# A DIAGNOSTIC CRITERION
# FOR APPROXIMATE FACTOR STRUCTURE

Patrick Gagliardini[a], Elisa Ossola[b] and Olivier Scaillet[c]*

First draft: February 2014.

**Abstract**

We build a simple diagnostic criterion for approximate factor structure in large cross-sectional equity datasets. Given a model for asset returns with observable factors, the criterion checks whether the error terms are weakly cross-sectionally correlated or share at least one unobservable common factor. It only requires computing the largest eigenvalue of the empirical cross-sectional covariance matrix of the residuals of a large unbalanced panel. The panel data model accommodates both time-invariant and time-varying factor structures. We develop the theory for large cross-section and time-series dimensions. No restriction is imposed on the relation between both dimensions. The empirical analysis runs on returns for about ten thousands US stocks from January 1968 to December 2011. Among several multi-factor models proposed in the literature, we cannot select a model with zero factors in the errors.

*JEL Classification:* C12, C13, C23, C51, C52, C58, G12.

*Keywords:* large panel, approximate factor model, asset pricing, model selection.

[a]University of Lugano and Swiss Finance Institute, [b]University of Lugano, [c]University of Geneva and Swiss Finance Institute.

*Acknowledgements: The second author gratