s_{T+1} \leq q^*_{\alpha/2}\right) \\ &= P\left(F^*_{T,s}(s_{T+1}) \leq 1-\alpha/2\right) - P\left(F^*_{T,s}(s_{T+1}) \leq \alpha/2\right). \end{aligned}$$

Given Corollary 10, we have that $F^*_{T,s}(s_{T+1}) = F_{\infty,s}(s_{T+1}) + o_P(1)$, and we can show that $F_{\infty,s}(s_{T+1}) \to^d U[0,1]$. Indeed, for any x,

$$P\left(F_{\infty,s}(s_{T+1}) \leq x\right) = P\left(s_{T+1} \leq F^{-1}_{\infty,s}(x)\right) \equiv F_{T,s}\left(F^{-1}_{\infty,s}(x)\right) \to F_{\infty,s}\left(F^{-1}_{\infty,s}(x)\right) = x.$$

A stronger result than that implied by Corollary 10 would be to prove that $P\left(y_{T+1} \in EQ^{1-\alpha}_{y_{T+1}}|z_T\right) \to 1-\alpha$, where $z_T = \left(F'_T, W'_T\right)'$. Nevertheless, to claim asymptotic validity of the bootstrap prediction intervals conditional on the regressors would require stronger assumptions, namely the assumption that ε_{T+1} is independent of z_T. Such a strong exogeneity assumption is unlikely to be satisfied in economics.

2.5.3 Multi-horizon forecasting, $h > 1$

Finally, we consider the case where the forecasting horizon, h, is larger than 1. The main complication in this case is the fact that the regression errors ε_{t+h} in the factor-augmented regression will generally be serially correlated to order $h-1$. This serial correlation affects the distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$ since the form of Ω is different in this case, as it includes autocovariances of the score process.

We modify our two algorithms above by drawing ε^*_{t+h} using the block wild bootstrap (BWB) algorithm proposed in Djogbenou et al. (2015). The idea is to separate the sample residuals $\hat{\varepsilon}_{t+h}$ into non-overlapping blocks of b consecutive observations. For simplicity, we assume that $\frac{T-h}{b}$, the number of such blocks, is an integer. Then, we generate our bootstrap errors by multiplying each residual within a block by the *same* draw of an external variable, i.e.

$$\varepsilon^*_{i+(j-1)b} = \hat{\varepsilon}_{i+(j-1)b}\eta_j$$

for $j = 1, \ldots, \frac{T-h}{b}$, $i = 1+h, \ldots, h+b$, and $\eta_j \sim i.i.d.\,(0,1)$. The fact that each residual within a block is multiplied by the same external draw preserves the time series dependence. We let $b = h$ because we use the fact that $\varepsilon_{t+h} \sim MA\,(h-1)$ under Assumption 4(b). For $h = 1$, this algorithm is the same as the wild bootstrap. Djogbenou et al. (2015) show that this algorithm allows for valid bootstrap inference in a regression model with estimated factors and general mixing conditions on the error term. The moving average structure obtained in a forecasting context (assuming correct specification) obviously satisfies these mixing conditions, and this ensures that this block wild bootstrap algorithm replicates the distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$ after rotating the estimated parameter in the bootstrap world. Thus, the result of Theorem 8 holds in this more general context since $h > 1$ does not affect factor estimation.

For the forecast of the new observation, y_{T+h}, the crucial condition for asymptotic validity of the bootstrap prediction intervals is to capture the marginal distribution of $\varepsilon_{T+h.}$. This means that the i.i.d. bootstrap can still be used in step 2 of algorithm 2 to generate ε^*_{t+h} despite the serial correlation in ε_{t+h}. Alternatively, we can also amend the block wild bootstrap by generating $\hat{\varepsilon}^*_{t+h}$ as above for $t = 1, \ldots, T-h$ and generating ε^*_{T+h} as a draw from the empirical distribution function of $\hat{\varepsilon}_t$, $t = 1, \ldots, T-h$. We will compare these two approaches in the simulation experiment below.

2.6 Simulations

In this section, we report results from a simulation experiment to analyze the properties of the normal asymptotic intervals as well as their bootstrap counterparts analyzed above. The data-generating process is similar to the one used in Gonçalves and Perron (2014). We consider the single factor model :

$$y_{t+h} = .5F_t + \varepsilon_{t+h} \tag{2.25}$$

where F_t is an autoregressive process :

$$F_t = .8F_{t-1} + u_t$$

with u_t drawn from a normal distribution independently over time with a variance of $(1 - .8^2)$. We use the backward representation of this autoregressive process to make sure that all sample paths have $F_T = 1$. We will consider two forecasting horizons, $h = 1$ and $h = 4$.

The regression error ε_{t+h} will be homoskedastic with expectation 0, variance 1 and will have a moving average structure to accommodate multi-horizon forecasting :

$$\varepsilon_{t+h} = \sum_{j=0}^{h-1} .8^j v_{t+h-j},$$

and to analyze the effects of deviations from normality, we report results for two distributions for v_t :

$$\text{Normal}: \quad v_t \sim \left(\frac{1}{\sum_{j=0}^{h-1} .8^{2j}}\right) N(0,1)$$

$$\text{Mixture}: \quad v_t \sim \left(\frac{1}{\sum_{j=0}^{h-1} .8^{2j}}\right) \frac{1}{\sqrt{10}} \left[pN(-1,1) + (1-p) N(9,1)\right],$$

where p is distributed as $Bernoulli\,(.9)$. The particular mixture distribution we are using is similar to the one proposed by Pascual, Romo and Ruiz (2004). Most of the data is drawn from a $N(-1,1)$ but about 10% will come from a second normal with a much larger mean of 9. The scaling term in parentheses ensures that the variance of ε_{t+h} is 1 regardless of h. We have also considered other distributions such as the uniform, exponential, and χ^2 but do not report these results for brevity.

The $(T \times N)$ matrix of panel variables is generated as :

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

where λ_i is drawn from a $U[0,1]$ distribution (independent across i) and e_{it} is heteroskedastic but independent over i and t. The variance of e_{it} is drawn from $U[.5, 1.5]$ for each i.

We consider asymptotic and bootstrap confidence intervals at a nominal level of 95%. Asymptotic inference is conducted by using a HAC estimator (quadratic spectral kernel with bandwidth set to h) to account for possible serial correlation.

We use Algorithms 1 and 2 described above to generate the bootstrap data with $B = 999$ bootstrap replications. The idiosyncratic errors are always drawn using the wild bootstrap in step 1. In step 3, three bootstrap schemes are analyzed to draw ε_t^* : the first one draws the residuals with replacement in an i.i.d. fashion, the second one uses the wild bootstrap, while the last one redraws the residuals using the block wild bootstrap with a block size equal to h. The first two methods are only valid when $h = 1$, while the last one is valid for both values of h. In all applications of the wild bootstrap and block wild bootstrap, the external variable has a standard normal distribution. With the wild bootstrap, we use the heteroskedasticity-robust variance estimator, while we use the HAC one with block size equal to h for the block wild bootstrap.

We consider two types of bootstrap intervals : symmetric percentile-t and equal-tailed percentile-t. We report experiments based on 5,000 replications and with three values for T (50, 100, and 200) and 4 values for N (50, 100, 150, and 200).

We report results graphically for the conditional mean $y_{T+h|T}$ and for the new observation y_{T+h}. We report the frequency of times the 95% confidence interval is to the left or right of the true parameter. Each figure has three rows corresponding to $T = 50$, $T = 100$, and $T = 200$ with N on the horizontal axis, and in the last column, we show the average length of the corresponding confidence intervals relative to the length of the "ideal"confidence intervals obtained with the 2.5% and 97.5% quantiles from the empirical distribution simulated for each N and T 1,000,000 times as endpoints. To keep the figures readable, we report results for two bootstrap methods in each figure. For the conditional mean, we report results for the wild bootstrap and block wild bootstrap (with differences thus only coming from the block size since the two methods are the same for a block size equal to 1). For the observation, we report results using the iid and block wild bootstrap since we require an i.i.d. assumption for the construction of intervals for this quantity.

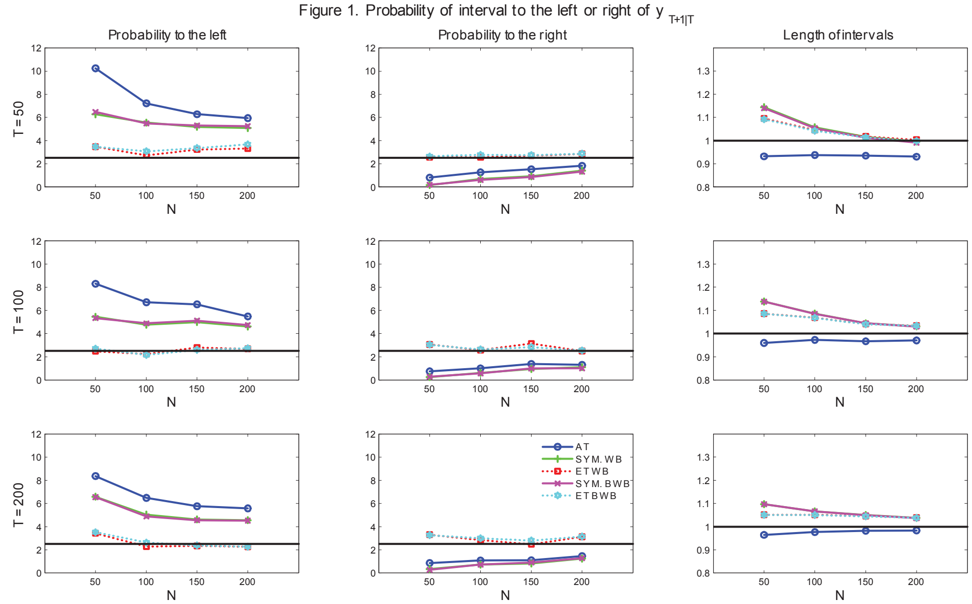
It turns out that the distribution of ε_{T+h} noticeably affects the results for y_{T+h} only. As a consequence, we only report results with Gaussian ε_{t+h} for the conditional mean. On the other hand, the results of y_{T+h} are dominated by the behavior of ε_{t+h}. Thus, the contribution of the conditional mean from the contribution of ε_{T+h} in the forecasts of y_{T+h} are clearly separated.

2.6.1 Forecasting horizon $h = 1$

We start by presenting results when we are interested in making a prediction for next period's value. For this horizon, because ε_{t+1} does not have serial correlation, the wild bootstrap and block wild bootstrap methods are identical with reported differences due to simulation error.

Conditional mean, $y_{T+1|T}$ The results for the conditional mean are presented in Figure 1. Asymptotic theory (blue line) shows large distortions that decrease with an increasing N. For example, for $N = T = 50$, the 95% confidence interval does not include the true mean in 11% of the replications instead of the nominal 5%. This number is reduced to 7.8% when $N = 200$ and $T = 50$. Moreover, we see that most of these instances are in one direction, when the confidence interval is to the left of the true value. This can be explained by a bias in the estimation of the parameter δ as documented by Gonçalves and Perron (2014) due to the estimation uncertainty in the factors. This bias is negative, thus shifting the distribution of the conditional mean to the left, leading to more rejections on the left side and fewer on the right side than predicted by the asymptotic normal distribution.

Figure 1. Probability of interval to the left or right of $y_{T+1|T}$



Note : The figures in the first two columns report the fraction of confidence intervals that lie to the left or to the right of the conditional mean for each method as a function of the cross-sectional dimension N. Each row corresponds to a different time series dimension. The last column reports the length of the confidence intervals relative to the length of the "ideal" intervals obtained as the 2.5% and 97.5% quantiles of the empirical distribution.

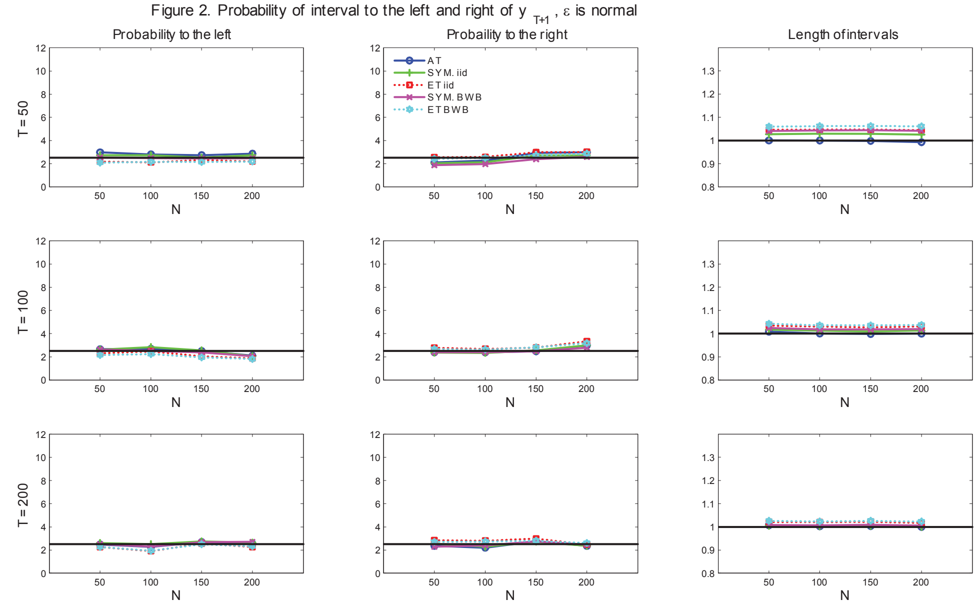
The presence of bias is reflected in the bootstrap distribution of $\hat{y}^*_{T+1|T}$ which is also shifted to the left. This is illustrated by a large difference between the bootstrap symmetric and equal-tailed intervals. The symmetric intervals reproduce the pattern of more coverage to the left than to the right, while equal-tailed intervals distribute coverage more or less equally in both tails. In both cases, the total rejection rates are closer to their nominal level than with asymptotic theory, for example with $N = T = 50$, the wild bootstrap does not include the true value in 6.7% of the replications with the symmetric intervals and 6.1% for the equal-tailed intervals.

This phenomenon is also reflected in the length of intervals. The asymptotic intervals are shortest (and least accurate). The equal-tailed intervals are typically slightly shorter than the corresponding symmetric intervals.

Forecast of y_{T+1} We next consider the prediction of y_{T+1} in Figures 2 and 3. As mentioned before, given our parameter configuration, the uncertainty is dominated by the underlying error term ε_{T+1} and not estimation uncertainty. This is the reason asymptotic intervals rely on the normality assumption. This provides a motivation for the bootstrap, and the effect of non-normality is highlighted in our figures.

Figure 2 shows that under normality, inference for y_{T+1} is quite accurate for all methods, and it is essentially unaffected by the values of N and T as predicted since it is dominated by the behavior of ε_{t+h}. All methods perform similarly, though we see that the asymptotic intervals that make the correct Gaussianity assumption are shorter than those based on the bootstrap. The iid bootstrap also produces slightly narrower intervals than the block wild bootstrap.

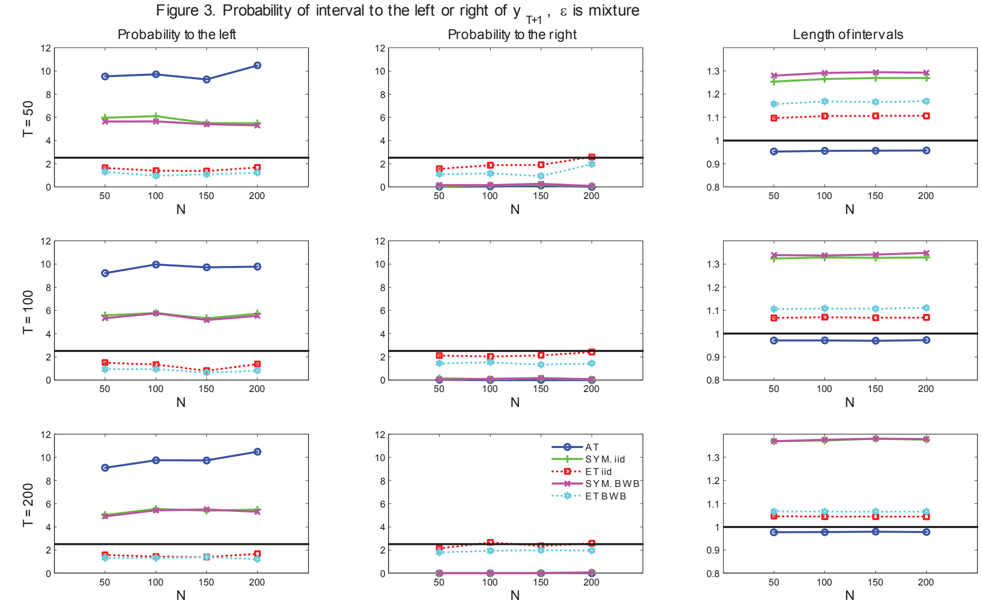
Figure 2. Probability of interval to the left and right of y_{T+1}, ε is normal



Note : The figures in the first two columns report the fraction of confidence intervals that lie to the left or to the right of the observation for each method as a function of the cross-sectional dimension N. Each row corresponds to a different time series dimension. The last column reports the length of the confidence intervals relative to the length of the "ideal" intervals obtained as the 2.5% and 97.5% quantiles of the empirical distribution.

Figure 3 provides the same information when the errors are drawn from a mixture of normals. We see problems with asymptotic theory, and these come almost exclusively in the form a confidence interval to the left of the true value. This is due to the fact that we have falsely imposed that errors are Gaussian, whereas the true distribution is bimodal. On the other hand, the bootstrap corrects these difficulties. The symmetric intervals do so by reducing coverage on the left side to between 5 and 6% and having almost no coverage to the right. The equal-tailed intervals distribute coverage more evenly by reducing undercoverage on the right side and pretty much eliminating the over-coverage on the left side. Because they allow for asymmetry, the equal-tailed intervals are shorter than the symmetric ones. Similarly, the i.i.d. bootstrap that makes the correct assumption that ε_{T+1} is i.i.d. produces slightly more accurate coverage and shorter intervals.

Figure 3. Probability of interval to the left or right of y_{T+1}, ε is mixture



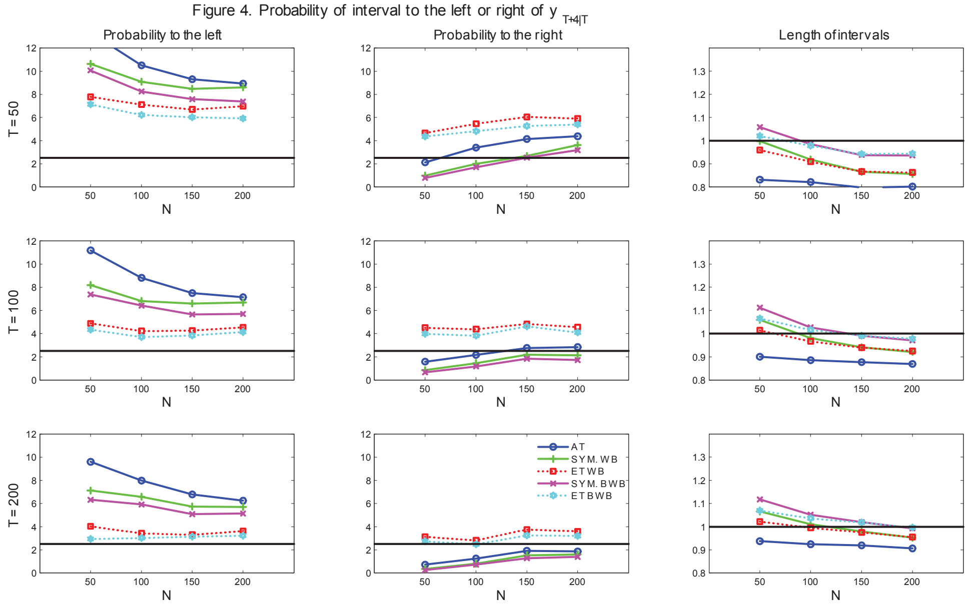
Note : see Figure 2.

2.6.2 Multi-horizon forecasting

In Figures 4-6, we report the same results as before but for $h = 4$ instead of $h = 1$. Because the error term is now a moving average of order 3, the wild bootstrap and block wild bootstrap (a block size equal to 4 is used) are no longer identical.

Figure 4 reports the results for the conditional mean, $y_{T+4|T}$. The main difference with Figure 1 is that there is a gap between the accuracy of the intervals based on the wild bootstrap and on the block wild bootstrap. As before, the equal-tailed intervals provide more accurate intervals and smaller length because they capture the bias in the distribution.

Figure 4. Probability of interval to the left or right of $y_{T+4|T}$

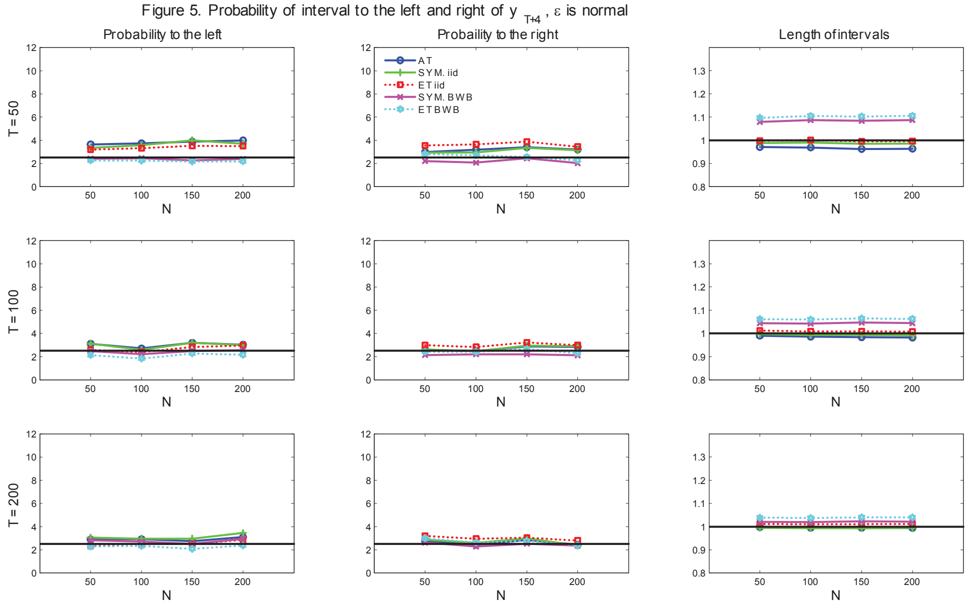


Note : see Figure 1.

While there is a difference in coverage between the wild bootstrap and the block wild bootstrap, it is not very large. This feature can be explained by the fact that factors are estimated. The forecast error variance has two parts, one due to the estimation of the parameters and one due to the estimation of the factors (see equation (2.4)). Serial correlation only affects the first term in that expression, and thus its effect is dampened by the presence of the second term which is usually not present in a typical forecasting context where predictors are observed.

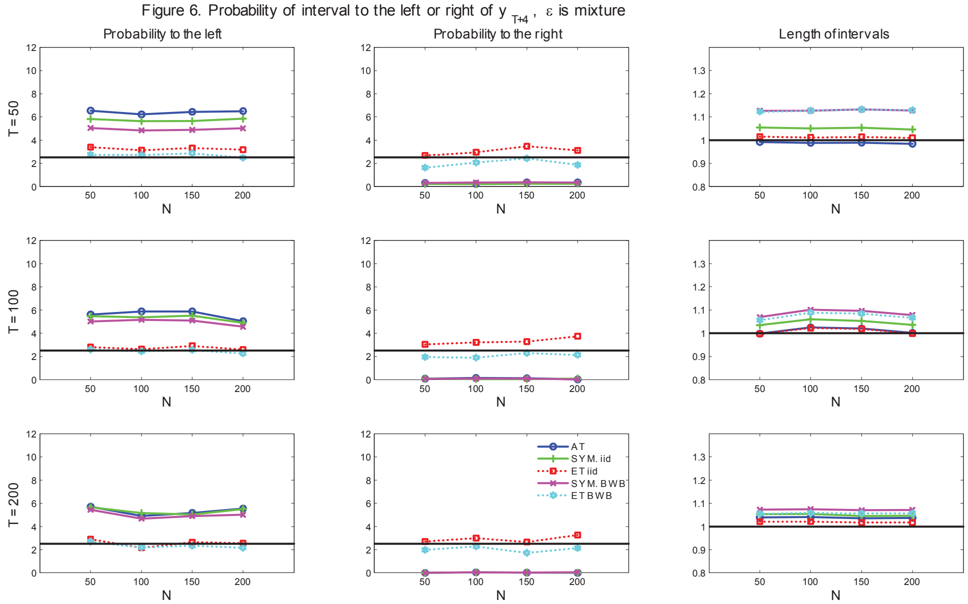
Figures 5 and 6 give the results for the new observation, y_{T+4}. Overall, we see that serial correlation does not seem to affect inference on y_{T+h} much. There are some effects when $T = 50$, but this seems related to difficulties in estimating the distribution of ε_{T+4} with serial correlation. Otherwise, the figures and conclusions are similar to those in Figures 2 and 3 with the exception of the fact that the block wild bootstrap leads to much wider intervals than the i.i.d. bootstrap with some improvement in coverage for $T = 50$.

Figure 5. Probability of interval to the left and right of y_{T+4}, ε is normal



Note : see Figure 2.

Figure 6. Probability of interval to the left or right of y_{T+4}, ε is mixture



Note : see Figure 2.

2.7 Empirical illustration

In this section, we use the dataset of Stock and Watson (2003) and Rossi and Sekhposyan (2014), updated to the first quarter of 2014, to illustrate the properties of asymptotic and bootstrap intervals.[1]

[1] We thank Tatevik Sekhposyan for providing us with the data.

We consider forecast intervals for changes in the inflation rate measured by the quarterly growth rate of the GDP deflator ($PGDP$) at annual rate :

$$\Delta\pi_t = \left[\ln\left(\frac{PDGP_t}{PGDP_{t-1}}\right) - \ln\left(\frac{PDGP_{t-1}}{PGDP_{t-2}}\right)\right] \times 400.$$

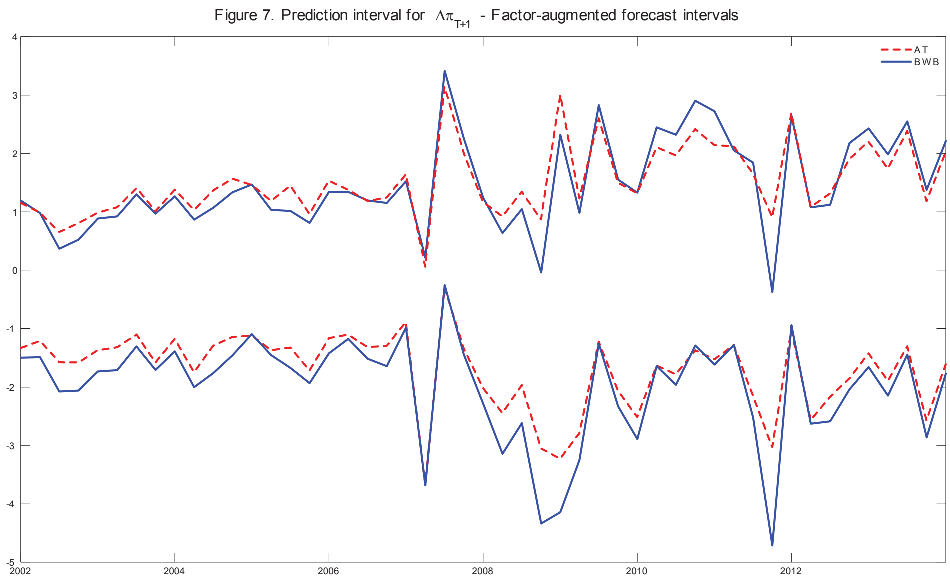
There is a total of $N = 29$ series on asset prices, measures of economic activity, wages and prices, and money used to construct forecasts, see Rossi and Sekhposyan (2014) for details. The inflation rate is not included in the data used for extracting the factors. In order to have a balanced panel, our sample covers the period 1973q1-2014q1.

We construct forecasts from the factor-augmented autoregressive model :

$$\Delta\hat{\pi}_{t+h} = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\phi}_j \Delta\pi_{t-j+1} + \sum_{j=1}^{r} \hat{\alpha}_j \tilde{F}_{j,t}.$$

We compute forecast intervals for $h = 1$ for the last 50 observations in the sample. This means that the forecasts are made each period from the third quarter of 2001 until the end of 2013. We use a rolling window of 40 observations to estimate factors and parameters as in Rossi and Sekhposyan. We also follow Rossi and Sekhposyan and first choose the AR order p for each time period using BIC and then augment with the estimated factors. In each period, we select the number of factors such that the factors explain a minimum of 60% of the total variance of the panel after centering and rescaling. Three factors are selected by this approach in 40 out of the 50 periods, and 4 for the remaining 10 periods.

The factor-based forecasts reduce the root mean squared error of the forecasts by about 13% relative to autoregressive forecasts. In Figure 7, we report prediction intervals for the factor-augmented forecasts. The dashed red lines represent the bounds of the (pointwise) 95% prediction interval based on the asymptotic theory of Bai and Ng (2006) for each date. This interval is symmetric around the point forecast by construction since it is based on the normal distribution. We also report bootstrap intervals based on the block wild bootstrap (BWB) for ε^*_{t+h} with block size equal to the bandwidth selected by the Andrews (1991) rule and the wild bootstrap for e^*_t. Other methods for drawing ε^*_{t+h} lead to very similar intervals, and we do not report them to ease exposition (they are available from the authors upon request). The reported intervals were constructed as equal-tailed percentile-t intervals and are based on $B = 9999$ bootstrap replications.

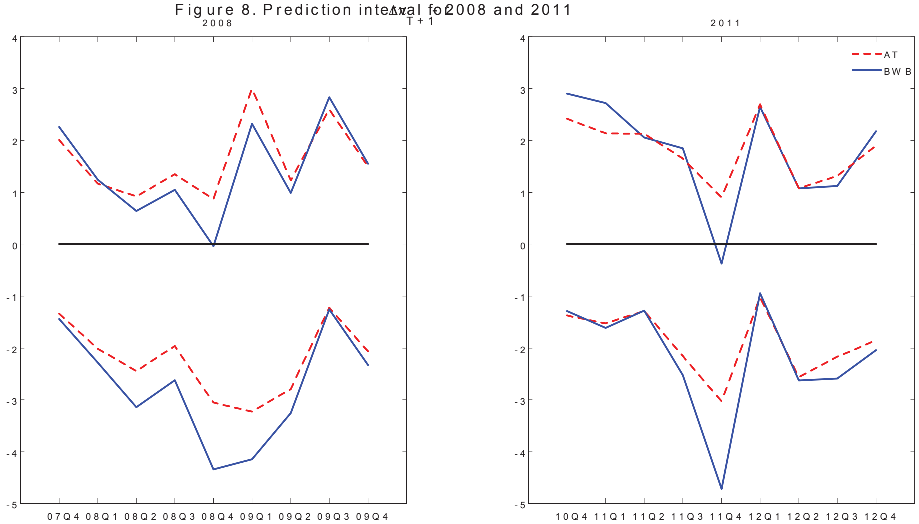


Note : The dashed red lines represent the bounds of the (pointwise) 95% prediction interval based on the asymptotic theory of Bai and Ng (2006) for each date. The solid blue line are bounds of the 95% equal-tailed percentile-t bootstrap intervals based on the block wild bootstrap (BWB) with block size equal to the bandwidth selected by the Andrews (1991) based on B=9999 bootstrap replications.

While both sets of intervals in Figure 7 are similar, there are noticeable differences that can be attributed either to bias in the estimation of the parameters or to non-normality in the distribution of the error term. Rossi and Sekhposyan (2014) find fairly strong evidence of non-normality of the forecast errors for this series, and this is likely an important source of the differences between the asymptotic and equal-tailed intervals.

The behavior of the bootstrap intervals during specific periods is quite interesting. For example, early in the sample, the bootstrap intervals are consistently shifted down relative to the asymptotic intervals. Figure 8 highlights two periods where the bootstrap interval lies completely below 0. The left panel presents the same intervals as Figure 7 around the fourth quarter of 2008. We see that the bootstrap intervals are shifted down for most of the reported period, and the upper limit of the bootstrap interval drops just below 0 (it is -.07%) in the fourth quarter of 2008. On the other hand, the asymptotic interval contains 0 with an upper limit of about 1%. This means that policy makers concerned about a sudden reduction in inflation following the collapse of Lehman Brothers would have underestimated the probability of a reduction in inflation had they based their decision on the asymptotic intervals.

Figure 8. Prediction interval for 2008 and 2011



Note : see Figure 7.

Similarly, the right panel of Figure 8 focuses on the intervals around the fourth quarter of 2011. As in 2008, the bootstrap interval for the fourth quarter of 2011 is shifted down and includes only negative values, whereas the corresponding asymptotic interval includes positive inflation changes. At that time, many central banks were concerned about deflation risk, and relying on asymptotic intervals would have given them the impression that large reductions in inflation were much less likely than suggested by the bootstrap interval (the change in inflation turned out to be -2 percentage points).

2.8 Conclusion

In this paper, we have proposed the bootstrap to construct valid prediction intervals for models involving estimated factors. We considered two objects of interest : the conditional mean $y_{T+h|T}$ and the realization y_{T+h}. The bootstrap improves considerably on asymptotic theory for the conditional mean when the factors are relevant because of the bias in the estimation of the regression coefficients. However, our simulation results suggest that allowing for serial correlation, as is relevant when the forecasting horizon is greater than 1, is not very important in practice. For the observation, the bootstrap allows the construction of valid intervals without having to make strong distributional assumptions such as normality as was done in previous work by Bai and Ng (2006) .

One key assumption that we had to make to establish our results is that the idiosyncratic errors in the factor models are cross-sectionally independent. This is certainly restrictive, but it allows for the use of the wild bootstrap on the idiosyncratic errors. Non-parametric bootstrapping under more general conditions remains a challenge. The results in this paper could be used to prove the validity of a scheme in that context by showing the conditions $\mathcal{A}$ and $\mathcal{B}$ are satisfied.

Chapitre 3

Model Selection in Factor-Augmented Regressions with Estimated Factors

$\subset$

3.1 Introduction

Factor-augmented regression (FAR) models are now widely used for generating forecasts since the seminal paper of Stock and Watson (2002) on diffusion indices. Unlike the traditional regressions, these models allow the inclusion of a large set of macroeconomic and financial variables as predictors, useful to span various information sets related to economic agents. Thereby, economic variables are considered as driven by some unobservable factors which are inferred from a large panel of observed data. Many empirical studies have been conducted using FAR. Among others, Stock and Watson (2002) forecast the inflation rate assuming some latent factors explain the comovement in their high dimensional macroeconomic data set. Furthermore, Ludvigson and Ng (2007) look at the risk-return relation in the equity market. From eight estimated factors resuming the information in their macro and financial data sets using Bai and Ng (2002) IC_{p2} criterion, they identify, based on the bayesian information criterion (BIC), three new factors termed "volatility", "risk premium" and "real" factors that predict future excess returns.

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

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

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

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

The simulations show that the leave-one-out cross-validation often selects a model larger than the true one while the modified BIC of Groen and Kapetanios (2013) tends to under-parameterize for smaller sample sizes. Nevertheless, the Monte Carlo leave-d-out cross-validation and the bootstrap selection pick with higher probability the estimated factors spanning the true factors. To illustrate the methods, an empirical application that revisits the relationship between macroeconomic and financial factors, and excess stock returns for the U.S. market have been conducted. The factors are extracted from 147 financial series and 130 macroeconomic series. The financial series correspond to the 147 variables in Jurado, Ludvigson and Ng (2015). The quarterly macroeconomic data set is constructed following McCracken and Ng (2015) and spans the first quarter of 1960 to the third quarter of 2014. After controlling for the consumption-wealth variable (Ludvigson and Lettau, 2001), the lagged realized volatility of the future excess returns and other factors, among the estimated factors from a large panel of U.S. macro and financial data, the factors heavily correlated with interest rate spreads and with the Fama-French factors have strong additional predictive power for excess returns. The out-of-sample performance for predicting excess returns with the new procedures is also compared to existing model selection ones.

This paper is organized as follows. In Section 3.2, we present the settings and assumptions. Section 3.3 addresses model selection. Section 3.4 reports the simulation study, and the section 3.5 presents the empirical application. The section 3.6 concludes. Mathematical proofs are in the Appendix 0.3 and the empirical application details in the Appendix 0.4.

3.2 Settings and assumptions

In this set-up, the econometrician has an information set up to time T i.e. $(y_t,\ W_t',\ X_t')_{t=1,\cdots,T}$ and his goal is to predict y_{T+1} with the following factor-augmented regression model

$$y_{t+1} = \delta' Z_t^0 + \varepsilon_{t+1},\ t = 1, ...,\ T-1 \tag{3.1}$$

where $Z_t^0 = \left(F_t^{0\prime},\ W_t'\right)'$. The $q \times 1$ vector W_t contains some observed regressors and F_t^0 represents r_0 factors not observed by the econometrician. These latent factors F_t^0 are a subset of the unobserved factors $F_t : r \times 1$ driven the large dimensional matrix X following a factor panel model specification given by

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

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

$$y_{t+1} = \alpha\,(m)'\,\widetilde{F}_t\,(m) + \beta' W_t + u_{t+1}\,(m) = \delta\,(m)'\,\hat{Z}_t\,(m) + u_{t+1}\,(m) \tag{3.2}$$

where m is any of the 2^r subsets of $\{1,\ \ldots,\ r\}$ denoted $\mathcal{M}$ including the empty set where no factor drives y. The size of $\widetilde{F}_t\,(m)$ is $r\,(m) \leq r$ and we assume the number of estimated factors selected in the first step known and equal to r. While Kleibergen and Zhan (2015) guide against the harmful effect of under parameterizing on the true R^2 and test statistics, Kelly and Pruitt (2015) correct for forecast using irrelevant factors by suggesting a three-pass regression filter procedure. Carrasco and Rossi (2015) recently develop regularization methods for in-sample inference and forecasting in misspecified factor models. Cheng and Hansen (2015) also study forecasting using a frequentist model averaging approach. However, none of these papers study the consistent estimation of the true latent factors space in order to predict y, based on the commonly used ordinary least squares of FAR with principal components. Although there is a large body of literature on selecting the number of factors that resume the information in the factor panel data set, 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. We denote $Z_t = \left(F_t', W_t'\right)'$, $t = 1, ..., T$, $\|M\| = \left(trace\,(M'M)\right)^{1/2}$ the Euclidean norm, $M > 0$ the positive definiteness for any square matrix M, and C a generic finite constant. The following standard assumptions are made.

Assumption A1 (factor model and idiosyncratic errors)

(a) $E\left\|F_t\right\|^4 \leq C$ and $\Sigma_F = \lim_{T\rightarrow\infty} E\left(\frac{1}{T}F'F\right) > 0$ where $F = \left(F_1, \cdots, F_T\right)'$.

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

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

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

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

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

Assumption A2 (moments and weak dependence among $\{z_t\}$, $\{\lambda_i\}$, $\{e_{it}\}$ and $\{\varepsilon_{t+1}\}$)

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

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

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

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

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

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

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

Assumption A3 (moments and CLT for the score vector)

(a) $E\left(\varepsilon_{t+1}|\mathcal{F}_t\right) = 0$, $E\left(\varepsilon_{t+1}^2|\mathcal{F}_t\right) = \sigma^2$, $E\left\|Z_t\right\|^4 < C$ and $E\left(\varepsilon_{t+1}^4\right) < C$ where

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

(b) $\Sigma_{ZZ} = \lim_{T\longrightarrow\infty} E\left(\frac{1}{T}\sum_{t=1}^{T}Z_tZ_t'\right) > 0$.

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

A1 and A2 are the same as in Bai and Ng (2002), Gonçalves and Perron (2014) and Cheng and Hansen (2015) in terms of factor-augmented regression specifications that allow for weak dependence and heteroscedasticity in the idiosyncratic errors. Assumption A3 is useful for deriving the asymptotic distribution of the estimator $\hat{\delta}$ of δ. It assumes that the forecast error is conditionally homoscedastic which is rather strong.

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

$$H = \tilde{V}^{-1} \frac{\tilde{F}'F}{T} \frac{\Lambda'\Lambda}{N}, \tag{3.3}$$

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

3.3 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 (3.2). In practice, we extract estimated factors $\tilde{F}_t$ which summarize information in the large $N \times T$ matrix X. Afterwards, a subset $\tilde{F}_t(m)$ is chosen for the prediction of y_{t+1}. Ludvigson and Ng (2007) select $\tilde{F}_t(m)$ using the bayesian information criterion (BIC) to predict excess stock returns. Because this criterion does not correct for factor estimation, Bai and Ng (2009) suggest a modified final prediction error (FPE) with an extra penalty to proxy the effect of factor estimation by approximating the mean squared error. 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. Given the information up to time t, the true conditional mean is

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

In the usual case with observed factors, Shao (1997) defines a model m as correct if its conditional mean equals that of the true unknown model i.e.

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

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

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

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 procedures using cross-validation and bootstrap procedures.

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.

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

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

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

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

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

However, by minimizing the CV_1, there is a positive probability that we do not select the smallest possible correct set of generated regressors. As it is next shown based on Lemma 3.1, this positive probability to select a larger correct model 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 to the space spanned by $Z(m) = (F(m), W)$ with $F(m)$ the generic limit of $\widetilde{F}(m)$ and $\mu = Z^0\delta$ the true conditionnal mean vector.

Lemma 11. *Under Assumptions A1-A3 and assuming*

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

for all m, *as* T, $N\to\infty$, *for* $m\in\mathcal{M}_2$,

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

where $V_T(m) = O_p\left(\frac{1}{C_{NT}^2}\right)$, *and for* $m\in\mathcal{M}_1$,

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

From Lemma 11, a for correct model,

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

otherwise

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

Lemma 11 contains the equations (3.5) and (3.6) of Shao (1993) when the factors are not observed but estimated. Contrary to that case where the regressors are observed, we have an additional term $V_T(m)$ corresponding to its factor estimation error, and $P(m)$ is associated to the space spanned by subsets of FH_0' a rotation of the true factor space. Consider two candidates estimated models m_1 and m_2. Suppose that m_1 is correct and m_2 is incorrect. Assume $p\lim\inf_{T\to\infty} \frac{1}{T-1}\mu'(I - P(m))\mu > 0$ for incorrect models. The CV_1 will prefer m_1 to m_2 since

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

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

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

with $V_T(m) = O_p\left(\frac{1}{C_{NT}^2}\right)$. The first term is independent of candidate set of estimated factors. The second term represents the complexity of the model as it increases with the model dimension. The term in parenthesis is a parameter estimation error with mean zero while comparing two competing correct models. The term $V_T(m)$ contains the factor estimation error in the $CV_1(m)$ which is not reflected by the term in parentheses. Because the complexity component is inflated in finite samples not only by this parameter estimation error but also the factor estimation one, we fail to pick accurately the smallest correct model. 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, a sufficient condition for a consistent estimated model selection is its ability to capture the complexity term useful to penalize the over-fitting.

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

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

We analyze through a simulation study how V_T contributes to worsen the probability of selecting a consistent set of estimated factors.

We consider the same data generating process (DGP) as the first DGP in the simulation section where $y_{t+1} = 1 + F_{1t} + 0.5F_{2t} + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim N(0,1)$ and $F^0 = (F_1,\ F_2) \subset F = (F_1,\ F_2,\ F_3,\ F_4)$. Given the specification for the latent factors and the factor loadings, the PC1 condition for identifying restrictions provided by Bai and Ng (2013) is asymptotically satisfied and makes possible to identify estimated factors. Hence, we extract four estimated factors and we expect to pick consistently the first two among the $2^4 = 16$ possibilities. The line "with parameter and factor estimation errors" on Figure 9 reports the frequency of selecting a larger set of estimated factors while minimizing the CV_1 criterion which includes the estimation errors.

Given the different sample sizes, it turns that the leave-one-out cross-validation selects very often a larger set of generated 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 forecast error in the leave-one-out cross-validation

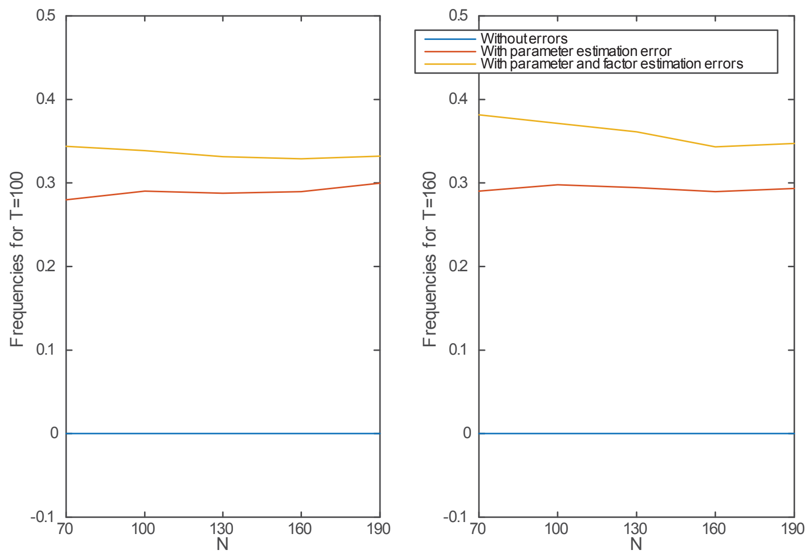
$$CV_{11}(m) = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{(r(m)+q)}{T-1}\sigma^2 + \frac{1}{T-1}\mu'(I-P(m))\mu$$

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

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

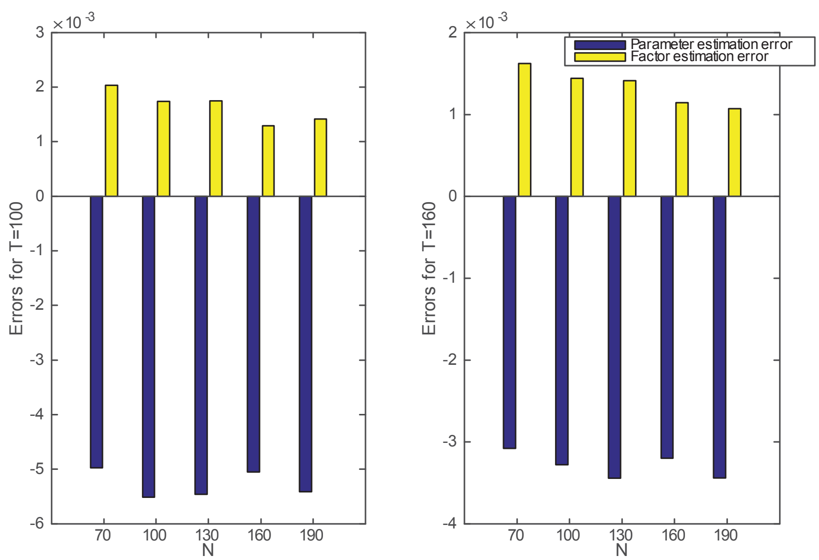
Once the parameter estimation error is included, the frequency of selecting a larger model increases. Moreover, when we include both the parameter and the factor estimation errors corresponding to the CV_1, that frequency increases more. The results show that this factor estimation error while asymptotically 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 (Figure 10) which is followed by a drop of the probabilities of over-parameterization.

Figure 9. Frequencies of selecting a larger set of estimated factors minimizing the CV_1, the CV_1 without the factor estimation error and the CV_1 without the parameter and the factor estimation errors over 10,000 simulations



Note : It reports the frequencies of selecting a larger set than first two estimated factors. The line blue represents the frequencies while minimizing the complexity component and the identifiability one plus $\frac{1}{T-1}\varepsilon'\varepsilon$. The orange and red corresponds respectively to the frequency when the parameter estimation error is added, and the "naive" leave-one-out cross-validation which includes both the parameter and the factor estimation errors.

Figure 10. Average parameter estimation error and factor estimation error in the CV_1



Note : This figure shows the average parameter and factor estimation errors in the leave-one-out cross-validation criterion as N and T vary over the simulations.

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

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

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

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

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

$$y_s = \begin{pmatrix} y_{s(1)} \\ y_{s(2)} \\ \vdots \\ y_{s(d)} \end{pmatrix}, \quad \hat{Z}_s(m) = \begin{pmatrix} \widetilde{F}'_{s(1)}(m) & W_{s(1)\cdot 1} & \cdots & W_{s(1)\cdot q} \\ \widetilde{F}'_{s(2)}(m) & W_{s(2)\cdot 1} & & W_{s(2)\cdot q} \\ \vdots & \vdots & & \vdots \\ \widetilde{F}'_{s(d)}(m) & W_{s(d)\cdot 1} & \cdots & W_{s(d)\cdot q} \end{pmatrix}.$$

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

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

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

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

for correct m and $CV_d(m) = \sigma^2 + \frac{1}{T-1}\mu'(I-P(m))\mu + o_p(1)$ for incorrect m. Hence for m_1 and m_2 correct estimated models such that $r(m_1) < r(m_2)$,

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

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

Assumption A4

(a) $p\lim\inf_{T\to\infty}\frac{1}{T-1}\mu'(I-P(m))\mu > 0$ for any $m \in \mathcal{M}_1$.

(b) $p\lim_{T\to\infty}\sup_{1\leq t\leq T-1}\left|Z_t(m)'\left[Z(m)'Z(m)\right]^{-1}Z_t(m)\right|=0$ for all m.

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

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

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

Assumption A4 (a) is an identifiability assumption in order to distinguish a correct set from an incorrect one. Groen and Kapetanios (2013) also made this assumption. By A4 (b), for any model, the diagonal elements of the projection matrix vanish asymptotically. This regularity condition can be seen as a form of a stationarity assumption for regressors in the different sub-model, which is typical in cross-validation literature. Assumption A4 (c) argues that the average difference between the Fisher information matrix of the validation and the construction samples are close as $N,T\to\infty$. A4 (d) complements A1 (e) as when $s^c=\{1,\ldots,T-1\}$, $\frac{1}{\sqrt{T\cdot\kappa}}\sum_{t\,\in\,s^c}\sum_{u=1}^{T}\frac{1}{N}\sum_{i,j}|\sigma_{ij,tu}|=\frac{1}{TN}\sum_{t,u,i,j}^{T}|\sigma_{ij,tu}|\leq C$. Assumption A4 (e) strengthen A2 (d) and is used for proving Lemma 18. The next theorem proves the consistency of the Monte Carlo leave-d-out cross-validation for FAR.

Theorem 12. *Under Assumptions A1, A2, A3 and A4 such that* $\frac{\kappa}{C_{NT}^2}\to 0$, $\frac{T^2}{\kappa^2 b}\to 0$, $\frac{T}{\kappa^2}\to C<\infty$ *and* κ, $d\to\infty$, *when* b, T, $N\to\infty$,

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

where

$$\hat{m}=\arg\min_m CV_d(m)$$

if $\mathcal{M}$ *contains at least one correct model.*

The proof of Theorem 12 is given in Appendix. This result is an extension of Shao (1993) to the case with generated regressors. Given the rate conditions, κ, $d\to\infty$ such that $\frac{\kappa}{T-1}\to 0$ and $\frac{d}{T-1}\to 1$. It follows from Theorem 12 that the consistency of the Monte Carlo leave-d-out cross-validation relies on κ much smaller than d. One could consider $\kappa=\min\{T,\ N\}^{3/4}$ and $d=(T-1)-\kappa$ as they are consistent with the conditions in Theorem 12. In particular, Shao (1993) suggests for the observed regressors framework $\kappa=T^{3/4}$. This difference is due to the presence of the factor estimation error which should converge faster to zero relative to the complexity term. An extreme case where this condition is not satisfied is the leave-one-out cross-validation where $\kappa=(T-1)-1$ and $d=1$.

The next paragraph studies an alternative selection procedure using the bootstrap methods.

3.3.2 Bootstrap rule for model selection

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

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

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

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

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

where

$$\hat{\delta}_\kappa^*(m) = \left(\hat{Z}^{*\prime}(m)\hat{Z}^*(m)\right)^{-1}\hat{Z}^{*\prime}(m)y^*(m) \tag{3.4}$$

with $\hat{Z}^*(m)$, $y^*(m)$ the bootstrap analog of $\hat{Z}(m)$, $y(m)$ obtained through the following Algorithm 3.3.

Algorithm 3.3

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

B) For each m :

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

3. For each bootstrap sample extract $\widetilde{F}^*$ from X^* and estimate $\hat{\delta}^*_\kappa(m)$ based on $\hat{Z}^* = \left(\widetilde{F}^*, W\right)$ and $y^*(m)$ using (3.4).

C) Obtain $\hat{m}$ as the model that minimizes the average of $\hat{\Gamma}_\kappa(m)$ over the B samples.

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

Theorem 13. *Under the Assumptions A1-A3. Suppose further that Assumptions 6-8 of Gonçalves and Perron (2014) and* $E^*|\eta_{it}|^4 \le C < \infty$ *hold. If* N, $T \longrightarrow \infty$ *and* $\kappa \longrightarrow \infty$ *such that* $\frac{\kappa}{C^2_{NT}} \longrightarrow 0$ *then*

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

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

From Theorem 13, it follows that $\hat{\delta}^*_\kappa(m)$ converges to the limit of $\Phi^*_0(m)\hat{\delta}(m)$ at a lower rate $\sqrt{\kappa} = o\left(\sqrt{T}\right)$. The proof in Appendix shows that our bootstrap scheme satisfies the high level conditions provided by Gonçalves and Perron (2014). This result allows us to use our new bootstrap scheme for the following optimality results.

Theorem 14. *Suppose that Assumptions A1, A2, A3, A4 (a) complemented by Assumptions 6-8 of Gonçalves and Perron (2014) hold. Suppose further* $\kappa \to \infty$ *such that* $\frac{\kappa}{C^2_{NT}} \to 0$ *as* T, $N \to \infty$ *and* $E^*|\eta_{it}|^4 \le C < \infty$ *then if* $\mathcal{M}_2$ *is not empty then*

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

where

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

This bootstrap result is the analog of Theorem 11. The following section compares the different procedures through a simulation study.

3.4 Simulation experiment

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

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

where $\alpha_0 = 1$, $F^0_t \subset F_t \sim i.i.d.N(0, I_4)$ and $\varepsilon_{t+1} \sim i.i.d.N(0,1)$. Three data generating process (DGP) are used.

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

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

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

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

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

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

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

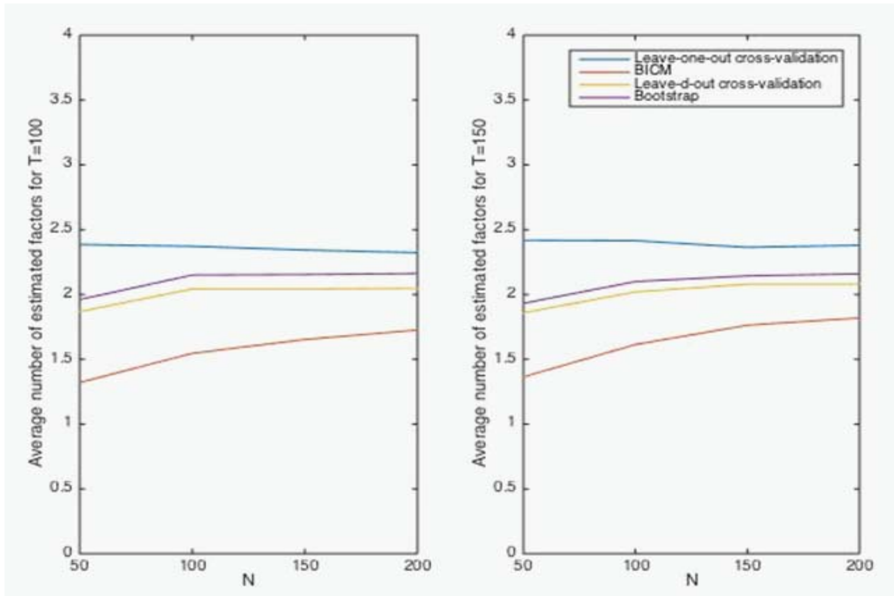
$$BICM(m) = \frac{T}{2} \ln\left(\hat{\sigma}^2(m)\right) + r(m) \ln(T) \left(1 + \frac{T}{N}\right) \text{ where } \hat{\sigma}^2(m) = \frac{1}{T-1-r(m)} \left\| y - \widetilde{F}(m) \hat{\delta}(m) \right\|^2,$$

is made by considering subsets of the first four principal component estimated factors.

Figure 11, 12 and 13 present the average number of selected estimated factors whereas Figure 14, 15 and 16 show the frequencies of selecting the true model over the $2^4 = 16$ possibilities including the case of no factor. Except for the largest estimated model, where the average number of estimated factors tends to be close to four, the CV_1 tends to overestimate the true number of factors. The BICM very often selects a smaller set of estimated factors than the true one. The leave-d-out cross-validation and the bootstrap procedure select in average a number of factor close to the true number.

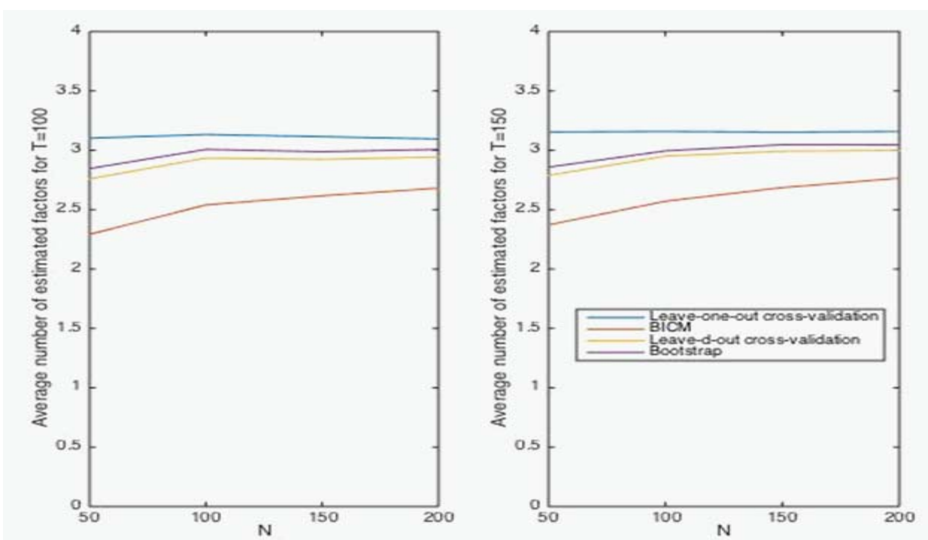
The suggested procedures offer a higher frequency of selecting factor estimates that span the true model for DGP 1 and 2. In particular, when $N = T = 100$, for DGP 1, the frequency of selecting the first two estimated factors is 55.3% using the modified BIC and 64.5% using the leave-one-out cross-validation. The bootstrap selection increases the frequency of the CV_1 by 13.6 points of percentage and the CV_d increases it by 18.5 points of percentage.These frequencies increase with the sample sizes. In general, the delete-one cross-validation very often selects a larger model than the true one and the modified BIC tends to pick smaller subset of the consistent model. As DGP 3 corresponds to the largest model, CV_1 unsurprisingly performs well.

Figure 11. Average number of estimated factors for DGP 1



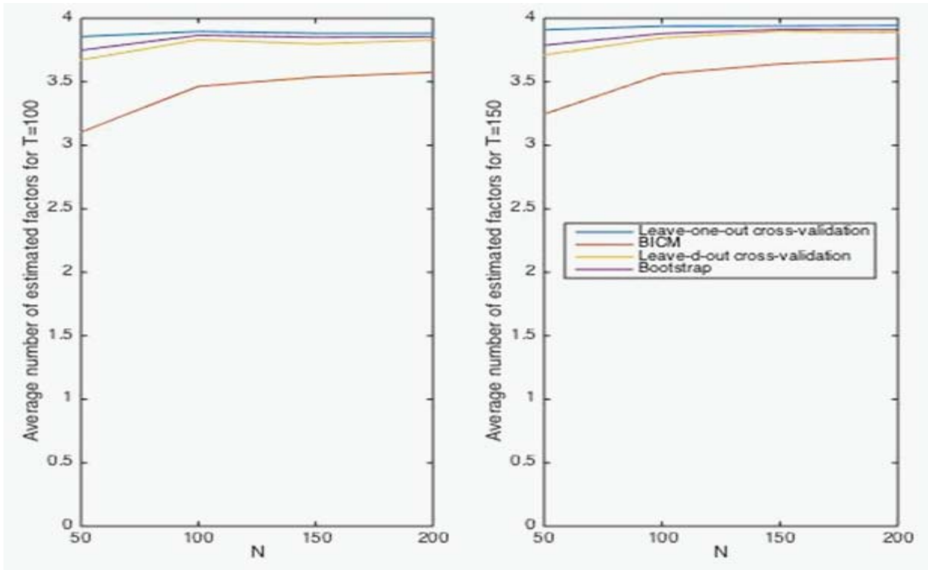
Note : This figure plots the average number of selected estimated factors over 1000 simulations. There are $2^4 = 16$ possible subsets of the larger one containing all factors. BICM relates to the modified BIC of Groen and Kapetanios (2013).

Figure 12. Average number of estimated factors for DGP 2



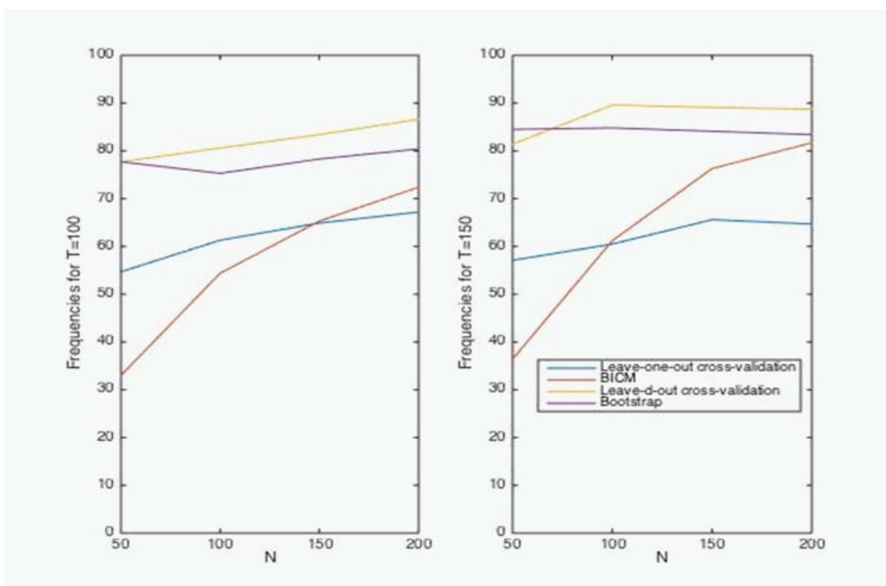
Note : see Figure 11.

Figure 13. Average number of estimated factors for DGP 3



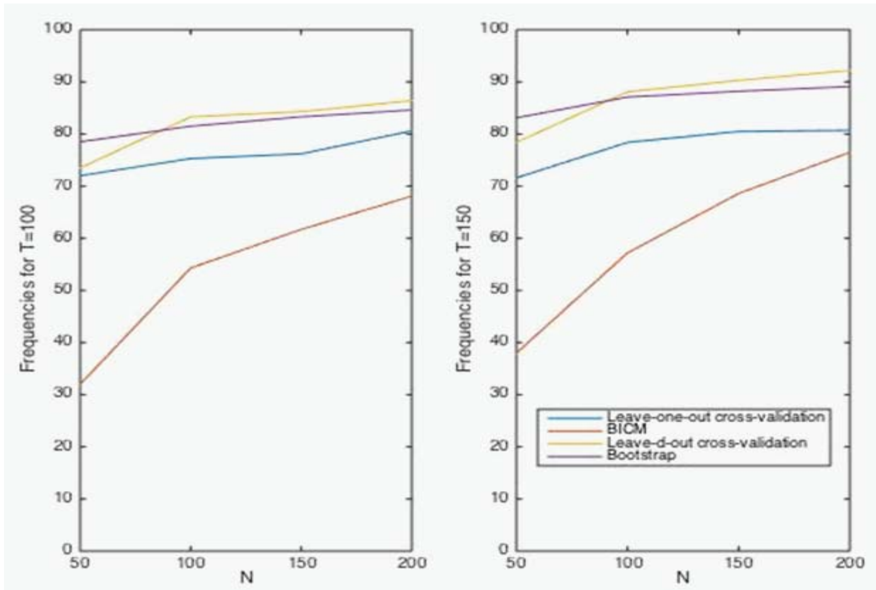
Note : see Figure 11.

Figure 14. Frequencies of selecting the two first estimated factors for DGP 1



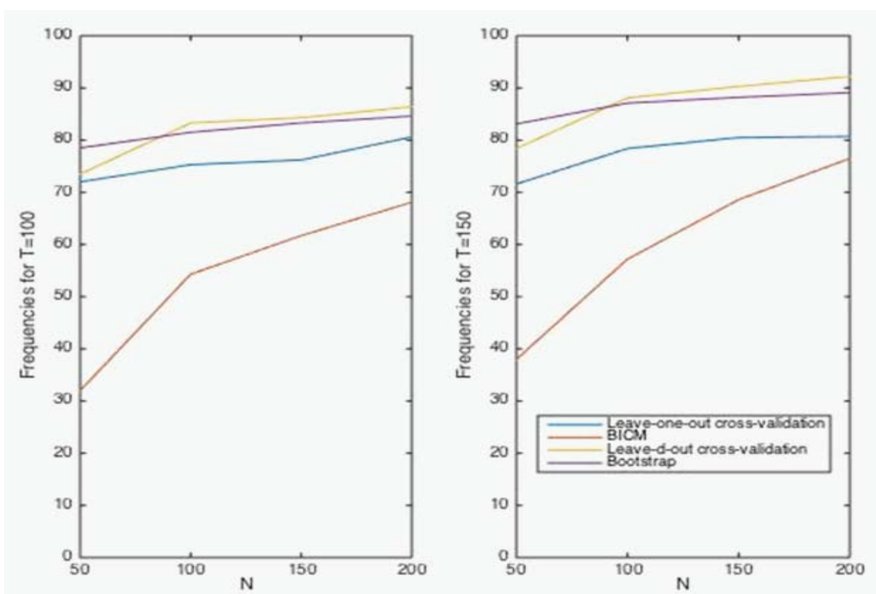
Note : The figure reports frequencies in percentage by methods.

Figure 15. Frequencies of selecting the three first estimated factors for DGP 2



Note : see Figure 14.

Figure 16. Frequencies of selecting the three first estimated factors for DGP 3



Note : see Figure 14.

3.5 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 data set are constructed using Jurado, Ludvigson and Ng (2015) financial data set and variables from Kenneth R. French website as described in the appendix [1]. The quarterly macro data are downloaded from the St. Louis Federal Reserve website and correspond to the monthly series constructed by McCracken and Ng (2015). Some of the quarterly data are also constructed based on McCracken and Ng (2015) data as explained in the Appendix. We examine how economic information summarized through a few numbers of estimated factors from real economic activities data and those related to financial markets can explain the future excess returns using various selection procedures. Recently, Gonçalves, McCracken and Perron (2015) also study the predictive ability of estimated factors from the macroeconomic data provided by McCracken and Ng (2015) to forecast excess returns to the S&P 500 Composite Index. They detect the interest rate factor as the strongest predictor of the equity premium. Indeed, as argued by Ludvigson and Ng (2007), restricting attention to a few sets of observed factors may not span all information related to financial market participants. Unlike Gonçalves, McCracken and Perron (2015), they considered both financial and macroeconomic data. Using the BIC, they found three new estimated factors termed "volatility", "risk premium" and "real" factors that have predictive power for the market excess returns after controlling for usual observed factors.

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

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

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

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

and the financial factor panel model

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

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

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

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

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

3.5.1 In-sample prediction of excess returns

The estimated regression takes the form

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

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

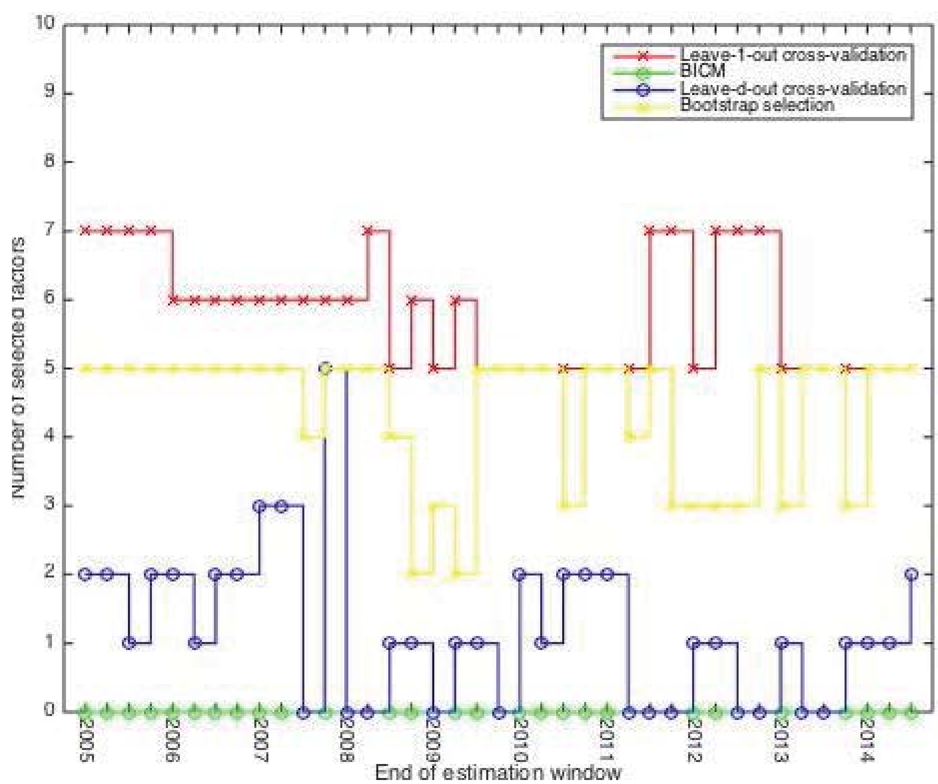
The Monte Carlo cross-validation and the bootstrap selection procedures select smaller set of generated regressors than the leave-one-out cross-validation. On the other hand, BICM selects the model with no financial or macro factor. Our cross-validation method selects two factors, the second financial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

3.5.2 Out-of-sample prediction of excess returns

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

We compare the different sets of model by computing the MSE relative to the benchmark which only contains the constant. The alternative model contains generated regressors selected using the CV_1, the BICM, CV_d or the bootstrap selection. Hence, we compute MSE_u/MSE_r the out-of-sample mean squared error of each unconstrained model relative to MOD_0. The forecast error is smaller in our new procedures (MSE_u/MSE_r = 0.8355 for $\hat{\Gamma}_\kappa$ and MSE_u/MSE_r = 0.9672 for CV_d) than for the CV_1 (MSE_u/MSE_r = 1.3163) and BICM (MSE_u/MSE_r = 1.0000). Another method of gauging the out-of-sample is to test the equal predictive ability of out-of-sample forecasts as considered by Gonçalves, McCracken and Perron (2015). Because in each recursion a new number of estimated factors is selected, there are no available critical values for such a situation.

Figure 17. Number of selected factors in the out-of-sample exercise



Note : This figure plots the number of selected estimated factors in term of the end of estimation windows for each model selection procedure.

Figure 17 indicates how the number of selected factors varies while the estimation sample changes. During the forecast exercise, while the BICM never selects an estimated factor, the leave-one-out cross-validation always chooses a larger model than our proposed methods. As it is argued in the previous sections, the new consistent model selection approaches prevent against too much under-fitting and over-fitting.

3.6 Conclusion

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

Conclusion générale

Cette thèse étudie l'utilisation des méthodes de bootstrap pour améliorer la prédiction de variables économiques en se basant sur les modèles à facteurs augmentés. Ces modèles permettent l'inclusion d'informations relatives à un grand nombre de prédicteurs à travers un petit nombre de facteurs communs. Étant donné que ces facteurs sont latents, ils sont estimés en utilisant entre autre la méthode des composantes principales. Ce travail contribue à cette litterature à au moins trois niveaux.

En premier lieu, nous justifions deux approches de bootstrap basés sur les résidus capables non seulement de reproduire le biais dans la distribution asymptotique des estimateurs des coefficients dû à l'estimation des facteurs, mais surtout de préserver la dépendence dans les erreurs lorque l'horizon de prévision est $h > 1$. La première approche, le «block wild bootstrap», subdivise les résidus en k blocs multipliés respectivement par k tirages i.i.d.$(0, 1)$. Ce faisant, nous pouvons maintenir la dépendance à l'intérieur de chaque bloc tout en considérant k blocs indépendants. La seconde, le «dependent wild bootstrap», est mis en oeuvre en multipliant chaque résidu par une variable qui est une moyenne pondérée locale de tirage externe. En deuxième lieu, nous proposons d'algorythme permettant de construire les intervalles de prediction pour une réalisation future d'une variable ainsi que pour sa moyenne conditionnellement à l'information disponible. Ce faisant, nous pouvons relâcher l'hypothèse forte de normalité des erreurs. En troixième lieu, nous nous sommes intéressés à comment choisir passimonieusement parmi les facteurs communs estimés de sorte à recouvrir l'information contenue dans les facteurs latents. L'examen des insuffisances des methodes de selection existantes ont permis de proposer un critère validation croisée «leave-d-out» et un critère bootstrap approprié. Comme le montre nos résultats en échantillon fini, ces deux approches préservent contre la selection d'un ensemble inapproprié de facteurs estimés.

Ce travail ouvre la porte à de nouvelles pistes de reflexion. En particulier, nos résultats peuvent aisément s'étendre dans le contexte où aucun sous-ensemble de facteurs estimés n'est correcte.

Bibliographie

Andrews, D. W K, 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

Andrews, D. W K and C. J. Monahan, 1992. "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953-66.

Bai, J., 2003. "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135-172.

Bai, J. and S. Ng, 2002. "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191-221.

Bai, J. and S. Ng, 2006. "Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-augmented Regressions," *Econometrica*, 74, 1133-1150.

Bai, J. and S. Ng, 2009, "Boosting Diffusion Indices," *Journal of Applied Econometrics*, 24, 607-629.

Bai, J. and S. Ng, 2013. "Principal Components Estimation and Identification of the Factors," *Journal of Econometrics*, 176, 18-29.

Beran, R. 1987. "Prepivoting to Reduce Level Error of Confidence Sets", *Biometrika*, 74, 3, 457-68.

Bickel, P. and D. Freedman, 1981. "Asymptotic Theory for the Bootstrap". *Annals of Statistics*, Vol. 9, No.6, 1196-1217.

Campbell, J. Y., 1991, "A Variance Decomposition for Stock Returns," *Economic Journal,* 101, 157-179.

Campbell, J. Y., Shiller, R. J., 1989, "The dividend-price ratio and expectations of future dividends and discount factors," *Review of Financial Studies,* 1 (3), 195–228.

Carrasco, M. and B. Rossi, 2015. "In-sample Inference and Forecasting in Misspecified Factor Models," Working paper.

Chang, Y. and J. Park, 2003. "A Sieve Bootstrap for the Test of a Unit Root", *Journal of Time Series Analysis*, 24, 379-400.

Cheng, X. and B. E. Hansen, 2013. "Forecasting with Factor-Augmented Regression : A Frequentist Model Averaging Approach," *Journal of Econometrics*, 186(2), 280-293.

Davidson, J., 1994. *Stochastic Limit Theory : An Introduction for Econometricians*, Oxford University Press.

De Jong, R. and J. Davidson, 2000. “Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices,” *Econometrica*, 68, 407-424.

Diebold, F. X., 2007. *Elements of Forecasting, Fourth Edition*, Thomson South-Western.

Djogbenou, A., S. Gonçalves, and B. Perron, 2014. “Bootstrap Inference for Regressions with Estimated Factors and Serial Correlation”, *Journal of Time Series Analysis*, forthcoming.

Fama, E. F., and K. R. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics,* 22, 3-27.

Fama, E. F., and K. R. French, 1989, "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics,* 25, 23-49.

Gallant, R. and H. White, 1987. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Blackwell.

Gonçalves, S., McCracken, M., and B. Perron, 2013. "Tests of Equal Predictive Ability in Factor-Augmented Models," Working paper.

Gonçalves, S., B. Perron, and A. Djogbenou, 2015,. "Bootstrap Prediction Intervals for Factor Models," *Journal of Business and Economics Statistic*, forthcoming.

Gonçalves, S. and B. Perron, 2014. “Bootstrapping Factor-Augmented Regressions,” *Journal of Econometrics*, 182, 156-173.

Groen, J. and Kapetanios, G,. 2013. "Model Selection Criteria for Factor-Augmented Regressions," *Oxford bulletin of economics and statistics, 75 (1), 37-63.*

Hall, P., 1992. *The Bootstrap and Edgeworth Expansion*, Springer.

Hansen, B. E., 2010, "Multi-Step Forecast Model Selection," University of Wisconsin, Working paper.

Hansen, B. E., and J. S., Racine, 2012, "Jackknife Model Averaging," *Journal of Econometrics,* 167, 38-46.

Jurado, K., Ludvigson, S., and S. Ng, 2015, "Measuring Uncertainty," *The American Economic Review*, forthcoming.

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, 101–116.

Li, K. C.,1987, "Asymptotic Optimality for Cp ; CL ; Cross-Validation and Generalized Cross-Validation : Discrete Index Set," *Annals of Statistics,* 15, 958-975.

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

Ludvigson, S. C., and S. Ng, 2007, "The Empirical Risk-Return Relation : a Factor Analysis Approach," *Journal of Financial Economics*, 83, 171-222

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

Lahiri, S.N., 2003. *Resampling Methods for Dependent Data,* Springer-Verlag.

McCracken, M. W., and S. Ng, 2015, "FRED-MD : A Monthly Database for Macroeco- nomic Research," http ://www.columbia.edu/ sn2294/papers/freddata.pdf.

Onatski, A., 2015, "Asymptotic Analysis of the Squared Estimation Error in Misspecified Factor Models," *Journal of Econometrics*, 186 (2), 388–406.

Pan, L. and D. Politis, 2014. “Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions”, *Journal of Statistical Planning and Inference*, forthcoming.

Pascual, L., J. Romo, and E. Ruiz, 2001. “Effects of Parameter Estimation on Prediction Densities : a Bootstrap Approach”, *International Journal of Forecasting*, 17 (1), 83-103.

Politis, D. 2013. “Model-free Model-fitting and Predictive Distributions" (with discussion and rejoinder), *Test*, 22, 183-250.

Shao, J., 1993. "Linear Model Selection by Cross-Validation," *Journal of the American Statistical Association,* 422, 486-494.

Shao, J., 1996. "Bootstrap Model Selection," *Journal of the American Statistical Association,* 434, 635-665.

Shao, J., 1997. "An asymptotic Theory for Linear Model Selection," *Statistica Sinica,* 7, 221-264.

Shao, X., 2010. “The Dependent Wild Bootstrap,” *Journal of the American Statistical Association,* 05, 218-235.

Shao, X., 2011. “A Bootstrap-assisted Spectral Test of White Noise under Unknown Dependence,” *Journal of Econometrics,* 162, 213-224.

Shiller, R. J., 1981, "Do stock srices move too much to be justified by subsequent changes individends ?," *American Economic Review,* 71, 421–436.

Stock, J. H. 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 Choice and Assessment of Statistical Predictions," *Journal of the Royal Statistical Society,* Ser.B, 36, 111-147.

Smeekes, S. and J.-P. Urbain, 2013. “Unit Root Testing Using Modified Wild Bootstrap Methods,” mimeo, Maastricht University.

Wolf, M. and D. Wunderli, 2015. “Bootstrap Joint Prediction Regions”, *Journal of Time Series Analysis*, 36, 352–376.

Yeh, A. B., 1998. “A Bootstrap Procedure in Linear Regression with Nonstationary Errors,” *The Canadian Journal of Statistics Association*, 26(1), 149-160.

Annexes

.1 Proofs of results in Sections 1.2 and 1.3

We first state an auxiliary result which strengthens the results in Lemma A.1 in Bai (2003, p. 159) and Theorem 1 in Bai and Ng (2002, p. 198), followed by its proof. We then prove the results in the main text.

Lemma 15. *Under Assumptions 1, 2, 3 and 4.a) strengthened by Assumption 3'.d), if* $N^{-1}T^{1/2} \to c < \infty$*, then* $\sum_{t=1}^{T} \left\| \tilde{F}_t - HF_t \right\|^4 = O_P(1)$.

Proof of Lemma 15. We have the following identity

$$\tilde{F}_t - HF_t = \tilde{V}^{-1}\left(\underbrace{T^{-1}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}}_{a_t} + \underbrace{T^{-1}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}}_{b_t} + \underbrace{T^{-1}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}}_{c_t} + \underbrace{T^{-1}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}}_{d_t} \right),$$

where $\gamma_{st} = E\left(N^{-1}\sum_{i=1}^{N} e_{is}e_{it}\right)$, $\zeta_{st} = N^{-1}\sum_{i=1}^{N}\left(e_{is}e_{it} - E\left(N^{-1}\sum_{i=1}^{N} e_{is}e_{it}\right)\right)$, $\eta_{st} = N^{-1}\sum_{i=1}^{N}\lambda_i' F_s e_{it}$, and $\xi_{st} = N^{-1}\sum_{i=1}^{N}\lambda_i' F_t e_{is}$. By the c-r inequality, it follows that

$$\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^4 \le 4^3\left\|\tilde{V}^{-1}\right\|^4\left(\sum_{t=1}^{T}\|a_t\|^4 + \sum_{t=1}^{T}\|b_t\|^4 + \sum_{t=1}^{T}\|c_t\|^4 + \sum_{t=1}^{T}\|d_t\|^4\right).$$

Note that $T^{-1}\sum_{t=1}^{T}\|a_t\|^4 \le T\left(T^{-1}\sum_{t=1}^{T}\|a_t\|^2\right)^2$ and $T^{-1}\sum_{t=1}^{T}\|a_t\|^2 = O_P(T^{-1})$ (Bai and Ng (2002, p. 213)), implying that $\sum_{t=1}^{T}\|a_t\|^4 = O_P(1)$. Similarly, by repeated application of Cauchy-Schwarz inequality,

$$\sum_{t=1}^{T}\|b_t\|^4 \le \left[T^{-2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left|\tilde{F}_s'\tilde{F}_u\right|\left(\sum_{t=1}^{T}\zeta_{st}^2\zeta_{ut}^2\right)^{1/2}\right]^2 \le \left(T^{-1}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2\right)^2 T^{-2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}^2\zeta_{ut}^2\right),$$

where we can show that $T^{-1}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2 = O_P(1)$ and $T^{-2}\sum_{s=1}^{T}\sum_{u=1}^{T}E\left(\sum_{t=1}^{T}\zeta_{st}^2\zeta_{ut}^2\right) = O(N^{-2}T)$. In particular,

$$T^{-2}\sum_{s=1}^{T}\sum_{u=1}^{T}E\left(\sum_{t=1}^{T}\zeta_{st}^2\zeta_{ut}^2\right) \le T\left[\max_{s,t}E\left(\zeta_{st}^4\right)\right] = O\left(N^{-2}T\right),$$

since $\max_{s,t} E\left(\zeta_{st}^4\right) = O\left(N^{-2}\right)$ by Assumption 2.c). Thus, $\sum_{t=1}^{T} \|b_t\|^4 = O_P\left(N^{-2}T\right)$. Thirdly,

$$\sum_{t=1}^{T} \|c_t\|^4 = \sum_{t=1}^{T} \left\| (NT)^{-1} \sum_{s=1}^{T} \tilde{F}_s F_s' \Lambda e_t \right\|^4 \leq \sum_{t=1}^{T} \left\| N^{-1} \Lambda e_t \right\|^4 \left\| T^{-1} \sum_{s=1}^{T} \tilde{F}_s F_s' \right\|^4,$$

implying that

$$\sum_{t=1}^{T} \|c_t\|^4 \leq \frac{T}{N^2} \frac{1}{T} \sum_{t=1}^{T} \left\| N^{-1/2} \sum_{i=1}^{N} \lambda_i e_t \right\|^4 \left(T^{-1} \sum_{s=1}^{T} \left\| \tilde{F}_s \right\|^2 \right)^2 \left(T^{-1} \sum_{s=1}^{T} \|F_s\|^2 \right)^2 = O_P\left(N^{-2}T\right),$$

given in particular Assumption 3'.d). The proof that $\sum_{t=1}^{T} \|d_t\|^4 = O_P\left(N^{-2}T\right)$ is similar and therefore omitted. Thus, $\sum_{t=1}^{T} \left\| \tilde{F}_t - HF_t \right\|^4 = O_P(1)$ as $N^{-1}T^{1/2} \to c < \infty$. □

Proof of Theorem 1. We apply Theorem 2.1 of GP (2014) by verifying their high level assumptions. Our Assumptions 1-4 coincide with their Assumptions 1-4, whereas by Theorem 5.3 of Gallant and White (1988, p. 76), we have that $\Omega^{-1/2} T^{-1/2} z' \varepsilon \to^d N(0, I)$, which verifies their Assumption 5.b). Finally, our moment conditions on z_t and ε_{t+h} imply those of GP (2014). □

Proof of Lemma 2. We can write $\hat{\Omega} = A_{1T} + A_{2T} + A_{3T} + A_{2T}' + A_{4T} + A_{5T} + A_{3T}' + A_{5T}' + A_{6T}$, with

$$A_{1T} = T^{-1} \sum_{s=1}^{T-h} \sum_{t=1}^{T-h} \hat{z}_t \varepsilon_{t+h} \hat{z}_s' \varepsilon_{s+h} k\left(\frac{s-t}{M_T}\right), \; A_{2T} = T^{-1} \sum_{s=1}^{T-h} \sum_{t=1}^{T-h} \hat{z}_t \varepsilon_{t+h} \left(\delta - \hat{\delta}\right)' \hat{z}_s \hat{z}_s' k\left(\frac{s-t}{M_T}\right),$$

$$A_{3T} = T^{-1} \sum_{s=1}^{T-h} \sum_{t=1}^{T-h} \hat{z}_t \varepsilon_{t+h} \alpha' H^{-1} \left(HF_s - \tilde{F}_s\right) \hat{z}_s' k\left(\frac{s-t}{M_T}\right),$$

$$A_{4T} = T^{-1} \sum_{s=1}^{T-h} \sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t' \left(\delta - \hat{\delta}\right) \left(\delta - \hat{\delta}\right)' \hat{z}_s \hat{z}_s' k\left(\frac{s-t}{M_T}\right),$$

$$A_{4T} = T^{-1} \sum_{s=1}^{T-h} \sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t' \left(\delta - \hat{\delta}\right) \left(\delta - \hat{\delta}\right)' \hat{z}_s \hat{z}_s' k\left(\frac{s-t}{M_T}\right),$$

$$A_{5T} = T^{-1} \sum_{s=1}^{T-h} \sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t' \left(\delta - \hat{\delta}\right) \alpha' H^{-1} \left(HF_s - \tilde{F}_s\right) \hat{z}_s' k\left(\frac{s-t}{M_T}\right), \text{ and}$$

$$A_{6T} = T^{-1} \sum_{s=1}^{T-h} \sum_{t=1}^{T-h} \hat{z}_t \left(HF_t - \tilde{F}_t\right)' \left(H^{-1}\right)' \alpha \alpha' H^{-1} \left(HF_s - \tilde{F}_s\right) \hat{z}_s' k\left(\frac{s-t}{M_T}\right).$$

Next we show $A_{1T} \to^P \Phi_0 \Omega \Phi_0'$ and $A_{iT} = o_P(1)$, for $i = 2, \ldots, 6$. Starting with A_{2T}, note that

$$\begin{aligned} A_{2T} &= T^{-1} \sum_{t=1}^{T-h} \hat{z}_t \varepsilon_{t+h} \left(\delta - \hat{\delta}\right)' \hat{z}_t \hat{z}_t' + T^{-1} \sum_{\tau=1}^{T-h-1} \sum_{t=1}^{T-h-\tau} \hat{z}_t \varepsilon_{t+h} \left(\delta - \hat{\delta}\right)' \hat{z}_{t+\tau} \hat{z}_{t+\tau}' k\left(\frac{\tau}{M_T}\right) \\ &\quad + T^{-1} \sum_{\tau=1}^{T-h-1} \sum_{t=1}^{T-h-\tau} \hat{z}_{t+\tau} \varepsilon_{t+h+\tau} \left(\delta - \hat{\delta}\right)' \hat{z}_t \hat{z}_t' k\left(\frac{\tau}{M_T}\right) \equiv A_{2T,1} + A_{2T,2} + A_{2T,3}. \end{aligned}$$

By repeated application of Cauchy-Schwarz inequality,

$$\|A_{2T,1}\| \leq \left\|\delta-\hat{\delta}\right\| \left(T^{-1}\sum_{t=1}^{T-h}\|\hat{z}_t\|^4\right)^{3/4} \left(T^{-1}\sum_{t=1}^{T-h}\|\varepsilon_{t+h}\|^4\right)^{1/4}.$$

Similarly, we can show that

$$\|A_{2T,2}\| \leq \left\|\delta-\hat{\delta}\right\| \sum_{\tau=1}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right| \left(T^{-1}\sum_{t=1}^{T-h}\|\hat{z}_t\|^4\right)^{3/4} \left(T^{-1}\sum_{t=1}^{T-h}\|\varepsilon_{t+h}\|^4\right)^{1/4},$$

where the same bound holds for $\|A_{2T,3}\|$. It follows that

$$\|A_{2T}\| \leq 2\left\|\delta-\hat{\delta}\right\| \sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right| \left(T^{-1}\sum_{t=1}^{T-h}\|\hat{z}_t\|^4\right)^{3/4} \left(T^{-1}\sum_{t=1}^{T-h}\|\varepsilon_{t+h}\|^4\right)^{1/4}.$$

By Theorem 1, $\left\|\delta-\hat{\delta}\right\| = O_P\left(T^{-1/2}\right)$. Since $M_T^{-1}\sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right| \to \int_0^{+\infty}|k(x)|\,dx < \infty$, we have that $\sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right| = O(M_T)$. We can also show that the two last factors are $O_P(1)$ (in particular, by Lemma 15 and the decomposition $\hat{z}_t = \Phi z_t + (\hat{z}_t - \Phi z_t)$, we have that $T^{-1}\sum_{t=1}^{T-h}\|\hat{z}_t\|^4 \leq 8\Phi^4 T^{-1}\sum_{t=1}^{T-h}\|z_t\|^4 + 8T^{-1}\sum_{t=1}^{T-h}\|\hat{z}_t-\Phi z_t\|^4 = O_P(1)$ given Markov's inequality and the moment conditions on z_t). Thus, $A_{2T} = O_P\left(T^{-1/2}M_T\right) = o_P(1)$ since $T^{-1/2}M_T \to 0$. Turning now to A_{3T}, and given that $\hat{z}_t = \Phi z_t + (\hat{z}_t - \Phi z_t)$, we can write $A_{3T} = A_{3T,1} + A_{3T,2}$, where

$$\begin{aligned} A_{3T,1} &= T^{-1}\sum_{s=1}^{T-h}\sum_{t=1}^{T-h}(\hat{z}_t-\Phi z_t)\,\varepsilon_{t+h}\alpha' H^{-1}\left(HF_s-\tilde{F}_s\right)\hat{z}_s' k\left(\frac{s-t}{M_T}\right) \text{ and} \\ A_{3T,2} &= T^{-1}\Phi\sum_{s=1}^{T-h}\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}\alpha' H^{-1}\left(HF_s-\tilde{F}_s\right)\hat{z}_s' k\left(\frac{s-t}{M_T}\right). \end{aligned}$$

Using the same arguments as for A_{2T}, we can show that $\|A_{3T,1}\| = O_P\left(T^{-1/2}M_T\right)$. Similarly,

$$\|A_{3T,2}\| \leq \|\Phi\|\left\|\alpha' H^{-1}\right\| \left(T^{-1}\sum_{s=1}^{T-h}\left\|HF_s-\tilde{F}_s\right\|^4 T^{-1}\sum_{s=1}^{T-h}\|\hat{z}_s\|^4\right)^{1/4} \left(T^{-1}\sum_{s=1}^{T-h}\left\|\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}k\left(\frac{s-t}{M_T}\right)\right\|^2\right)^{1/2},$$

which is $O_P\left(\left(T^{-1/2}M_T\right)^{1/2}\right)$. In particular, the last factor is $O_P(M_T)$. To see this, note that

$$\begin{aligned} E\left(T^{-1}\sum_{s=1}^{T-h}\left\|\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}k\left(\frac{s-t}{M_T}\right)\right\|^2\right) &= T^{-1}\sum_{s=1}^{T-h}\sum_{t_1=1}^{T-h}\sum_{t_2=1}^{T-h} E\left(z_{t_1}'\varepsilon_{t_1+h}z_{t_2}\varepsilon_{t_2+h}\right) k\left(\frac{s-t_1}{M_T}\right) k\left(\frac{s-t_2}{M_T}\right) \\ &\leq T^{-1}\sum_{t_1=1}^{T-h}\sum_{t_2=1}^{T-h}\left|E\left(z_{t_1}'\varepsilon_{t_1+h}z_{t_2}\varepsilon_{t_2+h}\right)\right| \sum_{\tau=-\infty}^{+\infty}\left|k\left(\frac{\tau}{M_T}\right)\right|^2 = O_P(M_T), \end{aligned}$$

given that the first factor is $O_P(1)$ (by a mixingale inequality, cf. Corollary 14.3 of Davidson (1994, p. 212) with $p = r > 2$ and $r > \frac{p}{p-1}$ and the fact that $E\left\|z_t\varepsilon_{t+h}\right\|^r \leq C < \infty$) and that the second factor is $O(M_T)$ since $M_T^{-1}\sum_{\tau=-\infty}^{+\infty}\left|k\left(\frac{\tau}{M_T}\right)\right|^2 \to \int_{-\infty}^{+\infty} k(x)^2\,dx \leq \int_{-\infty}^{+\infty}|k(x)|\,dx < \infty$. Thus, $A_{3T} = O_P\left(T^{-1/2}M_T\right) = o_P(1)$. Similar arguments show that $\|A_{4T}\| \leq O_P\left(T^{-1}M_T\right) = o_P(1)$, $A_{5T} = O_P\left(T^{-3/4}M_T\right) = o_P(1)$ and $A_{6T} = O_P\left(T^{-1/2}M_T\right) = o_P(1)$. Finally, we show that $A_{1T} \to^P \Phi_0\Omega\Phi_0'$. By replacing $\hat{z}_t$ with $\Phi z_t + (\hat{z}_t - \Phi z_t)$, we can write $A_{1T} = A_{1T,1} + A_{1T,2} + A_{1T,3} + A_{1T,3}'$, where

$$\begin{aligned}
A_{1T,1} &= \Phi T^{-1}\sum_{s=1}^{T-h}\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}z_s'\varepsilon_{s+h}k\left(\frac{s-t}{M_T}\right)\Phi', \\
A_{1T,2} &= T^{-1}\sum_{s=1}^{T-h}\sum_{t=1}^{T-h}\left(\hat{z}_t - \Phi z_t\right)\varepsilon_{t+h}\left(\hat{z}_s - \Phi z_s\right)'\varepsilon_{s+h}k\left(\frac{s-t}{M_T}\right), \\
A_{1T,3} &= \Phi T^{-1}\sum_{s=1}^{T-h}\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}\left(\hat{z}_s - \Phi z_s\right)'\varepsilon_{s+h}k\left(\frac{s-t}{M_T}\right).
\end{aligned}$$

By arguments similar to those already used, we can show that the last two terms are $O_P\left(T^{-1/2}M_T\right) = o_p(1)$. To show that $A_{1T,1} \to^P \Phi_0\Omega\Phi_0'$, note that under our assumptions, $E\left\|z_t\varepsilon_{t+h}\right\|^{2r} < C$ and $\left\{z_t\varepsilon_{t+h}\right\}$ is a strong mixing sequence of size $-\frac{3r}{r-2}$. The result then follows by Proposition 1 of Andrews (1991, p. 825) and the fact that $\Phi = \Phi_0 + o_P(1)$. Since $T^{-1}\hat{z}'\hat{z} = \Phi_0\Sigma_{zz}\Phi_0' + o_P(1)$ with $\Sigma_{zz} > 0$, we conclude that $\widehat{\Sigma}_\delta \to^P \Sigma_\delta = \left(\Phi_0'\right)^{-1}\Sigma_{zz}^{-1}\Omega\Sigma_{zz}^{-1}\Phi_0^{-1}$. □

Proof of Theorem 3. We verify Conditions A*-F* of GP (2014). Because our bootstrap scheme relies on the wild bootstrap to generate e_{it}^*, as in GP (2014), conditions that only involve this random variable were already verified by them. In particular, Conditions A*, B* and F* are satisfied under our assumptions (see proof of Theorem 4.1 of GP (2014)). Hence, we only need to verify Conditions C*, D* and E*. Starting with Condition C*(a), by the independence between e_{it}^* and ε_{s+h}^*, and the fact that e_{it}^* is independent across (i,t), it follows that

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T} E^*\left|(TN)^{-1/2}\sum_{s=1}^{T-h}\sum_{i=1}^{N}\varepsilon_{s+h}^*\left(e_{it}^*e_{is}^* - E\left(e_{it}^*e_{is}^*\right)\right)\right|^2 \\
=\;& T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T-h}\sum_{l=1}^{T-h} E^*\left(\varepsilon_{s+h}^*\varepsilon_{l+h}^*\right)N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} Cov^*\left(e_{it}^*e_{is}^*, e_{jt}^*e_{jl}^*\right) \\
=\;& T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T-h}\hat{\varepsilon}_{s+h}^2 N^{-1}\sum_{i=1}^{N}\widetilde{e}_{it}^2\widetilde{e}_{is}^2 Var^*\left(\eta_{it}\eta_{is}\right) \\
\leq\;& CN^{-1}\sum_{i=1}^{N}\left(T^{-1}\sum_{t=1}^{T}\widetilde{e}_{it}^2\right)\left(T^{-1}\sum_{s=1}^{T-h}\hat{\varepsilon}_{s+h}^2\widetilde{e}_{is}^2\right),
\end{aligned}$$

where the second equality uses the fact that $Cov^*\left(e_{it}^*e_{is}^*, e_{jt}^*e_{jl}^*\right) = 0$ for $i \neq j$ or $s \neq l$, and $E^*\left(\varepsilon_{s+h}^{*2}\right) = \hat{\varepsilon}_{s+h}^2$, given that $E^*\left(\eta_j^2\right) = 1$. The inequality relies on a bound for $Var^*\left(\eta_{it}\eta_{is}\right)$ under our assumptions. The result follows by an application of Cauchy-Schwarz inequality given in particular the fact that $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{e}_{it}^{4}=O_P(1)$ and $T^{-1}\sum_{s=1}^{T-h}\hat{\varepsilon}_{s+h}^{4}=O_P(1)$ under our assumptions. For Condition C*(b), we can show that

$$E^*\left\|(NT)^{-1/2}\sum_{t=1}^{T-h}\sum_{i=1}^{N}\tilde{\lambda}_i e_{it}^*\varepsilon_{t+h}^*\right\|^2=(NT)^{-1}\sum_{t=1}^{T-h}\hat{\varepsilon}_{t+h}^2\left(\sum_{i=1}^{N}\left\|\tilde{\lambda}_i\right\|^2\tilde{e}_{it}^2\right),$$

using the facts that $E^*\left(e_{it}^*e_{js}^*\right)=0$ whenever $i\neq j$ or $t\neq s$ and $E^*\left(\varepsilon_{t+h}^{*2}\right)=\hat{\varepsilon}_{t+h}^2$. The rest of the proof follows exactly the proof of GP (2014) (cf. Proof of their Theorem 4.1). The proof of Condition C*(c) follows the proof in GP (2014) closely with the only difference that we show that $T^{-1}\sum_{t=1}^{T-h}\varepsilon_{t+h}^{*4}=O_{p^*}(1)$ in probability. Indeed,

$$E^*\left(T^{-1}\sum_{t=1}^{T-h}\varepsilon_{t+h}^{*4}\right)=T^{-1}\sum_{j=1}^{k_T}\sum_{l=1}^{b_T}\hat{\varepsilon}_{(j-1)b_T+l+h}^4E^*\left(\nu_j^4\right)\leq CT^{-1}\sum_{j=1}^{k_T}\sum_{l=1}^{b_T}\hat{\varepsilon}_{(j-1)b_T+l+h}^4=CT^{-1}\sum_{t=1}^{T-h}\hat{\varepsilon}_{t+h}^4,$$

since $E^*\left(\nu_j^4\right)\leq C<\infty$. Because $T^{-1}\sum_{t=1}^{T-h}\hat{\varepsilon}_{t+h}^4=O_P(1)$ under our assumptions, this proves the desired result. For Condition D*(a), we have that for any $i=1,\ldots,b_T$ and $j=1,\ldots,k_T$, $E^*\left(\varepsilon_{i+(j-1)b_T+h}^*\right)=\hat{\varepsilon}_{(j-1)b_T+i+h}E^*\left(\nu_j\right)=0$ and $T^{-1}\sum_{t=1}^{T-h}E^*\left|\varepsilon_{t+h}^*\right|^2=T^{-1}\sum_{t=1}^{T-h}\hat{\varepsilon}_{t+h}^2=O_p(1)$. For Condition D*(b), let

$$\xi_j^*\equiv\Omega^{*-1/2}b_T^{-1/2}\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}^*=\Omega^{*-1/2}b_T^{-1/2}\sum_{l=1}^{b_T}\hat{z}_{(j-1)b+l}\hat{\varepsilon}_{(j-1)b_T+l+h}\nu_j,$$

where ν_j are i.i.d. $(0,1)$ across j. We can write $\Omega^{*-1/2}T^{-1/2}\sum_{t=1}^{T-h}\hat{z}_t\varepsilon_{t+h}^*=k_T^{-1/2}\sum_{j=1}^{k_T}\xi_j^*$, where ξ_j^* are conditionally independent for $j=1,\ldots,k_T$, with $E^*\left(\xi_j^*\right)=0$ and $Var^*\left(k_T^{-1/2}\sum_{j=1}^{k_T}\xi_j^*\right)=I$. It suffices to show that for some $d>1$, $Z_T\equiv k_T^{-d}\sum_{j=1}^{k_T}E^*\left\|\xi_j^*\right\|^{2d}=o_P(1)$. Replacing ξ_j^* by its definition and using the fact that $k_T=b_T^{-1}(T-h)$, we have that

$$Z_T\leq C\left\|\Omega^{*-1/2}\right\|^{2d}\frac{1}{T^d}\sum_{j=1}^{k_T}E^*\left(\left\|\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\varepsilon_{(j-1)b+l+h}^*\right\|\right)^{2d}.$$

Since $\Omega^*\to^P\Phi_0\Omega\Phi_0'$ (see Condition E* below) and $\Phi_0\Omega\Phi_0'>0$, $\left\|\Omega^{*-\frac{1}{2}}\right\|^{2d}=O_P(1)$. We show that the second factor is $o_P(1)$. Noting that

$$\varepsilon_{(j-1)b_T+l+h}^*=\varepsilon_{(j-1)b_T+l+h}\cdot\nu_j-\hat{z}_{(j-1)b_T+l}'\left(\hat{\delta}-\delta\right)\cdot\nu_j+\left(HF_{(j-1)b_T+l}-\tilde{F}_{(j-1)b_T+l}\right)'\left(H^{-1}\right)'\alpha\cdot\nu_j, \quad (5)$$

we have that

$$\frac{1}{T^d}\sum_{j=1}^{k_T} E^*\left\|\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\varepsilon^*_{(j-1)b_T+l+h}\right\|^{2d} \leq 3^{2d-1}\frac{1}{T^d}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}\right\|^{2d} E^*\left|\nu_j\right|^{2d}$$
$$+3^{2d-1}\frac{1}{T^d}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\hat{z}'_{(j-1)b_T+l}\left(\hat{\delta}-\delta\right)\right\|^{2d} E^*\left|\nu_j\right|^{2d}$$
$$+3^{2d-1}\frac{1}{T^d}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\left(HF_{(j-1)b_T+l}-\tilde{F}_{(j-1)b_T+l}\right)'\left(H^{-1}\right)'\alpha\right\|^{2d} E^*\left|\nu_j\right|^{2d} \equiv Z_{1T}+Z_{2T}+Z_{3T}.$$

Starting with Z_{1T}, by letting $\hat{z}_{(j-1)b+l} = \Phi z_{(j-1)b_T+l} + \left(\hat{z}_{(j-1)b_T+l} - \Phi z_{(j-1)b_T+l}\right)$ and using the c-r inequality,

$$\begin{aligned} Z_{1T} \quad &\leq \quad C\frac{1}{T^d}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}\right\|^{2d} \leq C\left\|\Phi\right\|^{2d}\frac{2^{2d-1}}{T^d}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}z_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}\right\|^{2d} \\ &\quad +C\frac{2^{2d-1}}{T^d}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\left(\hat{z}_{(j-1)b_T+l}-\Phi z_{(j-1)b_T+l}\right)\varepsilon_{(j-1)b_T+l+h}\right\|^{2d}. \end{aligned}$$

Noting that for any $d>1$, $\sum_{j=1}^{k_T}|a_j|^{2d} \leq \left(\sum_{j=1}^{k_T}|a_j|^2\right)^d$ for any a_j, we can bound the second term of Z_{1T} by

$$C\frac{1}{T^d}\left(\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\left(\hat{z}_{(j-1)b+l}-\Phi z_{(j-1)b_T+l}\right)\varepsilon_{(j-1)b_T+l+h}\right\|^2\right)^d = O_P\left(\left(T^{-1/2}b_T\right)^d\right) = o_P(1),$$

given an application of the c-r inequality and the fact that $\sum_{t=1}^{T-h}\left\|\hat{z}_t-\Phi z_t\right\|^4 = O_P(1)$ and $\sum_{t=1}^{T-h}\varepsilon_t^4 = O_P(T)$. Similarly, we can show that the first term of Z_{1t} is $o_P(1)$. This follows by showing that its expectation is of order $O\left(\left(T^{-1}b_T\right)\right)^{d-1} = o(1)$ for some $1<d<2$, given standard inequalities (in particular, we rely on Corollary 14.3 of Davidson (1994)). For Z_{2T}, by repeated application of the Cauchy-Schwarz and the c-r inequalities, we have that it is bounded by

$$\begin{aligned} &\frac{1}{T^d}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\hat{z}'_{(j-1)b_T+l}\left(\hat{\delta}-\delta\right)\right\|^{2d} \leq \frac{b_T^d}{T^d}\left\|\hat{\delta}-\delta\right\|^{2d}\sum_{j=1}^{k_T}\left(\sum_{l=1}^{b_T}\left\|\hat{z}_{(j-1)b_T+l}\right\|^4\right)^d \\ = \quad &O\left(b_T^d/T^d\right)\cdot O_P\left(1/T^d\right)\cdot O_P\left(k_T\cdot b_T^d\right) = O_P\left(b_T^{2d-1}/T^{2d-1}\right) = o_P(1). \end{aligned}$$

That $Z_{3T} = O_P\left(\left(T^{-1/2}b_T\right)^d\right)$ follows by similar arguments and we omit the details. For Condition E*, since ν_j is independent over $j=1,\ldots,k_T$, with $Var^*(\nu_j)=1$, we have that

$$\begin{aligned} \Omega^* \quad &= \quad T^{-1}Var^*\left(\sum_{j=1}^{k_T}\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\hat{\varepsilon}_{(j-1)b_T+l+h}\cdot\nu_j\right) \\ &= \quad T^{-1}\sum_{j=1}^{k_T}\left(\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\hat{\varepsilon}_{(j-1)b_T+l+h}\right)\left(\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\hat{\varepsilon}_{(j-1)b_T+l+h}\right)' \equiv \Omega^*_{1T}+\Omega^*_{2T}+\Omega^{*\prime}_{2T}+\Omega^*_{3T}, \end{aligned}$$

where

$$\Omega_{1T}^{*} = T^{-1}\sum_{j=1}^{k_T}\left(\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}\right)\left(\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}\right)'$$

$$\Omega_{2T}^{*} = T^{-1}\sum_{j=1}^{k_T}\left(\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}\right)$$

$$\times\left(\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\left[-\hat{z}_{(j-1)b_T+l}'\left(\hat{\delta}-\delta\right)+\alpha' H^{-1}\left(HF_{(j-1)b_T+l}-\tilde{F}_{(j-1)b_T+l}\right)\right]\right)'$$

$$\Omega_{3T}^{*} = \frac{1}{T}\sum_{j=1}^{k_T}\left(\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\left[-\hat{z}_{(j-1)b_T+l}'\left(\hat{\delta}-\delta\right)+\alpha' H^{-1}\left(HF_{(j-1)b_T+l}-\tilde{F}_{(j-1)b_T+l}\right)\right]\right)$$

$$\times\left(\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\left[-\hat{z}_{(j-1)b_T+l}'\left(\hat{\delta}-\delta\right)+\alpha' H^{-1}\left(HF_{(j-1)b_T+l}-\tilde{F}_{(j-1)b_T+l}\right)\right]\right)'.$$

Starting with Ω_{1T}^{*}, and using the fact that $\hat{z}_{(j-1)b_T+l} = \Phi z_{(j-1)b_T+l} + \left(\hat{z}_{(j-1)b_T+l} - \Phi z_{(j-1)b_T+l}\right)$, we can write $\Omega_{1T}^{*} = \Omega_{1.1T}^{*} + \Omega_{1.2T}^{*} + \Omega_{1.2T}^{*\prime} + \Omega_{1.3T}^{*}$, where

$$\Omega_{1.1T}^{*} = T^{-1}\Phi\sum_{j=1}^{k_T}\left(\sum_{l=1}^{b_T}z_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}\right)\left(\sum_{l=1}^{b_T}z_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}\right)'\Phi',$$

$$\Omega_{1.2T}^{*} = T^{-1}\Phi\sum_{j=1}^{k_T}\left(\sum_{l=1}^{b_T}z_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}\right)\left(\sum_{l=1}^{b_T}\left(\hat{z}_{(j-1)b_T+l}-\Phi z_{(j-1)b_T+l}\right)\varepsilon_{(j-1)b_T+l+h}\right)'$$

and

$$\Omega_{1.3T}^{*} = T^{-1}\sum_{j=1}^{k_T}\left(\sum_{l=1}^{b_T}\left(\hat{z}_{(j-1)b_T+l}-\Phi z_{(j-1)b_T+l}\right)\varepsilon_{(j-1)b_T+l+h}\right)\left(\sum_{l=1}^{b_T}\left(\hat{z}_{(j-1)b_T+l}-\Phi z_{(j-1)b_T+l}\right)\varepsilon_{(j-1)b_T+l+h}\right)'.$$

We can show that $\Omega_{1.1T}^{*} \to^{P} \Phi_0\Omega\Phi_0'$ by an application of Theorem 3.1 of Lahiri (2003, p. 49) and the fact that $\Phi = \Phi_0 + o_P(1)$. For $\Omega_{1.2T}^{*}$, by Cauchy-Schwarz inequality, $\Omega_{1.2T}^{*}$ is bounded by

$$\|\Phi\|\left(\frac{1}{T}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}z_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\left(\hat{z}_{(j-1)b_T+l}-\Phi z_{(j-1)b_T+l}\right)\varepsilon_{(j-1)b_T+l+h}\right\|^2\right)^{\frac{1}{2}},$$

which is $O_P\left(b_T^2/T\right)^{\frac{1}{4}} = o_P(1)$ as the first factor is $O_P(1)$ by a mixingale inequality. The second factor is bounded by

$$T^{-1}b_T\sum_{j=1}^{k_T}\sum_{l=1}^{b_T}\left\|\left(\hat{z}_{(j-1)b_T+l}-\Phi z_{(j-1)b_T+l}\right)\varepsilon_{(j-1)b_T+l+h}\right\|^2 \le T^{-1}b_T\sum_{t=1}^{T-h}\left\|\left(\widehat{z}_t-\Phi z_t\right)\varepsilon_{t+h}\right\|^2$$

$$\le b_T\left(T^{-1}\sum_{t=1}^{T-h}\left\|\widehat{z}_t-\Phi z_t\right\|^4\right)^{\frac{1}{2}}\left(T^{-1}\sum_{t=1}^{T-h}\varepsilon_{t+h}^4\right)^{\frac{1}{2}} = O_P\left(T^{-1/2}b_T\right),$$

since $\sum_{t=1}^{T-h}\left\|\hat{z}_t-\Phi z_t\right\|^4=O_P(1)$. The same argument can be used to show that $\left\|\Omega_{1.3T}^*\right\|=O_P\left(T^{-1/2}b_T\right)=o_P(1)$, implying that $\Omega_{1T}^*=o_P(1)$. Next, we show that $\Omega_{2T}^*=o_P(1)$. Letting $\hat{X}_j\equiv\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\varepsilon_{(j-1)b_T+l+h}$, by Cauchy-Schwarz inequality, we can bound $\left\|\Omega_{2T}^*\right\|$ by

$$\left(\frac{1}{T}\sum_{j=1}^{k_T}\left\|\hat{X}_j\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\left(-\hat{z}_{(j-1)b_T+l}'\left(\hat{\delta}-\delta\right)+\alpha'H^{-1}\left(HF_{(j-1)b_T+l}-\tilde{F}_{(j-1)b_T+l}\right)\right)\right\|^2\right)^{\frac{1}{2}},$$

where the first factor is equal $\left(trace\left(\Omega_{1T}^*\right)\right)^{1/2}=o_P(1)$, as we showed before. For the second factor, and ignoring the square root, applying twice the c-r inequality yields the bound of

$$\frac{2}{T}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\hat{z}_{(j-1)b_T+l}'\left(\widehat{\delta}-\delta\right)\right\|^2+\frac{2}{T}\sum_{j=1}^{k_T}\left\|\sum_{l=1}^{b_T}\hat{z}_{(j-1)b_T+l}\alpha'H^{-1}\left(HF_{(j-1)b_T+l}-\tilde{F}_{(j-1)b_T+l}\right)\right\|^2$$

by $\Omega_{2.1T}^*+\Omega_{2.2T}^*$ where

$$\Omega_{2.1T}^*=\frac{2}{T}b_T\sum_{j=1}^{k_T}\sum_{l=1}^{b_T}\left\|\hat{z}_{(j-1)b_T+l}\hat{z}_{(j-1)b_T+l}'\left(\hat{\delta}-\delta\right)\right\|^2\leq\frac{2}{T}b_T\left\|\hat{\delta}-\delta\right\|^2\sum_{t=1}^{T-h}\left\|\hat{z}_t\right\|^4=O_P\left(T^{-1}b_T\right),$$

and

$$\Omega_{2.2T}^*\quad=\quad\frac{2b_T}{T}\sum_{j=1}^{k_T}\sum_{l=1}^{b_T}\left\|\hat{z}_{(j-1)b_T+l}\alpha'H^{-1}\left(HF_{(j-1)b_T+l}-\tilde{F}_{(j-1)b_T+l}\right)\right\|^2,$$

bounded by

$$\frac{2b_T}{T}\left\|\alpha'H^{-1}\right\|^2\left(\sum_{t=1}^{T-h}\left\|\hat{z}_t\right\|^4\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T-h}\left\|HF_t-\tilde{F}_t\right\|^4\right)^{\frac{1}{2}}=O_P\left(\frac{b_T}{T^1/2}\right)$$

since in particular $\sum_{t=1}^{T-h}\left\|HF_t-\tilde{F}_t\right\|^4=O_P(1)$. Hence, $\Omega_{2T}^*=o_P(1)$. Finally, note that $\left\|\Omega_{3T}^*\right\|\leq\Omega_{2.1T}^*+\Omega_{2.2T}^*=o_P(1)$, which completes the proof. □

Proof of Theorem 4. The proof follows closely that of Theorem 3, so we only highlight the main differences. As in that proof, only Conditions C*, D* and E* of GP (2014) need to be verified, now with $\varepsilon_{s+h}^*=\hat{\varepsilon}_{s+h}\cdot w_{s+h}^*$, where w^* is l_T-dependent with mean zero and covariance matrix K, a $(T-h)\times(T-h)$ matrix with typical element given by $K_{ij}=k_{dwb}\left(\frac{j-i}{l_T}\right)$, where $k_{dwb}(\cdot)$ is a kernel function and l_T is a bandwidth parameter. Conditions C*(a) and (b) follow immediately by noting that $E^*\left(\varepsilon_{s+h}^{*2}\right)=\hat{\varepsilon}_{s+h}^2$ since $Var^*(w^*)=K$ with diagonal elements equal to one. Condition C*(c) follows by noting that $E^*\left(\varepsilon_{t+h}^{*4}\right)=\hat{\varepsilon}_{t+h}^4E^*\left(w_{t+h}^{*4}\right)$, where $E^*\left(w_{t+h}^{*4}\right)\leq C<\infty$ and $T^{-1}\sum_{t=1}^{T-h}\hat{\varepsilon}_{t+h}^4=O_P(1)$. Condition D*(a) follows exactly as in the proof of Theorem 3, with w_t^* replacing v_j. To prove D*(b), by a decomposition of ε_{t+h}^* similar to that in (5) with v_j replaced with w_{t+h}^* and the fact that $\hat{z}_t=\Phi z_t+(\hat{z}_t-\Phi z_t)$, $\Omega^{*-1/2}T^{-1/2}\sum_{t=1}^{T-h}\hat{z}_t\varepsilon_{t+h}^*=$

$J_{1T}+J_{2T}+J_{3T}+J_{4T}$, where

$$\begin{aligned} J_{1T} &= \Omega^{*-1/2}T^{-1/2}\Phi\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}w_{t+h}^*,\ J_{2T}=\Omega^{*-1/2}T^{-1/2}\sum_{t=1}^{T-h}\left(\hat{z}_t-\Phi z_t\right)\varepsilon_{t+h}w_{t+h}^*, \\ J_{3T} &= -\Omega^{*-1/2}T^{-1/2}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\left(\hat{\delta}-\delta\right)w_{t+h}^*,\text{ and } J_{4T}=\Omega^{*-1/2}T^{-1/2}\sum_{t=1}^{T-h}\hat{z}_t\left(HF_t-\tilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^*. \end{aligned}$$

We first show that $J_{iT}=o_{P^*}(1)$ for $i=2$, 3 and 4. Starting with J_{2T}, we have that

$$\|J_{2T}\|\leq\left\|\Omega^{*-\frac{1}{2}}\right\|T^{-1/2}\left\|\sum_{t=1}^{T-h}\left(\hat{z}_t-\Phi z_t\right)\varepsilon_{t+h}w_{t+h}^*\right\|=O_{P^*}\left(T^{-1/2}\left(T^{1/2}l_T\right)^{1/2}\right)=O_{P^*}\left(T^{-1/4}l_T^{1/2}\right)=o_{P^*}(1),$$

since $T^{-1}l_T^2\to0$ and $J_{2.1,T}\equiv E^*\left\|\sum_{t=1}^{T-h}\left(\hat{z}_t-\Phi z_t\right)\varepsilon_{t+h}w_{t+h}^*\right\|^2=O_P\left(T^{1/2}l_T\right)$. Indeed, noting that $E^*\left(w_{t+h}^*w_{s+h}^*\right)=k_{dwb}\left(\frac{t-s}{l_T}\right)$,

$$\begin{aligned} J_{2.1,T} &= \sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\left(\hat{z}_t-\Phi z_t\right)'\left(\hat{z}_s-\Phi z_s\right)\varepsilon_{t+h}\varepsilon_{s+h}E^*\left(w_{t+h}^*w_{s+h}^*\right) \\ &\leq 2T^{1/2}l_T\left(\sum_{t=1}^{T-h}\|\hat{z}_t-\Phi z_t\|^4\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T-h}\varepsilon_{t+h}^4\right)^{1/2}l_T^{-1}\sum_{\tau=0}^{T-h}\left|k_{dwb}\left(\frac{\tau}{l_T}\right)\right|=O_P\left(T^{1/2}l_T\right). \end{aligned}$$

For J_{3T}, we have that

$$\|J_{3T}\|\leq\left\|\Omega^{*-1/2}\right\|T^{-1/2}\left\|\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\left(\hat{\delta}-\delta\right)w_{t+h}^*\right\|=O_{P^*}\left(T^{-1/2}l_T^{1/2}\right)=o_{P^*}(1),$$

where $J_{3.1,T}\equiv E^*\left\|\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\left(\hat{\delta}-\delta\right)w_{t+h}^*\right\|^2=O_P(l_T)$. Indeed,

$$J_{3.1,T}=\left(\delta-\hat{\delta}\right)'\left(\sum_{s=1}^{T-h}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\hat{z}_s\hat{z}_s'E^*\left(w_{t+h}^*w_{s+h}^*\right)\right)\left(\delta-\hat{\delta}\right)\leq 2T\left\|\delta-\hat{\delta}\right\|^2\sum_{\tau=0}^{T-h}\left|k_{dwb}\left(\frac{\tau}{l_T}\right)\right|T^{-1}\sum_{t=1}^{T-h}\|z_t\|^4,$$

which is $O_P(l_T)$. Similarly, we can show that $J_{4,T}=O_{P^*}\left(T^{-1/4}l_T^{1/2}\right)=o_{P^*}(1)$. It remains to show that $J_{1,T}\to^{d^*}N(0,I)$ in probability. For this purpose, we use Theorem 3.1 of Shao (2010) by verifying his assumptions. In particular, as $\{z_t\varepsilon_{t+h}\}$ are strong mixing of size $-\frac{3r}{r-2}$ for some $r>2$ with $E\|z_t\varepsilon_{t+h}\|^{2r}<C<\infty$, we have that $\sum_{j=1}^{\infty}\alpha(j)^{\frac{r}{r+2}}<\infty$ verifying his Assumption 3.1. We also have that $\sum_{j=1}^{\infty}j^2\alpha(j)^{\frac{r-2}{r}}<\infty$ and $E\|z_t\varepsilon_{t+h}\|^{2r}<C<\infty$, thus verifying Shao's Assumption 3.2 (by Lemma 1 of Andrews (1991)). Finally, we verify Condition E* of GP (2014). Following Lemma 2, we can write $\Omega^*=B_{1T}+B_{2T}+B_{3T}+B_{2T}'+B_{4T}+B_{5T}+B_{3T}'+B_{5T}'+B_{6T}$, with

$$
\begin{aligned}
B_{1T} &= T^{-1}E^*\left(\sum_{t=1}^{T-h}\hat{z}_t\varepsilon_{t+h}w_{t+h}^*\right)\left(\sum_{s=1}^{T-h}\hat{z}_s\varepsilon_{s+h}w_{s+h}^*\right)',\\
B_{2T} &= T^{-1}E^*\left(\sum_{t=1}^{T-h}\hat{z}_t\varepsilon_{t+h}w_{t+h}^*\right)\left(\sum_{s=1}^{T-h}\hat{z}_s\hat{z}_s'\left(\delta-\hat{\delta}\right)w_{s+h}^*\right)'\\
B_{3T} &= T^{-1}E^*\left[\left(\sum_{t=1}^{T-h}\hat{z}_t\varepsilon_{t+h}w_{t+h}^*\right)\left(\sum_{s=1}^{T-h}\hat{z}_s\left(HF_s-\tilde{F}_s\right)'\left(H^{-1}\right)'\alpha w_{s+h}^*\right)'\right],\\
B_{4T} &= T^{-1}E^*\left[\left(\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\left(\delta-\hat{\delta}\right)w_{t+h}^*\right)\left(\sum_{s=1}^{T-h}\hat{z}_s\hat{z}_s'\left(\delta-\hat{\delta}\right)w_{s+h}^*\right)'\right],\\
B_{5T} &= T^{-1}E^*\left[\left(\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\left(\delta-\hat{\delta}\right)w_{t+h}^*\right)\left(\sum_{s=1}^{T-h}\hat{z}_s\left(HF_s-\tilde{F}_s\right)'\left(H^{-1}\right)'\alpha w_{s+h}^*\right)'\right]
\end{aligned}
$$

and

$$
B_{6T}=T^{-1}E^*\left[\left(\sum_{t=1}^{T-h}\hat{z}_t\left(HF_t-\tilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^*\right)\left(\sum_{s=1}^{T-h}\hat{z}_s\left(HF_s-\tilde{F}_s\right)'\left(H^{-1}\right)'\alpha w_{s+h}^*\right)'\right].
$$

We show that each of B_{iT}, $i=2,3,\ldots,6$ are $o_P(1)$ and that $B_{1T}\to^P\Phi_0\Omega\Phi_0'$. Starting with B_{2T},

$$
\|B_{2T}\|\le T^{-1}\left(E^*\left\|\sum_{t=1}^{T-h}\hat{z}_t\varepsilon_{t+h}w_{t+h}^*\right\|^2\right)^{1/2}\left(E^*\left\|\sum_{s=1}^{T-h}\hat{z}_s\hat{z}_s'\left(\delta-\hat{\delta}\right)w_{s+h}^*\right\|\right)^{1/2}\equiv T^{-1}\left(B_{2.1,T}\right)^{1/2}\left(J_{3.1,T}\right)^{1/2},
$$

where $J_{3.1,T}=O_P(l_T)$, as shown above, and

$$
\begin{aligned}
B_{2.1,T} &= E^*\left\|\sum_{t=1}^{T-h}\Phi z_t\varepsilon_{t+h}w_{t+h}^*+\sum_{t=1}^{T-h}\left(\hat{z}_t-\Phi z_t\right)\varepsilon_{t+h}w_{t+h}^*\right\|^2\\
&\le 2E^*\left\|\sum_{t=1}^{T-h}\Phi z_t\varepsilon_{t+h}w_{t+h}^*\right\|^2+2E^*\left\|\sum_{t=1}^{T-h}\left(\hat{z}_t-\Phi z_t\right)\varepsilon_{t+h}w_{t+h}^*\right\|^2,
\end{aligned}
$$

where we can show that the first term is $O_P(T)$ and the second term is identical to $J_{2.1,T}=O_P\left(T^{1/2}l_T\right)$. Hence, $B_{2.1,T}=O_P(T)$ given that $l_T=o\left(\sqrt{T}\right)$ by assumption. This implies that $\|B_{2T}\|=O_P\left(T^{-1/2}l_T^{1/2}\right)=o_P(1)$. For B_{3T}, by Cauchy-Schwarz inequality,

$$
\|B_{3T}\|\le T^{-1}\left(B_{2.1,T}\right)^{1/2}\left(B_{3.1,T}\right)^{1/2}=O_p\left(\left(l_T^2/T\right)^{1/4}\right),
$$

where $B_{2.1,T}=O_P(T)$ and $B_{3.1,T}\equiv E^*\left\|\sum_{s=1}^{T-h}\hat{z}_s\left(HF_s-\tilde{F}_s\right)'\left(H^{-1}\right)'\alpha w_{s+h}^*\right\|^2=O_P\left(T^{1/2}l_T\right)$, as can be shown using similar arguments as above. For B_{4T}, note that $\|B_{4T}\|\le T^{-1}J_{3.1,T}=$

$O_P\left(T^{-1}l_T\right) = o_p\left(1\right)$. For B_{5T}, by Cauchy-Schwarz inequality, we have that $\|B_{5T}\| \leq T^{-1}\left(J_{3.1,T}\right)^{1/2}\left(B_{3.1,T}\right)^{1/2}$ $O_P\left(T^{-1}\right)O_P\left(l_T^{1/2}\right)O_P\left(T^{1/4}l_T^{1/2}\right) = O_P\left(T^{-3/4}l_T\right) = o_P\left(1\right)$. For B_{6T}, note that $\|B_{6T}\| \leq T^{-1}B_{3.1,T} = O_P\left(T^{-1/2}l_T\right) = o_P\left(1\right)$. Finally, note that

$$B_{1T} = T^{-1}\Phi E^*\left(\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}w_{t+h}^*\right)\left(\sum_{s=1}^{T-h} z_t\varepsilon_{t+h}w_{t+h}^*\right)'\Phi' + B_{1.2,T},$$

where $\|B_{1.2,T}\| \leq T^{-1}J_{2.1,T} = O_P\left(T^{-1/2}l_T\right) = o_P\left(1\right)$. The first term converges in probability to $\Phi_0\Omega\Phi_0'$ given that $\Phi \to^P \Phi_0$ and

$$T^{-1}E^*\left(\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}w_{t+h}^*\right)\left(\sum_{s=1}^{T-h} z_t\varepsilon_{t+h}w_{t+h}^*\right)' = T^{-1}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h} z_t z_s'\varepsilon_{t+h}\varepsilon_{s+h}k_{dwb}\left(\frac{s-t}{l_T}\right).$$

□

Proof of Lemma 5. The proof follows closely that of Lemma 2 and therefore we omit the details.

□

.2 Proofs of results in Chapter 2

The proof of Theorem 6 requires the following auxiliary result, which is the bootstrap analogue of Lemma A.2 of Bai (2003). It is based on the following identity that holds for each t :

$$\tilde{F}_t^* - H^* \tilde{F}_t = \tilde{V}^{*-1} \left(\underbrace{\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s^* \gamma_{st}^*}_{\equiv A_{1t}^*} + \underbrace{\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s^* \zeta_{st}^*}_{\equiv A_{2t}^*} + \underbrace{\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s^* \eta_{st}^*}_{\equiv A_{3t}^*} + \underbrace{\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s^* \xi_{st}^*}_{\equiv A_{4t}^*} \right),$$

where

$$\begin{aligned} \gamma_{st}^* &= E^* \left(\frac{1}{N} \sum_{i=1}^{N} e_{is}^* e_{it}^* \right), \ \zeta_{st}^* = \frac{1}{N} \sum_{i=1}^{N} \left(e_{is}^* e_{it}^* - E^* \left(e_{is}^* e_{it}^* \right) \right), \\ \eta_{st}^* &= \frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_i' \tilde{F}_s e_{it}^* = \tilde{F}_s' \frac{\tilde{\Lambda}' e_t^*}{N} \text{ and } \xi_{st}^* = \frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_i' \tilde{F}_t e_{is}^* = \eta_{ts}^*. \end{aligned}$$

Lemma 16. *Assume Assumptions 1 and 2 hold. Under Condition $\mathcal{A}$, we have that for each t, in probability, as $N, T \to \infty$,*

(a) $T^{-1} \sum_{s=1}^{T} \tilde{F}_s^* \gamma_{st}^* = O_{P^*} \left(\frac{1}{\sqrt{T} \delta_{NT}} \right) + O_{P^*} \left(\frac{1}{T^{3/4}} \right);$

(b) $T^{-1} \sum_{s=1}^{T} \tilde{F}_s^* \zeta_{st}^* = O_{P^*} \left(\frac{1}{\sqrt{N} \delta_{NT}} \right);$

(c) $T^{-1} \sum_{s=1}^{T} \tilde{F}_s^* \eta_{st}^* = O_{P^*} \left(\frac{1}{\sqrt{N}} \right);$

(d) $T^{-1} \sum_{s=1}^{T} \tilde{F}_s^* \xi_{st}^* = O_{P^*} \left(\frac{1}{\sqrt{N} \delta_{NT}} \right).$

Remark 3. *The term $O_{P^*} \left(1/T^{3/4} \right)$ that appears in (a) is of a larger order of magnitude than the corresponding term in Bai (2003, Lemma A.2(i)), which is $O_P (1/T)$. The reason why we obtain this larger term is that we rely on Bonferroni's inequality and Chebyshev's inequality to bound $\max_{1 \le s \le T} \|F_s\| = O_P \left(T^{1/4} \right)$ using the fourth order moment assumption on F_s (cf. Assumption 1(a)). In general, if $E \|F_s\|^q \le M$ for all s, then $\max_{1 \le s \le T} \|F_s\| = O_P \left(T^{1/q} \right)$ and we will obtain a term of order $O_{P^*} \left(1/T^{1-1/q} \right)$.*

Proof of Lemma 16. The proof follows closely that of Lemma A.2 of Bai (2003). The only exception is (a), where an additional $O \left(\frac{1}{T^{3/4}} \right)$ term appears. In particular, we write

$$T^{-1} \sum_{s=1}^{T} \tilde{F}_s^* \gamma_{st}^* = T^{-1} \sum_{s=1}^{T} \left(\tilde{F}_s^* - H^* \tilde{F}_s \right) \gamma_{st}^* + H^* T^{-1} \sum_{s=1}^{T} \tilde{F}_s \gamma_{st}^* = a_t^* + b_t^*.$$

We use Cauchy-Schwartz and Condition $\mathcal{A}$.1 to bound a_t^* as follows

$$\begin{aligned} \|a_t^*\| &\le \left(T^{-1} \sum_{s=1}^{T} \left\| \tilde{F}_s^* - H^* \tilde{F}_s \right\|^2 \right)^{1/2} \left(T^{-1} \sum_{s=1}^{T} |\gamma_{st}^*|^2 \right)^{1/2} \\ &= O_{P^*} \left(\frac{1}{\delta_{NT}} \right) O_P \left(\frac{1}{\sqrt{T}} \right) = O_{P^*} \left(\frac{1}{\delta_{NT} \sqrt{T}} \right), \end{aligned}$$

where $T^{-1}\sum_{s=1}^{T}\left\|\tilde{F}_s^* - H^*\tilde{F}_s\right\|^2 = O_{P^*}\left(\delta_{NT}^{-2}\right)$ by Lemma 3.1 of Gonçalves and Perron (2014) (note that this lemma only requires Conditions A*(b), A*(c), and B*(d), which correspond to our Condition $\mathcal{A}$.1, $\mathcal{A}$.2 and $\mathcal{A}$.5). For b_t^*, we have that (ignoring H^*, which is $O_{P^*}(1)$),

$$b_t^* = T^{-1}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}^* = T^{-1}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st}^* + HT^{-1}\sum_{s=1}^{T}F_s\gamma_{st}^* = b_{1t}^* + b_{2t}^*,$$

where $b_{1t}^* = O_P\left(1/\delta_{NT}\sqrt{T}\right)$ using the fact that $T^{-1}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$ under Assumptions 1 and 2 and the fact that $T^{-1}\sum_{s=1}^{T}|\gamma_{st}^*|^2 = O_P(1/T)$ for each t by Condition $\mathcal{A}$.1. For b_{2t}^*, note that (ignoring $H = O_P(1)$),

$$\|b_{2t}^*\| \leq \underbrace{\left(\max_s \|F_s\|\right)}_{O_P\left(T^{1/4}\right)}\underbrace{T^{-1}\sum_{s=1}^{T}|\gamma_{st}^*|}_{O_P\left(\frac{1}{T}\right)} = O_P\left(\frac{1}{T^{3/4}}\right),$$

where we have used the fact that $E\|F_s\|^4 \leq M$ for all s (Assumption 1) to bound $\max_s \|F_s\|$. Indeed, by Bonferroni's inequality and Chebyshev's inequality, we have that

$$P\left(T^{-1/4}\max_s \|F_s\| > M\right) \leq \sum_{s=1}^{T} P\left(\|F_s\| > T^{1/4}M\right) \leq \sum_{s=1}^{T}\frac{E\|F_s\|^4}{M^4 T} \leq \frac{1}{M^3} \to 0$$

for M sufficiently large. For (b), we follow exactly the proof of Bai (2003) and use Condition $\mathcal{A}$.2 to bound $T^{-1}\sum_{s=1}^{T}\zeta_{st}^{*2} = O_{P^*}\left(\frac{1}{N}\right)$ for each t; similarly, we use Condition $\mathcal{A}$.3 to bound $\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}^*$ for each t. For (c), we bound $T^{-1}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}^* = N^{-1}H^*\sum_{i=1}^{N}\tilde{\lambda}_i e_{it}^* = O_{P^*}\left(1/\sqrt{N}\right)$ by using Condition $\mathcal{A}$.6. This same condition is used to bound $T^{-1}\sum_{s=1}^{T}\eta_{st}^{*2} = O_{P^*}(1/N)$ for each t. Finally, for part (d), we use Condition $\mathcal{A}$.4 to bound $T^{-1}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}^* = O_{P^*}\left(\frac{1}{\sqrt{NT}}\right)$ for each t and we use Condition $\mathcal{A}$.5 to bound $T^{-1}\sum_{s=1}^{T}\xi_{st}^{*2} = O_{P^*}(1/N)$ for each t.

Proof of Theorem 6. By Lemma 16, it follows that the third term in $\sqrt{N}\left(\tilde{F}_t^* - H^*\tilde{F}_t\right)$ is the dominant one (it is $O_{P^*}(1)$); the first term is $O_{P^*}\left(\frac{\sqrt{N}}{\sqrt{T}\delta_{NT}}\right) + O_P\left(\frac{\sqrt{N}}{T^{3/4}}\right) = O_{P^*}\left(\frac{\sqrt{N}}{T^{3/4}}\right) = o_{P^*}(1)$ if $\sqrt{N}/T^{3/4} \to 0$ whereas the second and the fourth terms are $O_{P^*}(1/\delta_{NT}) = o_{P^*}(1)$ as $N, T \to \infty$. Thus, we have that

$$\begin{aligned}
\sqrt{N}\left(\tilde{F}_t^* - H^*\tilde{F}_t\right) &= \tilde{V}^{*-1}\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^*\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_i'\tilde{F}_s e_{it}^* + o_{P^*}(1) \\
&= \left[\tilde{V}^{*-1}\left(\frac{\tilde{F}^{*\prime}\tilde{F}}{T}\right)\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)\right]\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_i e_{it}^* + o_{P^*}(1) \\
&= H^*\tilde{V}^{-1}\Gamma_t^{*1/2}\underbrace{\Gamma_t^{*-1/2}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_i e_{it}^*}_{\to^{d^*} N(0,I_r) \text{ by Condition } \mathcal{A}.6} + o_{P^*}(1), \qquad (6)
\end{aligned}$$

given the definition of H^* and the fact that $\tilde{V} = \frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}$. Since det $(\Gamma_t^*) > \epsilon > 0$ for all N and some ϵ, Γ_t^{*-1} exists and we can define $\Gamma_t^{*-1/2} = \left(\Gamma_t^{*1/2}\right)^{-1}$ where $\Gamma_t^{*1/2}\Gamma_t^{*1/2} = \Gamma_t^*$. Let $\Pi_t^{*-1/2} = \Gamma_t^{*-1/2}\tilde{V}$ and note that $\Pi_t^{*-1/2}$ is symmetric and it is such that $\left(\Pi_t^{*-1/2}\right)\left(\Pi_t^{*-1/2}\right) = \tilde{V}\Gamma_t^{*-1}\tilde{V} = \Pi_t^{*-1}$. The result follows by multiplying (6) by $\Pi_t^{*-1/2}H^{*-1}$ and using Condition $\mathcal{A}$.6.

Proof of Corollary 7. Condition $\mathcal{B}$ and the fact that $\tilde{V} \to^P V$ under our assumptions imply that $\Pi_t^* \to^P \Pi_t \equiv V^{-1}Q\Gamma_t Q'V^{-1}$. This suffices to show the result.

Proof of Theorem 8. Using the decomposition (2.19) and the fact that

$$\hat{z}_T^* = \Phi^*\hat{z}_T + \begin{pmatrix} \tilde{F}_T^* - H^*\tilde{F}_T \\ 0 \end{pmatrix},$$

where $\Phi^* = diag\left(H^*, I_q\right)$, it follows that

$$\hat{y}_{T+1|T}^* - y_{T+1|T}^* = \frac{1}{\sqrt{T}}\hat{z}_T'\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right) + \frac{1}{\sqrt{N}}\hat{\alpha}'\sqrt{N}\left(H^{*-1}\tilde{F}_T^* - \tilde{F}_T\right) + r_T^*,$$

where the remainder is

$$r_T^* = \frac{1}{\sqrt{T}}\left(\tilde{F}_T^* - H^*\tilde{F}_T\right)'\sqrt{T}\left(\hat{\alpha}^* - H^{*-1\prime}\hat{\alpha}\right) = O_{P^*}\left(\frac{1}{\sqrt{TN}}\right).$$

First, we argue that

$$\frac{\hat{y}_{T+1|T}^* - y_{T+1|T}^*}{\sqrt{B_T^*}} \to^{d^*} N(0,1), \tag{7}$$

where B_T^* is the asymptotic variance of $\hat{y}_{T+1|T}^* - y_{T+1|T}^*$, i.e. $B_T^* = aVar^*\left(\hat{y}_{T+1|T}^* - y_{T+1|T}^*\right) = \frac{1}{T}\hat{z}_T'\Sigma_\delta\hat{z}_T + \frac{1}{N}\hat{\alpha}'\Pi_T\hat{\alpha}$. To show (7), we follow the arguments of Bai and Ng (2006, proof of their Theorem 3) and show that (1) $Z_{1T}^* = \sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right) \to^{d^*} N\left(-c\Delta_\delta, \Sigma_\delta\right)$; (2) $Z_{2T}^* = \sqrt{N}\left(H^{*-1}\tilde{F}_T^* - \tilde{F}_T\right) \to^{d^*} N\left(0, \Pi_T\right)$; (3) Z_{1T}^* and Z_{2T}^* are asymptotically independent (conditional on the original sample). Condition (1) follows from Gonçalves and Perron (2014) under Assumptions 1-6; (2) follows from Corollary 7 provided $\sqrt{N}/T^{11/12} \to 0$ and conditions $\mathcal{A}$ and $\mathcal{B}$ hold for the wild bootstrap (which we verify next); (3) holds because we generate e_t^* independently of ε_{t+1}^*.

Proof of Condition $\mathcal{A}$ for the wild bootstrap. We verify for $t = T$. We have that $\sum_{s=1}^T \left|\gamma_{sT}^*\right|^2 = \left(\frac{1}{N}\sum_{i=1}^N \tilde{e}_{iT}^2\right)^2$. Thus, it suffices to show that $\frac{1}{N}\sum_{i=1}^N \tilde{e}_{iT}^2 = O_P(1)$. This follows by using the decomposition

$$\tilde{e}_{it} = e_{it} - \lambda_i'H^{-1}\left(\tilde{F}_t - HF_t\right) - \left(\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right)'\tilde{F}_t,$$

which implies that

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^N \left|\tilde{e}_{it}\right|^2 &\leq 3\frac{1}{N}\sum_{i=1}^N \left|e_{it}\right|^2 + 3\frac{1}{N}\sum_{i=1}^N \left\|\lambda_i\right\|^2\left\|H^{-1}\right\|^2\left\|\tilde{F}_T - HF_T\right\|^2 \\ &+3\frac{1}{N}\sum_{i=1}^N \left\|\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right\|^2\left\|\tilde{F}_T\right\|^2. \end{aligned}$$

The first term is $O_P(1)$ given that $E|e_{it}|^2 = O(1)$; the second term is $O_P(1)$ since $E\|\lambda_i\|^2 = O(1)$ and given that $\left\|\tilde{F}_T - HF_T\right\|^2 = O_P(1/N) = o_P(1)$; and the third term is $O_P(1)$ given Lemma C.1.(ii) of Gonçalves and Perron (2014) and the fact that $\left\|\tilde{F}_T\right\|^2 = O_P(1)$. Next, we verify $\mathcal{A}.2$. For $t = T$, following the proof of Theorem 4.1 in Gonçalves and Perron (2014) (condition A*(c)), we have that

$$\begin{aligned}
&\frac{1}{T}\sum_{s=1}^{T} E^*\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(e^*_{iT}e^*_{is} - E^*\left(e^*_{iT}e^*_{is}\right)\right)\right|^2 \\
= &\frac{1}{T}\sum_{s=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^2_{iT}\tilde{e}^2_{is}\underbrace{Var\left(\eta_{iT}\eta_{is}\right)}_{\leq\bar{\eta}} \leq \bar{\eta}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^2_{iT}\left(\frac{1}{T}\sum_{s=1}^{T}\tilde{e}^2_{is}\right) \\
\leq &\bar{\eta}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^4_{iT}\right)^{1/2}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\tilde{e}^4_{is}\right)^{1/2} = O_P(1), \qquad (8)
\end{aligned}$$

where the first factor in (8) can be bounded by an argument similar to that used above to bound $\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^2_{iT}$, and the second factor can be bounded by Lemma C.1 (iii) of Gonçalves and Perron (2014). $\mathcal{A}.3$ follows by an argument similar to that used by Gonçalves and Perron (2014) to verify Condition B*(b). In particular,

$$\begin{aligned}
&E^*\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_s\left(e^*_{is}e^*_{iT} - E^*\left(e^*_{is}e^*_{iT}\right)\right)\right\|^2 \\
= &\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s'\tilde{F}_s\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^2_{iT}\tilde{e}^2_{is}Var^*\left(\eta_{iT}\eta_{is}\right) \leq \bar{\eta}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^2_{iT}\left(\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s'\tilde{F}_s\tilde{e}^2_{is}\right) \\
\leq &\bar{\eta}\left[\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^4_{iT}\right]^{1/2}\left[\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^4\frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{s=1}^{T}\tilde{e}^4_{is}\right]^{1/2} = O_P(1),
\end{aligned}$$

under our assumptions. Conditions $\mathcal{A}.4$ and $\mathcal{A}.5$ correspond to Gonçalves and Perron's (2014) Conditions B*(c) and B*(d), respectively. Finally, we prove Condition $\mathcal{A}.6$ for $t = T$. Using the fact that $e^*_{iT} = \tilde{e}_{iT}\eta_{iT}$, where $\eta_{iT} \sim$ i.i.d. $(0,1)$ across i, note that

$$\Gamma^*_T = Var^*\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_i e^*_{iT}\right) = \frac{1}{N}\sum_{i=1}^{N}\tilde{\lambda}_i\tilde{\lambda}_i'\tilde{e}^2_{iT} \to^P Q\Gamma_T Q',$$

by Theorem 6 of Bai (2003), where $\Gamma_T \equiv \lim_{N\to\infty} Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{iT}\right) > 0$ by assumption. Thus, Γ^*_T is uniformly positive definite. We now need to verify that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\ell'\Gamma_T^{*-1/2}\tilde{\lambda}_i e^*_{iT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\underbrace{\ell'\Gamma_T^{*-1/2}\tilde{\lambda}_i\tilde{e}_{iT}\eta_{iT}}_{=\omega^*_{iT}} \to^{d^*} N(0,1),$$

in probability, for any ℓ such that $\ell'\ell = 1$. Since ω^*_{iT} is an heterogeneous array of independent random variables (given that η_{it} is i.i.d.), we apply a CLT for heterogeneous independent arrays.

Note that $E^*\left(\omega_{iT}^*\right)=0$ and

$$
\begin{aligned}
Var^*\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \omega_{iT}^*\right) &= \ell'\left(\Gamma_T^*\right)^{-1/2} Var^*\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\lambda}_i \tilde{e}_{iT}\eta_{iT}\right)\left(\Gamma_T^*\right)^{-1/2}\ell \\
&= \ell'\left(\Gamma_T^*\right)^{-1/2}\left(\frac{1}{N}\sum_{i=1}^N \tilde{\lambda}_i\tilde{\lambda}_i'\tilde{e}_{iT}^2\right)\left(\Gamma_T^*\right)^{-1/2}\ell = \ell'\ell = 1.
\end{aligned}
$$

Thus, it suffices to verify Lyapunov's condition, i.e. for some $r>1$, $\frac{1}{N^r}\sum_{i=1}^N E^*\left|\omega_{iT}^*\right|^{2r} \to^P 0$. We have that

$$
\begin{aligned}
\frac{1}{N^r}\sum_{i=1}^N E^*\left|\omega_{iT}^*\right|^{2r} &\leq \frac{1}{N^{r-1}}\|\ell\|^{2r}\left\|\left(\Gamma_T^*\right)^{-1/2}\right\|^{2r}\frac{1}{N}\sum_{i=1}^N\left\|\tilde{\lambda}_i\right\|^{2r}\left|\tilde{e}_{iT}\right|^{2r}\underbrace{E^*\left|\eta_{iT}\right|^{2r}}_{\leq M<\infty} \\
&\leq C\frac{1}{N^{r-1}}\left\|\left(\Gamma_T^*\right)^{-1/2}\right\|^{2r}\left(\frac{1}{N}\sum_{i=1}^N\left\|\tilde{\lambda}_i\right\|^{4r}\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^N\left|\tilde{e}_{iT}\right|^{4r}\right)^{1/2} = O_P\left(\frac{1}{N^{r-1}}\right) = o_P(1)
\end{aligned}
$$

Proof of Condition $\mathcal{B}$ for the wild bootstrap. $\Gamma_T^* = \frac{1}{N}\sum_{i=1}^N \tilde{\lambda}_i\tilde{\lambda}_i'\tilde{e}_{iT}^2 \to^P Q\Gamma_T Q'$, by Theorem 6 of Bai (2003).

The result for the studentized statistic (where we replace B_T^* with an estimate $\hat{B}_T^*$) then follows by showing that $\hat{z}_T^{*\prime}\hat{\Sigma}_\delta^*\hat{z}_T^* - \hat{z}_T'\Sigma_\delta\hat{z}_T \to^{P^*} 0$, and $\hat{\alpha}^{*\prime}\hat{\Sigma}_{\tilde{F}_T}^*\hat{\alpha}^* - \hat{\alpha}'\hat{\Sigma}_{\tilde{F}_T}\hat{\alpha} \to^{P^*} 0$, in probability. This can be shown using the arguments in Bai and Ng (2006, Theorems 3.1) and Bai (2003, Theorem 6).

Proof of Lemma 9. Recall that $F_\varepsilon(x) = P(\varepsilon_t \leq x)$ and define the following empirical distribution functions,

$$
F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}(x) = \frac{1}{T-1}\sum_{t=1}^{T-1} 1\left\{\hat{\varepsilon}_{t+1} - \bar{\hat{\varepsilon}} \leq x\right\} \text{ and } F_{T,\varepsilon}(x) = \frac{1}{T-1}\sum_{t=1}^{T-1} 1\left\{\varepsilon_{t+1}\leq x\right\},
$$

where $\bar{\hat{\varepsilon}} = \frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}$. Note that $F_{T,\varepsilon^*}(x) = F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}(x)$. It follows that

$$
d_2\left(F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}, F_\varepsilon\right) \leq d_2\left(F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}, F_{T,\varepsilon}\right) + d_2\left(F_{T,\varepsilon}, F_\varepsilon\right),
$$

where $d_2\left(F_{T,\varepsilon}, F_\varepsilon\right) = o_{a.s.}(1)$ by Lemma 8.4 of Bickel and Freedman (1981). Thus, it suffices to show that $d_2\left(F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}, F_{T,\varepsilon}\right) = o_P(1)$. Let I be distributed uniformly on $\{1,\ldots,T-1\}$ and define $X_1 = \hat{\varepsilon}_{I+1} - \bar{\hat{\varepsilon}}$ and $Y_1 = \varepsilon_{I+1}$. We have that

$$
\begin{aligned}
\left(d_2\left(F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}, F_{T,\varepsilon}\right)\right)^2 &\leq E\left(X_1 - Y_1\right)^2 = E_I\left(\hat{\varepsilon}_{I+1} - \bar{\hat{\varepsilon}} - \varepsilon_{I+1}\right)^2 = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1} - \bar{\hat{\varepsilon}} - \varepsilon_{t+1}\right)^2 \\
&= \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1} - \varepsilon_{t+1}\right)^2 - 2\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1} - \varepsilon_{t+1}\right)\bar{\hat{\varepsilon}} + \left(\bar{\hat{\varepsilon}}\right)^2 \equiv A_1 + A_2 + A_3.
\end{aligned}
$$

We can write

$$
\hat{\varepsilon}_{t+1} - \varepsilon_{t+1} = -\left(\tilde{F}_t - HF_t\right)'\hat{\alpha} - \left(\Phi z_t\right)'\left(\hat{\delta} - \delta\right),
$$

where $\Phi = diag\left(H, I_q\right)$. This implies that

$$A_1 \leq 2\frac{1}{T-1}\sum_{t=1}^{T-1}\left\|\tilde{F}_t - HF_t\right\|^2\|\hat{\alpha}\|^2 + 2\frac{1}{T-1}\sum_{t=1}^{T-1}\|\Phi z_t\|^2\left\|\hat{\delta}-\delta\right\|^2 = O_P\left(\frac{1}{\delta_{NT}^2}\right) + O_P\left(\frac{1}{T}\right) = o_P\left(1\right).$$

Similarly,

$$\bar{\hat{\varepsilon}} = \frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1} = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1} - \varepsilon_{t+1}\right) + \frac{1}{T-1}\sum_{t=1}^{T-1}\varepsilon_{t+1} = O_P\left(\frac{1}{\delta_{NT}}\right) + o_P\left(1\right),$$

where the first term is bounded by an argument similar to that used to bound A_1 (via the Cauchy-Schwartz inequality). This implies that A_2 and A_3 are also $o_P\left(1\right)$.

Proof of Corollary 10. Lemma 9 implies that $s_{T+1}^* \to^{d^*} 1 - F_\varepsilon\left(-x\sigma_\varepsilon\right)$, in probability. Since $s_{T+1} \to^d 1 - F_\varepsilon\left(-x\sigma_\varepsilon\right)$ and F_ε is everywhere continuous under Assumption 7, Polya's Theorem implies the result.

.3 Proofs of results in Chapter 3

Lemma 17. *Under Assumption A1-A3,*

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

and

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

Proof of Lemma 17. We have the following decomposition

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

where

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

and

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

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

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

where $b_{f1} = O_p\left(\frac{1}{C_{NT}T^{1/2}}\right)$, $b_{f2} = O_p\left(\frac{1}{N^{1/2}T^{1/2}}\right)$, $b_{f3} = O_p\left(\frac{1}{N^{1/2}T^{1/2}}\right)$ and $b_{f4} = \frac{1}{N}Q\Gamma Q'V^{-1} + O_p\left(\frac{1}{N^{1/2}T^{1/2}}\right)$. Hence, $\frac{1}{T-1}F'\left[\widetilde{F}-FH'\right] = O_p\left(\frac{1}{C_{NT}^2}\right)$, similarly $\frac{1}{T-1}W'\left[\widetilde{F}-FH'\right] = O_p\left(\frac{1}{C_{NT}^2}\right)$, thus $L_{1T} = O_p\left(\frac{1}{C_{NT}^4}\right)$. Since $\frac{1}{T-1}Z^{0\prime}Z(m) = O_p(1)$, similarly, $L_{2T} = O_p\left(\frac{1}{C_{NT}^2}\right)$. To finish, $L_{3T} = O_p\left(\frac{1}{C_{NT}^2}\right)$ as

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

Indeed,

$$\begin{aligned}
&\left(\frac{1}{T-1}\widetilde{Z}(m)'\widetilde{Z}(m)\right)^{-1} - \left(\frac{1}{T-1}Z(m)'Z(m)\right)^{-1} \\
= &\left(\frac{1}{T-1}\widetilde{Z}(m)'\widetilde{Z}(m)\right)^{-1}\left(\frac{1}{T-1}Z(m)'Z(m) - \frac{1}{T-1}\widetilde{Z}(m)'\widetilde{Z}(m)\right)\left(\frac{1}{T-1}Z(m)'Z(m)\right)^{-1} \\
= &\left(\frac{1}{T-1}\widetilde{Z}(m)'\widetilde{Z}(m)\right)^{-1}\left(\frac{1}{T-1}Z(m)'\left(Z(m) - \widetilde{Z}(m)\right) + \frac{1}{T-1}\left(Z(m) - \widetilde{Z}(m)\right)'\widetilde{Z}(m)\right) \\
&\times\left(\frac{1}{T-1}Z(m)'Z(m)\right)^{-1}
\end{aligned}$$

since $\frac{1}{T-1}\left(Z(m) - \widetilde{Z}(m)\right)'\widetilde{Z}(m) = O_p\left(\frac{1}{C_{NT}^2}\right)$ and $\frac{1}{T-1}Z(m)'\left(Z(m) - \widetilde{Z}(m)\right) = O_p\left(\frac{1}{C_{NT}^2}\right)$. □

Proof of Lemma 11. To prove Lemma 11, we will first need to show that

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

We have the following decomposition

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

Hence

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

Thus

$$
\left|\max_{1\leq t\leq T-1}\left\|\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right\|-\max_{1\leq t\leq T-1}\left\|Z_t(m)'\left(Z'(m)Z(m)\right)^{-1}Z_t(m)\right\|\right|\leq A_1+A_2+A_3
$$

where

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

and

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

Since

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

and

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

$A_1=O_p\left(\frac{1}{C_{NT}^2}\right)$. Because,

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

$A_2=O_p\left(\frac{1}{C_{NT}^2}\right)$. From Proposition 2 of Bai (2003),

$$
\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)-Z_t(m)\right\|=O_p\left(\frac{1}{T^{1/2}}\right)+O_p\left(\frac{T^{1/2}}{N^{1/2}}\right),
$$

it follows that $A_3 = O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N^{1/2}}\right)$ as $\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)\right\| = O_p\left(T^{1/2}\right)$ since

$$\frac{1}{T-1}\max_{1\leq t\leq T-1}\left\|\hat{Z}_t(m)\right\|^2 = O_p(1)$$

. We then deduce

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

and

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

as by A4 (b)

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

The remaining part of the proof goes similarly as the proof of (3.4) and (3.4) of Shao (1993). Firstly,

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

By Taylor expansion,

$$\begin{aligned}\left(1-\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right)^{-1} &= 1+2\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\\ &\quad +O_p\left[\left(\hat{Z}_t'(m)\left(\hat{Z}'(m)\hat{Z}(m)\right)^{-1}\hat{Z}_t(m)\right)^2\right].\end{aligned}$$

Hence

$$CV_1(m) = A_4 + 2A_5 + o_p(A_5)$$

where

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

since

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

Given the decomposition $\hat{\varepsilon}(m) = \varepsilon + \mu - \widetilde{\mu}(m)$ where $\mu = F^0\alpha + W\beta$ and $\widetilde{\mu}(m) = \widetilde{P}(m)\mu + \widetilde{P}(m)\varepsilon$,

$$A_4 = \frac{1}{T-1}\varepsilon'\varepsilon + \widetilde{L}_T(m) - 2r_{1T}(m)$$

with

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

and

$$r_{1T}(m) = \frac{1}{T-1}\left(\widetilde{\mu}(m) - \mu\right)'\varepsilon = \frac{1}{T-1}\varepsilon'\widetilde{P}(m)\varepsilon - \frac{1}{T-1}\mu'\left(I - \widetilde{P}(m)\right)\varepsilon.$$

Hence,

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

Under our Assumption $A_1 - A_3$,

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

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

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

It then follows that

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

(1) For incorrect model,

$$A_4 = \frac{1}{T-1}\varepsilon'\varepsilon + \frac{1}{T-1}\mu'\left(I - P(m)\right)\mu + o_p(1)$$

since $\frac{1}{T-1}\varepsilon' P(m)\varepsilon = o_p(1)$ and $\frac{1}{T-1}\mu'\left(I - P(m)\right)\varepsilon = o_p(1)$. We also have that

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

Hence, $A_5 = o_p(1)$ because the first term in the right hand side is $o_p(1)$ given Assumption $A4(b)$ and the second is equal to A_2 which is $O_p(1)$. Thus, for $m \in \mathcal{M}_2$

$$\begin{aligned} CV_1(m) &= \frac{1}{T-1}\varepsilon'\varepsilon + \frac{1}{T-1}\mu'\left(I - P(m)\right)\mu + o_p(1) \\ &= \sigma^2 + \frac{1}{T-1}\mu'\left(I - P(m)\right)\mu + o_p(1). \end{aligned}$$

(2) Because $\mu'\left(I - P(m)\right)\mu = 0$, $\mu'\left(I - P(m)\right)\varepsilon = 0$ for correct model,

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

More specifically, $A_5 = \frac{(r(m)+q)}{T-1}\sigma^2 + o_p\left(\frac{1}{T-1}\right)$ for models in $\mathcal{M}_2$. Indeed, because

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

and $\frac{1}{T-1}\hat{Z}'(m)\hat{Z}'(m) = \Sigma_{ZZ}(m) + o_p(1)$, it follows that

$$\begin{aligned} A_5 &= \frac{1}{T-1} trace \left[\left(\Sigma_{ZZ}(m)^{-1} + o_p(1) \right) \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{Z}_t(m) \hat{Z}_t'(m) \hat{\varepsilon}_{t+1}^2(m) \right] \\ &= \frac{1}{T-1} trace \left[\left(\Sigma_{ZZ}(m)^{-1} + o_p(1) \right) \left(\sigma^2 \Sigma_{ZZ}(m) + o_p(1) \right) \right] = \frac{(r(m)+q)}{T-1}\sigma^2 + o_p\left(\frac{1}{T-1}\right). \end{aligned}$$

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

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

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

Lemma 18. *Under Assumptions A1-A4, as* b, T, $N \to \infty$,

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

Démonstration. The proof uses the following identity

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

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

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

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

Using the Cauchy-Schwartz inequality, we have

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

In addition, $\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{F}_s\right\|^2 = O_p(1)$ 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 $\frac{1}{\sqrt{T\cdot\kappa}}\sum_{t\,\in\,s^c}\sum_{u=1}^{T}|\gamma_{ut}| \le C$ easily using the proof of Lemma 1 (i) of Bai and Ng (2002)). In consequence, $\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{1t}\|^2 = O_p\left(\frac{\sqrt{\kappa}}{\sqrt{T}}\right)$. By repeated application of Cauchy-Schwartz inequality,

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

Hence,

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

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

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

since $\max_{u,t}E\left(\zeta_{ut}^4\right) = O\left(\frac{1}{N^2}\right)$ by Assumption A1.(e) Thus, $\sum_{t\,\in\,s^c}\|A_{2t}\|^2 = O_p\left(\frac{\kappa}{N^2}\right)$. Thirdly, as $\frac{1}{b\cdot\kappa}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\frac{1}{N^{1/2}}\Lambda e_t\right\|^4 = O_p(1)$,

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

implies that

$$\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{3t}\|^4 \leq \frac{\kappa}{N^2}\left[\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\frac{1}{T_c}\sum_{t\,\in\,s^c}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_ie_t\right\|^4\right]\left(\frac{1}{T}\sum_{u=1}^{T}\left\|\widetilde{F}_u\right\|^2\right)^2\left(\frac{1}{T}\sum_{u=1}^{T}\|F_u\|^2\right)^2 = O_p\left(\frac{\kappa}{N^2}\right),$$

since $\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{F}_s\right\|^2 = O_p(1)$, $\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2 = O_p(1)$. The proof that $\frac{1}{b}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|A_{4t}\|^4 = O_p\left(\frac{\kappa}{N^2}\right)$ uses $\frac{1}{T}\sum_{u=1}^{T}\left\|\widetilde{F}_u\right\|^2 = O_p(1)$, $\frac{1}{b\cdot\kappa}\sum_{s\,\in\,\mathcal{R}}\sum_{t\,\in\,s^c}\|F_t\|^4 = O_p(1)$, $\frac{1}{T}\sum_{u=1}^{T}\left\|\frac{1}{\sqrt{N}}\Lambda' e_u\right\|^2 = O_p(1)$ and the following bound

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

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

Proof of Theorem 12. We have the following decomposition

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

where

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

and

$$B_3=\frac{1}{d\times b}\sum_{s\in\mathcal{R}}\left\|(y_s-P_{s^c}(m)\,y_{s^c})\right\|^2$$

with

$$P_{s^c}(m)=Z_s(m)\left(Z_{s^c}(m)'\,Z_{s^c}(m)\right)^{-1}Z_{s^c}(m)'$$

and

$$\widetilde{P}_{s^c}(m)=\hat{Z}_s(m)\left(\hat{Z}_{s^c}(m)'\,\hat{Z}_{s^c}(m)\right)^{-1}\hat{Z}_{s^c}(m)'.$$

The proofs will be done into two parts.

Part 1 : After decomposing $\left(P_{s^c}(m)-\widetilde{P}_{s^c}(m)\right)y_{s^c}$ and using the cr-inequality, it follows that

$$B_1\le \underbrace{4\frac{1}{d\times b}\sum_{s\in\mathcal{R}}\|B_{11s}\|^2}_{B_{11}}+\underbrace{4\frac{1}{d\times b}\sum_{s\in\mathcal{R}}\|B_{12s}\|^2}_{B_{12}}+\underbrace{4\frac{1}{d\times b}\sum_{s\in\mathcal{R}}\|B_{13s}\|^2}_{B_{13}}+\underbrace{4\frac{1}{d\times b}\sum_{s\in\mathcal{R}}\|B_{14s}\|^2}_{B_{14}}.$$

with

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

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

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

and

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

Starting with B_{11}, it follows that

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

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

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

Because $\frac{1}{d}\left\|\hat{Z}(m)\right\|^2 = O_p(1)$, let's show that

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

and

$$B_{112} = \frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\,y_{s^c}\right\|^4 = O_p\left(\frac{T}{\kappa^3}\right).$$

We have the following decomposition

$$\begin{aligned}
& \left(\frac{1}{\kappa}Z'_{s^c}(m)\,Z_{s^c}(m)\right)^{-1} - \left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\,\hat{Z}_{s^c}(m)\right)^{-1} \\
= & \left(\frac{1}{\kappa}Z'_{s^c}(m)\,Z_{s^c}(m)\right)^{-1}\left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\,\hat{Z}_{s^c}(m) - \frac{1}{\kappa}Z'_{s^c}(m)\,Z_{s^c}(m)\right)\left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\,\hat{Z}_{s^c}(m)\right)^{-1} \\
= & \left(\frac{1}{\kappa}Z'_{s^c}(m)\,Z_{s^c}(m)\right)^{-1}\left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\left(\hat{Z}_{s^c}(m) - Z'_{s^c}(m)\right)\right)\left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\,\hat{Z}_{s^c}(m)\right)^{-1} \\
& +\left(\frac{1}{\kappa}Z'_{s^c}(m)\,Z_{s^c}(m)\right)^{-1}\left(\frac{1}{\kappa}\left(\hat{Z}_{s^c}(m) - Z'_{s^c}(m)\right)'Z_{s^c}(m)\right)\left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\,\hat{Z}_{s^c}(m)\right)^{-1}
\end{aligned}$$

Hence by the c-r inequality,

$$\begin{aligned}
B_{111} \leq\ & \left\|(\Sigma_{ZZ}(m))^{-1}\right\|^4 \frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\left(\frac{1}{\kappa}\hat{Z}'_{s^c}(m)\left(\hat{Z}_{s^c}(m) - Z'_{s^c}(m)\right)\right) + \left(\frac{1}{\kappa}\left(\hat{Z}_{s^c}(m) - Z'_{s^c}(m)\right)'Z_{s^c}(m)\right)\right\|^4 \\
& \times\left\|(\Sigma_{ZZ}(m))^{-1}\right\|^4(1+o_p(1)) \\
\leq\ & 2^3\left\|(\Sigma_{ZZ}(m))^{-1}\right\|^4 \frac{1}{b\cdot\kappa^4}\sum_{s\in\mathcal{R}}\left\|\hat{Z}'_{s^c}(m)\left(\hat{Z}_{s^c}(m) - Z'_{s^c}(m)\right)\right\|^4\left\|(\Sigma_{ZZ}(m))^{-1}\right\|^4(1+o_p(1)) \\
& +2^3\left\|(\Sigma_{ZZ}(m))^{-1}\right\|^4 \frac{1}{b\cdot\kappa^4}\sum_{s\in\mathcal{R}}\left\|\left(\hat{Z}_{s^c}(m) - Z'_{s^c}(m)\right)'Z_{s^c}(m)\right\|^4\left\|(\Sigma_{ZZ}(m))^{-1}\right\|^4(1+o_p(1))
\end{aligned}$$

since $\left(\frac{1}{\kappa}Z'_{s^c}(m)\,Z_{s^c}(m)\right)^{-1} = (\Sigma_{ZZ}(m))^{-1} + o_p(1)$. In addition, as $\|Z_{s^c}(m)\| \leq \|Z(m)\|$ and

$\sum_t \left\|\hat{Z}_t(m) - Z_t(m)\right\|^4 = O_p(1)$ by Lemma A1 of Djogbenou, Gonçalves and Perron (2015),

$$\begin{aligned}\frac{1}{b\kappa^4}\sum_{s\in\mathcal{R}}\left\|\left(\hat{Z}_{s^c}(m) - Z'_{s^c}(m)\right)' Z_{s^c}(m)\right\|^4 &\leq \frac{1}{b\kappa^4}\sum_{s\in\mathcal{R}}\left(\sum_{t\,\in\,s^c}\left\|\left(\hat{Z}_t(m) - Z_t(m)\right) Z_t(m)'\right\|^2\right)^2 \\ &\leq \left(\sum_{t=1}^{T-1}\left\|\hat{Z}_t(m) - Z_t(m)\right\|^4\right)\frac{1}{b\kappa^3}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|Z_t(m)\right\|^4 \\ &= O_p(1)\times O_p\left(\frac{1}{\kappa^2}\right) = O_p\left(\frac{1}{\kappa^2}\right).\end{aligned}$$

Thus, $B_{111} = O_p\left(\frac{1}{\kappa^2}\right)$. Given the fact that $\frac{1}{T}\sum_{t=1}^{T-1}\left\|y_{t+1}\right\|^4 = O_p(1)$ and $\frac{1}{b\cdot\kappa}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\hat{Z}_t(m)\right\|^4 = O_p(1)$ and the bound

$$B_{112} \leq \frac{1}{\kappa^2}\frac{1}{b}\sum_{s\in\mathcal{R}}\left(\frac{1}{\kappa}\sum_{t\,\in\,s^c}\left\|\hat{Z}_t(m)\, y_{t+1}\right\|^2\right)^2 \leq \frac{T}{\kappa^3}\left(\frac{1}{T}\sum_{t=1}^{T-1}\left\|y_{t+1}\right\|^4\right)\frac{1}{b\cdot\kappa}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\hat{Z}_t(m)\right\|^4,$$

it follows that $B_{112} = O_p\left(\frac{T}{\kappa^3}\right)$. In consequence, $B_{11} = O_p\left[\left(\frac{T}{\kappa^5}\right)^{1/2}\right]\cdot O_p\left[\left(\frac{1}{\kappa^2}\right)^{1/2}\right] = O_p\left[\left(\frac{T}{\kappa^2}\frac{1}{\kappa^5}\right)^{1/2}\right]$ and $B_{11} = O_p\left[\frac{1}{\kappa^{3/2}}\right]$ since $\frac{T}{\kappa^2} \to C < \infty$. Let's look at B_{12}. Since $\left\|Z_s(m) - \hat{Z}_s(m)\right\| \leq \left\|Z(m) - \hat{Z}(m)\right\|$ and $\left(\frac{1}{T_c}Z'_{s^c}(m)\, Z_{s^c}(m)\right)^{-1} = \left(\Sigma_{ZZ}(m)\right)^{-1} + o_p(1)$, it follows that

$$\begin{aligned}B_{12} &\leq \frac{1}{d}\left\|Z(m) - \hat{Z}(m)\right\|^2\frac{1}{\kappa^2}\left\|\left(\Sigma_{ZZ}(m)\right)^{-1}\right\|^2\left(1 + o_p(1)\right)\frac{1}{b}\sum_{s\in\mathcal{R}}\left\|\left(\hat{Z}_{s^c}(m) - Z_{s^c}(m)\right)' y_{s^c}\right\|^2 \\ &\leq \frac{1}{\kappa}\frac{1}{d}\left\|Z(m) - \hat{Z}(m)\right\|^2\left\|\left(\Sigma_{ZZ}(m)\right)^{-1}\right\|^2\left(1 + o_p(1)\right) \\ &\quad\times\left[\frac{1}{\kappa\cdot b}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|\hat{Z}_t(m) - Z_t(m)\right\|^4\left(\frac{1}{\kappa\cdot b}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|y_{t+1}\right\|^4\right)\right]^{1/2}.\end{aligned}$$

Because, $\frac{1}{d}\left\|Z(m) - \hat{Z}(m)\right\|^2 = O_p\left(\frac{1}{C_{NT}^2}\right)$, $\frac{\kappa^2}{T} \to c < \infty$ and $\frac{\kappa^2}{N^2} \to 0$

$$\frac{1}{\kappa\cdot b}\sum_{s\in\mathcal{R}}\left\|\hat{Z}_{s^c}(m) - Z_{s^c}(m)\right\|^4 = \left[O_p\left(\frac{\kappa^2}{T}\right) + O_p\left(\frac{\kappa^2}{N^2}\right)\right]O_p\left(\frac{1}{\kappa^2}\right) = O_p\left(\frac{1}{\kappa^2}\right)$$

and $\frac{1}{\kappa\cdot b}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|y_{t+1}\right\|^4 = O_p(1)$, it follows that

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

For B_{13}, we have the bound

$$\begin{aligned} B_{13} &\leq \frac{1}{d\cdot b}\sum_{s\in\mathcal{R}}\|Z_s(m)\|^2\left\|(Z_{s^c}'(m)\,Z_{s^c}(m))^{-1}\right\|^2\left\|\left(Z_{s^c}(m)-\hat{Z}_{s^c}(m)\right)'y_{s^c}\right\|^2 \\ &\leq \frac{1}{\kappa}\left[\frac{d}{\kappa^2}\frac{1}{d\cdot b}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s}\|Z_t(m)\|^4\right]^{1/2}\left\|(\Sigma_{ZZ}(m))^{-1}\right\|^2(1+o_p(1)) \\ &\quad\times\left[\frac{1}{d}\sum_t\|y_{t+1}\|^4\right]^{1/2}\left[\frac{1}{\kappa\cdot b}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\left\|Z_t(m)-\hat{Z}_t(m)\right\|^4\right]^{1/2} \end{aligned}$$

Hence

$$B_{13}=O_p\left(\frac{1}{\kappa}\right)\left[O_p\left(\frac{1}{T^{1/2}}\right)+O_p\left(\frac{1}{N}\right)\right]=O_p\left(\frac{1}{\kappa^2}\right)$$

since $\frac{\kappa^2}{T}\to C<\infty$ and $\frac{\kappa^2}{N^2}\to 0$. To finish, we have that $B_{14}=O_p\left(\frac{1}{\kappa\cdot C_{NT}^2}\right)$, using the bound

$$\begin{aligned} B_{14} &\leq \frac{1}{d\cdot b}\sum_{s\in\mathcal{R}}\left\|Z_s(m)-\hat{Z}_s(m)\right\|^2\left\|(Z_{s^c}'(m)\,Z_{s^c}(m))^{-1}\right\|^2\left\|Z_{s^c}(m)'\,y_{s^c}\right\|^2, \\ &\leq \frac{1}{\kappa}\frac{1}{d}\left\|Z(m)-\hat{Z}(m)\right\|^2\left\|(\Sigma_{ZZ}(m))^{-1}\right\|^2(1+o_p(1)) \\ &\quad\times\left(\frac{1}{b\cdot\kappa}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\|Z_t(m)\|^4\right)^{1/2}\left(\frac{1}{b\cdot\kappa}\sum_{s\in\mathcal{R}}\sum_{t\,\in\,s^c}\|y_{t+1}\|^4\right)^{1/2}, \end{aligned}$$

$B_{14}=O_p\left(\frac{1}{\kappa\cdot C_{NT}^2}\right)\times O_p(1)\times O_p(1)=O_p\left(\frac{1}{\kappa\cdot C_{NT}^2}\right)$.

Finally $B_1=o_p\left(\frac{1}{\kappa}\right)$. By similar arguments, $B_2=o_p\left(\frac{1}{\kappa}\right)$.

Part 2 :

We first show in this part that

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

with $\widetilde{\delta}(m)$ the OLS estimator by regressing y_s on $Z_s(m)$ and $Q_s(m)=Z_s(m)\,(Z'(m)\,Z(m))^{-1}\,Z_s'(m)$. Indeed,

$$\begin{aligned} y_s-P_{s^c}(m)\,y_{s^c} &= (I_d-Q_s(m))^{-1}(I_d-Q_s(m))\,(y_s-P_{s^c}(m)\,y_{s^c}) \\ &= (I_d-Q_s(m))^{-1}\left[y_s-Q_s(m)\,y_s-(I_d-Q_s(m))\,P_{s^c}(m)\,y_{s^c}\right]. \end{aligned}$$

Noting that

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

It follows that

$$
\begin{aligned}
& y_s - Q_s(m)\, y_s - (I_d - Q_s(m))\, P_{s^c}(m)\, y_{s^c} \\
= \;& y_s - Z_s(m)\left(Z'(m)\, Z(m)\right)^{-1} Z_s(m)\, y_s + Z_s(m)\left(Z'(m)\, Z(m)\right)^{-1} Z_{s^c}(m)'\, y_{s^c} \\
= \;& y_s - Z_s(m)\left(Z'(m)\, Z(m)\right)^{-1} Z(m)\, y = y_s - Z_s(m)\, \widetilde{\delta}(m).
\end{aligned}
$$

Thus

$$
y_s - \widetilde{y}_s(m) = \left(I_{T_v} - \widetilde{Q}_s(m)\right)^{-1}\left(y_s - Z_s(m)\, \widetilde{\delta}(m)\right)
$$

and

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

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

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

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

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

Finally,

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

And the result follows. □

Proof of Theorem 13. The proof follows similarly as the one of Theorem 3.1 in Djogbenou, Gonçalves and Perron (2015) by showing that the high level condition of Gonçalves and Perron (2014) are satisfied. Noting that

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

with $A^*(m) = \Phi_0^*(m) \sqrt{\kappa} \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{Z}_t(m)\, \varepsilon^*_{t+1}$, $B^*(m) = \sqrt{\kappa} \frac{1}{T-1} \sum_{t=1}^{T-1} \left(\widetilde{F}^*_t(m) - H_0^*(m)\, \widetilde{F}_t(m)\right) \varepsilon^*_{t+1}$ and $C^*(m) = \sqrt{\kappa} \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{Z}^*_t(m) \left(\widetilde{F}^*_t(m) - H_0^*(m)\, \widetilde{F}_t(m)\right)' \left(H_0^*(m)\right)^{-1\prime} \widehat{\alpha}$ where $\Phi_0^*(m) = p\lim \Phi^*(m)$ and $H_0^*(m) = p\lim H^*(m)$ are diagonal. This decomposition uses the fact that in the bootstrap world Φ_0^* is diagonal (see Gonçalves and Perron, 2014) and $H_0^*(m)$ is a square submatrix of H_0^*. 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. (1) One can write that

$$
B^*(m) = \frac{\sqrt{\kappa}}{T-1} \sum_{t=1}^{T-1} \left(\widetilde{F}^*_t(m) - H_0^*(m)\, \widetilde{F}_t(m)\right) \varepsilon^*_{t+1} = \frac{1}{\sqrt{T-1}} \sum_{t=1}^{T-1} \left(\widetilde{F}^*_t(m) - H_0^*(m)\, \widetilde{F}_t(m)\right) \xi^*_{t+1}
$$

with $\xi^*_{t+1} = \frac{1}{\sqrt{1-\frac{r+q}{T-1}}}\left(\hat{\varepsilon}_{t+1}(M) - \overline{\hat{\varepsilon}(M)}\right)$. Given Lemma B2.of Gonçalves and Perron (2014), $B^*(m) = O_p\left(\frac{1}{C_{NT}}\right)$ as long as $B^* = \frac{1}{\sqrt{T-1}}\sum_{t=1}^{T-1}\left(\widetilde{F}^*_t - H^*_0\widetilde{F}_t\right)\xi^*_{s+1} = O_p\left(\frac{1}{C_{NT}}\right)$ if Conditions $A^* - D^*$ are verified with ξ^*_{t+1} replacing ε^*_{t+1}. Indeed, A^* and B^* are satisfied since e^*_{it} relies on the wild bootstrap and we only need to check Conditions C*, D*. Starting with condition $C^*(a)$, since e^*_{it} and ε^*_{s+1} are independent and e^*_{it} is independent across (i,t),

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

where the first equality uses the fact that $Cov^*\left(e^*_{it}e^*_{is}, e^*_{jt}e^*_{jl}\right) = 0$ for $i \neq j$ or $s \neq l$, and the inequality uses the fact that

$$E^*\left(\varepsilon^{*2}_{s+h}\right) = \frac{1}{1-\frac{r+q}{T-1}}\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\widehat{\varepsilon}^2_{s+1} - \overline{\widehat{\varepsilon}}^2\right) \leq \frac{1}{T-1-r-q}\sum_{s=1}^{T-1}\widehat{\varepsilon}^2_{s+1}.$$

and that $Var^*\left(\eta_{it}\eta_{is}\right)$ is bounded under our assumptions. Since

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{e}^4_{it}/NT = O_P(1) \text{ and } \frac{1}{T-1-r-q}\sum_{t=1}^{T-1}\widehat{\varepsilon}^2_{t+1} = O_P(1)$$

under our assumptions, the result follows. We verify condition $C^*(b)$. We have

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

where the first equality uses the fact that $E^*\left(e^*_{it}e^*_{js}\right) = 0$ whenever $i \neq j$ or $t \neq s$, and the third equality the fact that $E^*\left(\varepsilon^{*2}_{t+1}\right) \leq \frac{1}{T-1-r-q}\sum_{s=1}^{T-1}\widehat{\varepsilon}^2_{s+1}$ and $\left(\frac{1}{T}\sum_{t=1}^{T-1}\widetilde{e}^2_{it}\right)^2 \leq \frac{1}{T}\sum_{t=1}^{T-1}\widetilde{e}^4_{it}$. The result follows since each term of the last inequality is $O_p(1)$ (see GP(2014)). To prove condition $C^*(c)$, I follow closely the proof in GP (2014) and it will be sufficient to show that $\frac{1}{T-1}\sum_{t=1}^{T-1}\xi^{*4}_{t+1} = O_{p^*}(1)$ in probability. Using the definition of $E^*\left(\xi^{*4}_{t+1}\right)$ and the c-r inequality,

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

Because $\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^{4}=O_P\left(1\right)$ and $\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}=O_P\left(1\right)$ under our assumptions, $E^*\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\varepsilon_{t+1}^{*4}\right)=O_{p^*}(1)$ and $C^*\left(c\right)$ follows. For condition $D^*(a)$, we have $E^*\left(\xi_{t+1}^*\right)=\frac{T-1}{T-1-r-q}\frac{1}{T-1}\sum_{s=1}^{T-1}\left(\hat{\varepsilon}_{s+1}-\overline{\hat{\varepsilon}}\right)=0$ and

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

To finish, I also verify condition $D^*(b)$. Noting $\Psi^*\equiv\Omega^{*-\frac{1}{2}}\hat{Z}_t\xi_{t+1}^*$ and

$$\Omega^*=p\lim_{N,T\to\infty}E^*\left(\frac{1}{T}\sum_{t=1}^{T-1}\hat{Z}_t\hat{Z}_t'\xi_{t+1}^{*2}\right),$$

we can write

$$\Omega^{*-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\hat{Z}_t\xi_{t+1}^*=\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\Psi_t^*$$

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

$$p\lim_{N,T\to\infty}Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Psi_t^*\right)=\Omega^{*-\frac{1}{2}}\left[p\lim_{N,T\to\infty}E^*\left(\frac{1}{T}\sum_{t=1}^{T-1}\hat{Z}_t\hat{Z}_t'\xi_{t+1}^{*2}\right)\right]\Omega^{*-\frac{1}{2}}=I.$$

It remains to show that for some $1<s<2$, $\Upsilon_T\equiv\frac{1}{T^d}\sum_{t=1}^{T-1}E^*\left\|\Psi_t^*\right\|^{2s}=o_p(1)$. We have

$$\begin{aligned}\Upsilon_T &= \frac{1}{T^s}\sum_{t=1}^{T-1}E^*\left(\left\|\Omega^{*-\frac{1}{2}}\hat{Z}_t\xi_{t+1}^*\right\|^2\right)^s\leq\frac{1}{T^{s-1}}\frac{1}{T}\sum_{t=1}^{T-1}\left(E^*\left\|\Omega^{*-\frac{1}{2}}\hat{Z}_t\xi_{t+1}^*\right\|^2\right)^s\\ &\leq \frac{1}{T^{s-1}}\left(\frac{1}{T}\sum_{t=1}^{T-1}E^*\left\|\Omega^{*-\frac{1}{2}}\hat{Z}_t\xi_{t+1}^*\right\|^2\right)^s \text{ as } s<2.\end{aligned}$$

which induces that $\Upsilon_T=o_p(1)$ as

$$\Upsilon_T\leq\frac{1}{T^{s-1}}\left(trace\left[\Omega^{*-\frac{1}{2}}\frac{1}{T}\sum_{t=1}^{T-1}E^*\left(\hat{Z}_t\hat{Z}_t'\xi_{t+1}^{*2}\right)\Omega^{*-\frac{1}{2}}\right]\right)^s=\frac{r+q}{T^{s-1}}\to0 \text{ since as } s>1.$$

Thus $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\hat{Z}_t\xi_{t+1}^*\to_{d^*}N\left(0,\ \Omega^*\right)$. (2) By Lemma B4 of Gonçalves and Perron (2014),

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

as it does not involve the residual bootstrap for the time series dimension. Hence,

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

(3) By similar steps to condition $D^*(b)$, $\Omega^*(m)^{-\frac{1}{2}} A^*(m) \rightarrow^{d^*} N(0,\ I)$ where

$$\Omega^*(m)^{-\frac{1}{2}} A^*(m) = \Omega^*(m)^{-\frac{1}{2}} \Phi_0^*(m) \sqrt{\kappa} \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{Z}_t(m) \varepsilon_{t+1}^* = \Omega^*(m)^{-\frac{1}{2}} \Phi_0^*(m) \frac{1}{\sqrt{T-1}} \sum_{t=1}^{T-1} \hat{Z}_t(m) \xi_{t+1}^*,$$

$$\Omega^*(m) = Var^*\left(\Phi_0^*(m) \frac{1}{\sqrt{T-1}} \sum_{t=1}^{T-1} \hat{Z}_t(m) \xi_{t+1}^*\right)$$

and

$$\Psi_t^*(m) \equiv \Omega^{*-\frac{1}{2}}(m) \hat{Z}_t(m) \xi_{t+1}^*.$$

Finally,

$$A^*(m) \rightarrow_p N\left(0,\ \sigma^2 \Phi_0^*(m) \left(p\lim \frac{1}{T-1} Z(m)' Z(m)\right) \Phi_0^*(m)\right).$$

Indeed,

$$\Omega^*(m) = \Phi_0^*(m) \left[\frac{1}{T-1-(r+q)} \sum_{t=1}^{T-1} \hat{Z}_t(m) \hat{Z}_t(m)' \left(\hat{\varepsilon}_{t+1}(M) - \overline{\hat{\varepsilon}(M)}\right)^2\right] \Phi_0^*(m)'$$

with

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

and $\frac{1}{T-1-(r+q)} \sum_{t=1}^{T-1} \hat{Z}_t(m) \hat{Z}_t(m)' \hat{\varepsilon}_{t+1}^2(M)$ a submatrix of

$$\frac{1}{T-1-(r+q)} \sum_{t=1}^{T-1} \hat{Z}_t(M) \hat{Z}_t(M)' \hat{\varepsilon}_{t+1}(M)^2 \rightarrow_p \sigma^2 p\lim \frac{1}{T-1} Z(M)' Z(M)$$

by Assumption A3. Hence, it follows that $\Omega^*(m) \rightarrow_p \sigma^2 \Phi_0^*(m) \left[p\lim \frac{1}{T-1} Z(m)' Z(m)\right] \Phi_0^*(m)$.

From (1), (2), (3) and the fact that

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

by the asymptotic equivalence Lemma,

$$\begin{aligned} \sqrt{\kappa}\left(\hat{\delta}_d^*(m) - \Phi_0^*(m)^{-1} \hat{\delta}(m)\right) &= \left(\frac{1}{T-1} \hat{Z}^{*\prime}(m) \hat{Z}^*(m)\right)^{-1} \left[A^*(m) + o_{p^*}(1)\right] \\ &\rightarrow_{d^*} N\left(0, \Phi_0^*(m)^{-1} \left[p\lim \frac{1}{T-1} Z(m)' Z(m)\right]^{-1} \Phi_0^*(m)^{-1\prime}\right). \end{aligned}$$

□

Proof of Theorem 14. In the first part of our proof, I show that if it exists a matrix $A(m)$ such that $F_t^0 = A(m) F_t(m)$ and no matrix $A(m')$ such that $F_t^0 = A(m') F_t(m')$, then

$$P\left(\hat{\Gamma}_\kappa(m) < \hat{\Gamma}_\kappa(m')\right) \to 1 \text{ as } T,\ N \to \infty.$$

We have that for any m,

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

where

$$\begin{aligned}
D_1 &= \frac{1}{T-1}\left\|y - \hat{Z}(m)\hat{\delta}(m)\right\|^2 \qquad (9)\\
D_2 &= E^*\left(\frac{1}{T-1}\left\|\hat{Z}(m)\hat{\delta}(m) - \hat{Z}^*(m)\hat{\delta}^*_\kappa(m)\right\|^2\right)
\end{aligned}$$

and

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

Using the decomposition

$$\begin{aligned}
\hat{Z}^*(m)\hat{\delta}^*_\kappa(m) - \hat{Z}(m)\hat{\delta}(m) &= \hat{Z}^*(m)\left(\hat{\delta}^*_\kappa(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right) \\
&\quad + \left(\hat{Z}^*(m) - \hat{Z}(m)\Phi_0^*(m)\right)\Phi_0^{*-1\prime}(m)\hat{\delta}(m)
\end{aligned}$$

where $\Phi_0^*(m)$ is a $(r(m)+q) \times (r(m)+q)$ submatrix of $\Phi_0^* = p^* \lim \Phi^* = diag(\pm 1)$ (see Gonçalves and Perron, 2014),

$$D_2 = D_{21} + D_{22} + 2D_{23}$$

where

$$\begin{aligned}
D_{21} &= E^*\left[\left(\hat{\delta}^*_\kappa(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right)' \frac{1}{T-1}\hat{Z}^{*\prime}(m)\hat{Z}^*(m)\left(\hat{\delta}^*_\kappa(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right)\right] \\
D_{22} &= E^*\left[\hat{\delta}'(m)\Phi_0^{*-1}(m)\frac{1}{T-1}\left(\hat{Z}^*(m) - \hat{Z}(m)\Phi_0^*(m)\right)'\left(\hat{Z}^*(m) - \hat{Z}(m)\Phi_0^*(m)\right)\Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right] \\
D_{23} &= \frac{1}{T-1}E^*\left[\hat{\delta}'(m)\Phi_0^{*-1}(m)\left(\hat{Z}^*(m) - \hat{Z}(m)\Phi_0^*(m)\right)'\hat{Z}(m)\left(\hat{\delta}^*_\kappa(m) - \Phi_0^{*-1\prime}(m)\hat{\delta}(m)\right)\right].
\end{aligned}$$

Starting with D_{21}, we have that

$$\begin{aligned}
& D_{21} \\
= \ & E^{*}\left[\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*-1\prime}(m)\,\hat{\delta}(m)\right)^{\prime}\left[\Phi_{0}^{*}(m)\,\frac{1}{T-1}\hat{Z}^{\prime}(m)\,\hat{Z}(m)\,\Phi_{0}^{*\prime}(m)\right]\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*-1\prime}(m)\,\hat{\delta}(m)\right)\right] \\
= \ & trace\left[\left(\Phi_{0}^{*}(m)\,\hat{Z}^{\prime}(m)\,\hat{Z}(m)\,\Phi_{0}^{*\prime}(m)\right)E^{*}\left[\kappa\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*-1\prime}(m)\,\hat{\delta}(m)\right)\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*-1\prime}(m)\,\hat{\delta}(m)\right)^{\prime}\right]\right] \\
& \times\left[1+o_{p^{*}}(1)\right]\times\frac{1}{T-1}\frac{1}{\kappa} \\
= \ & \frac{1}{\kappa}trace\left\{\left(\Phi_{0}^{*}(m)\,\Sigma_{ZZ}(m)\,\Phi_{0}^{*\prime}(m)\right)Avar^{*}\left[\sqrt{\kappa}\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*-1\prime}(m)\,\hat{\delta}(m)\right)\right]\right\}+o_{p^{*}}\left(\frac{1}{\kappa}\right).
\end{aligned}$$

By Gonçalves and Perron (2014), it follows that

$$\sqrt{\kappa}\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*-1\prime}(m)\,\hat{\delta}(m)\right)\to^{d}N\left(0,\ \sigma^{2}\Phi_{0}^{*-1\prime}(m)\,\Sigma_{ZZ}(m)^{-1}\,\Phi_{0}^{*-1}(m)\right)$$

as $\frac{\sqrt{\kappa}}{N}\to 0$. Hence,

$$\begin{aligned}
& p\lim Avar^{*}\left[\sqrt{\kappa}\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*-1\prime}(m)\,\hat{\delta}(m)\right)\right] \\
= \ & p\lim E^{*}\left[\kappa\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*-1\prime}(m)\,\hat{\delta}(m)\right)\left(\hat{\delta}_{\kappa}^{*}(m)-\Phi_{0}^{*-1\prime}(m)\,\hat{\delta}(m)\right)^{\prime}\right] \\
= \ & \sigma^{2}\Phi_{0}^{*-1\prime}(m)\,\Sigma_{ZZ}(m)^{-1}\,\Phi_{0}^{*-1}(m).
\end{aligned}$$

Thereby,

$$\begin{aligned}
D_{21} &= \sigma^{2}\frac{1}{\kappa}trace\left\{\left(\Phi_{0}^{*}(m)\,\Sigma_{ZZ}(m)\,\Phi_{0}^{*\prime}(m)\right)\Phi_{0}^{*-1\prime}(m)\,\Sigma_{ZZ}(m)^{-1}\,\Phi_{0}^{*-1}(m)\right\}+o_{p}\left(\frac{1}{\kappa}\right) \\
&= \sigma^{2}\frac{1}{\kappa}trace\left\{\Sigma_{ZZ}(m)\,\Sigma_{ZZ}(m)^{-1}\right\}+o_{p}\left(\frac{1}{\kappa}\right).
\end{aligned}$$

For a correct model m,

$$D_{21}=\frac{\sigma^{2}\left(r(m)+q\right)}{\kappa}+o_{p}\left(\frac{1}{\kappa}\right).$$

For D_{22}, we have that

$$D_{22}=\hat{\delta}(m)^{\prime}\,\Phi_{0}^{*-1}(m)\,E^{*}\left[\frac{1}{T-1}\left(\hat{Z}^{*}(m)-\hat{Z}(m)\,\Phi_{0}^{*}(m)\right)^{\prime}\left(\hat{Z}^{*}(m)-\hat{Z}(m)\,\Phi_{0}^{*}(m)\right)\right]\Phi_{0}^{*-1\prime}(m)\,\hat{\delta}(m).$$

We also have that

$$\begin{aligned}
& E^{*}\left[\frac{1}{T-1}\left(\hat{Z}^{*}(m)-\hat{Z}(m)\,\Phi_{0}^{*}(m)\right)^{\prime}\left(\hat{Z}^{*}(m)-\hat{Z}(m)\,\Phi_{0}^{*}(m)\right)\right] \\
= \ & \frac{1}{T-1}\sum_{t=1}^{T-1}E^{*}\left(\widetilde{F}_{t}^{*}(m)-H_{0}^{*}(m)\,F_{t}(m)\right)\left(\widetilde{F}_{t}^{*}(m)-H_{0}^{*}(m)\,F_{t}^{\prime}(m)\right)^{\prime}
\end{aligned}$$

which is a subset of $D_{221}=\frac{1}{T-1}\sum_{t=1}^{T-1}E^{*}\left(\widetilde{F}_{t}^{*}-H_{0}^{*}\widetilde{F}_{t}\right)\left(\widetilde{F}_{t}^{*}-H_{0}^{*}\widetilde{F}_{t}\right)'$. Because, it can be shown that D_{221} is of the same order with

$$D_{222}=\frac{1}{T}\sum_{t=1}^{T}E^{*}\left(\widetilde{F}_{t}^{*}-H^{*}\widetilde{F}_{t}\right)\left(\widetilde{F}_{t}^{*}-H^{*}\widetilde{F}_{t}\right)'$$

it suffices to find the order in probability of D_{222}. Following the step of the proof of Lemma 3.1 in Gonçalves and Perron (2014), we have that,

$$\left\|D_{222}\right\|\leq\frac{1}{T}\sum_{t=1}^{T}E^{*}\left\|\widetilde{F}_{t}^{*}-H^{*}\widetilde{F}_{t}\right\|^{2}\leq\frac{4}{T}\sum_{t=1}^{T}\left(E^{*}\left\|A_{1t}^{*}\right\|^{2}+E^{*}\left\|A_{2t}^{*}\right\|^{2}+E^{*}\left\|A_{3t}^{*}\right\|^{2}+E^{*}\left\|A_{4t}^{*}\right\|^{2}\right)$$

given that $\left\|V^{*-1}\right\|$ is bounded and ignored, where

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

with $\gamma_{st}^{*}=E^{*}\left(\frac{1}{N}\sum_{i=1}^{N}e_{is}^{*}e_{it}^{*}\right)$, $\zeta_{st}^{*}=\frac{1}{N}\sum_{i=1}^{N}\left(e_{is}^{*}e_{it}^{*}-E^{*}\left(\frac{1}{N}\sum_{i=1}^{N}e_{is}^{*}e_{it}^{*}\right)\right)$, $\eta_{st}^{*}=\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_{i}'\widetilde{F}_{s}e_{it}^{*}$, and $\xi_{st}^{*}=\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_{i}'\widetilde{F}_{t}e_{is}^{*}$.

Firstly,

$$\frac{1}{T}\sum_{t=1}^{T}E^{*}\left\|A_{1t}^{*}\right\|^{2}\leq\frac{1}{T}E^{*}\left[\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{F}_{s}^{*}\right\|^{2}\right)\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|\gamma_{st}^{*}\right\|^{2}\right)\right]=r\frac{1}{T}\cdot\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|\gamma_{st}^{*}\right\|^{2}\right)$$

since

$$\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{F}_{s}^{*}\right\|^{2}=trace\left(\frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_{s}^{*}\widetilde{F}_{s}^{*\prime}\right)=trace\left(I_{r}\right)=r.$$

Because the high level condition $A^{*}\left(b\right)$ of Gonçalves and Perron (2014) is satisfied $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|\gamma_{st}^{*}\right\|^{2}=O_{p}\left(1\right)$. In consequence, $\frac{1}{T}\sum_{t=1}^{T}E^{*}\left\|A_{1t}^{*}\right\|^{2}=O_{p}\left(\frac{1}{T}\right)$. Secondly,

$$\frac{1}{T}\sum_{t=1}^{T}E^{*}\left\|A_{2t}^{*}\right\|^{2}\leq E^{*}\left[\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{F}_{s}^{*}\right\|^{2}\right)\left(\frac{1}{T^{2}}\sum_{s=1}^{T}\sum_{t=1}^{T}\zeta_{st}^{*2}\right)\right]\leq r\cdot E^{*}\left(\frac{1}{T^{2}}\sum_{s=1}^{T}\sum_{t=1}^{T}\zeta_{st}^{*2}\right)=O_{p}\left(\frac{1}{N}\right)$$

by condition $A^{*}\left(c\right)$ of GP (2014). Thirdly, using the same argument as Gonçalves and Perron (2014),

$$\frac{1}{T}\sum_{t=1}^{T}E^{*}\left\|A_{3t}^{*}\right\|^{2}\leq E^{*}\left[\left\|\frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_{s}^{*}\widetilde{F}_{s}'\right\|^{2}\frac{1}{T^{2}}\sum_{s=1}^{T}\left\|\frac{1}{N}\widetilde{\Lambda}'e_{t}^{*}\right\|^{2}\right]\leq r\cdot E^{*}\left[\left(\frac{1}{T^{2}}\sum_{s=1}^{T}\left\|\frac{1}{N}\widetilde{\Lambda}'e_{t}^{*}\right\|^{2}\right)\right]=O_{p}\left(\frac{1}{N}\right).$$

Because similarly, $\frac{1}{T}\sum_{t=1}^{T}E^{*}\left\|A_{4t}^{*}\right\|^{2}=O_{p}\left(\frac{1}{N}\right)$, we can deduce that, $\left\|D_{221}\right\|=O_{p}\left(C_{NT}^{-2}\right)$. It follows that $D_{22}=O_{p}\left(C_{NT}^{-2}\right)$. We have that $\left\|D_{23}\right\|\leq\left\|D_{21,NT}\right\|^{\frac{1}{2}}\left\|D_{22,NT}\right\|^{\frac{1}{2}}=O_{p}\left(\frac{1}{\sqrt{T_{c}}}\right)O_{p}\left(C_{NT}^{-1}\right)=O_{p}\left(\frac{1}{\sqrt{T_{c}}\cdot C_{NT}}\right)$. Finally, $D_{2}=\frac{\sigma^{2}(r_{s}+q)}{T_{c}}+O_{p}\left(\frac{1}{\sqrt{T_{c}}\cdot C_{NT}}\right)$. We also have that

$$D_{3}=-2\frac{1}{T-1}\left(y-\hat{Z}\left(m\right)\hat{\delta}\left(m\right)\right)'\hat{Z}\left(m\right)\Phi_{0}^{*\prime}\left(m\right)E^{*}\left(\hat{\delta}_{d}^{*}\left(m\right)\left(1+o_{p^{*}}\left(1\right)\right)\right)=o_{p}\left(\frac{1}{T-1}\right)$$

as $\left(y-\hat{Z}(m)\hat{\delta}(m)\right)'\hat{Z}(m)=0$.Let's look at D_1. Denoting

$$\widetilde{M}(m)=I_{T-1}-\widetilde{P}(m) \text{ and } M(m)=I_{T-1}-P(m),$$

we have that

$$\begin{aligned}
D_1 &= \frac{1}{T-1}y'\widetilde{M}(m)y=\frac{1}{T-1}y'M(m)y+O_p\left(\frac{1}{C_{NT}^2}\right)\\
&= \frac{1}{T-1}\varepsilon'M(m)\varepsilon+\frac{1}{T-1}\delta'Z^{0\prime}M(m)Z^0\delta+2\frac{1}{T-1}\delta'Z^{0\prime}M(m)\varepsilon+O_p\left(\frac{1}{C_{NT}^2}\right)\\
&= \frac{1}{T-1}\varepsilon'\varepsilon+\frac{1}{T-1}\delta'Z^{0\prime}M(m)Z^0\delta+O_p\left(\frac{1}{C_{NT}^2}\right).
\end{aligned}$$

Indeed, $\frac{1}{T-1}y'\widetilde{M}(m)y=\frac{1}{T-1}y'M(m)y+O_p\left(\frac{1}{C_{NT}^2}\right)$, $\frac{1}{T-1}\varepsilon'M(m)\varepsilon=\frac{1}{T-1}\varepsilon'\varepsilon+O_p\left(\frac{1}{C_{NT}^2}\right)$. Hence

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

Given the assumptions that there exists matrix $A(m)$ such that $F_t^0=A(m)F_t(m)$ and no matrix $A(m')$ such that $F_t^0=A(m')F_t(m')$, it follows that $M(m)Z^0=0$ and $M(m')Z^0\neq 0$. Therefore, $\frac{1}{T-1}\delta'Z^{0\prime}M(m)Z^0\delta=\frac{1}{T-1}\delta'Z^{0\prime}M(m)\varepsilon=0$,

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

and

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

as $\frac{1}{T-1}\delta'Z^{0\prime}M(m')\varepsilon=o_p(1)$. Since $p\lim\inf_{N,T\to\infty}\frac{1}{T-1}\delta'Z^{0\prime}M(m')Z^0\delta>0$ if $M(m')Z^0\neq 0$ given $A_4(a)$, we have that

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

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

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

Hence,

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

□

.4 Appendix for next quarter excess returns prediction

Table 2 : Variation explained by estimated macro in X_1 and financial factors in X_2

	Macro factors $\left(\widetilde{F}\right)$		Financial factors $\left(\widetilde{G}\right)$	
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 3 : Estimation results for $m_{t+1} = \alpha_1' (m) \widetilde{F}_t (m) + \alpha_2' (m) \widetilde{G}_t (m) + \beta Z_t + u_{t+1} (m)$

Regressors	CV_1	BICM	CV_d	$\hat{\Gamma}_\kappa$
$constant$	10.90★★	6.53	11.49★★	10.31★★
$(t-stat)$	(2.65)	(1.48)	(3.15)	(2.30)
CAY_t	20.37★	27.97★★	22.71★	22.62★★
$(t-stat)$	(1.67)	(2.36)	(2.02)	(1.77)
$RREL_t$	0.50★	−0.33★	0.04	−0.16
$(t-stat)$	(1.59)	(−1.75)	(0.19)	(−0.61)
$d-p_t$	1.85★★	1.02	1.89★★	1.76★★
$(t-stat)$	(2.69)	(1.40)	(3.03)	(2.35)
VOL_t	0.15★	0.0475	0.09	0.15★★
$(t-stat)$	(1.83)	(0.46)	(0.83)	(1.99)
$\widetilde{F}_{1t}$	0.71★★			
$(t-stat)$	(2.04)			
$\widetilde{F}_{3t}$	1.34★★		0.98★★	0.97★★
$(t-stat)$	(3.67)		(2.99)	(2.59)
$\widetilde{F}_{4t}$	−0.65★★			
$(t-stat)$	(−2.37)			
$\widetilde{G}_{2t}$	−0.60★★		−0.65★★	−0.64★★
$(t-stat)$	(−2.48)		(−2.81)	(−2.50)
$\widetilde{G}_{3t}$	0.48★★			0.59★★
$(t-stat)$	(2.01)			(2.48)
$\widetilde{G}_{4t}$	0.71★★			0.69★★
$(t-stat)$	(2.40)			(2.35)
$\widetilde{G}_{6t}$	−0.55★★			−0.55★★
$(t-stat)$	(−2.13)			(−1.99)
R^2	0.219	0.048	0.121	0.19
$F-test$	6.37		8.65	7.24
$F-cv$	2.05		3.04	2.26

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

Datas :

The macro data are formed following McCracken and Ng (2015). Four series are dropped to obtain balanced data set 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 St. Louis Federal Reserve database. Since not all the data are available on FRED web site or some have missing value, we complete my data set by aggregating appropriately monthly data in 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, 7 correspond respectively to level, first difference, second difference, log transformation, first difference of the log, second difference of the log and growth rate.

The financial data series are indexed from 1 to 147 corresponding to Jurado, Ludvigson and Ng (2015) database. This data set 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 Jurado, Ludvigson and Ng (JLN) data are monthly, we download quarterly available from Kenneth R. French database and construct the remaining one using similar steps to JLN (2015). The quarterly returns of portfolios are obtained by computing the three month returns from monthly version in Kenneth R. French database. We also applied $Log(1 + x/100)$ times 400 instead of 1200 by JLN to have the corresponding annual version. Except the logged dividend price ratio which corresponds in my database to the end of the correspond quarter in JLN (2015), the variables in group 1 are summed over the quarter from JLN (2015). As in Ludvigson and Ng (2007), the quarterly CP factor of Cochrane and Piazzesi (2005) is its average over the quarter.

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
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	CPI 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	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

Financial data set

Group 1 : Yield and dividends

No.	Code	Description	Tcode
1	D_log(DIV)	Log diff. of sum of the dividends in the last 4 quarters	1
2	D_log(P)	Log diff. of portfolio price when dividends are not reinvested	1
3	D_DIVreinvested	Log diff. of sum of the dividends in the last 4 quarters	1
4	D_Preinvested	Log diff. of 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 : (S, H) minus (S, L) 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 B/M	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

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

No.	Code	Description	Tcode
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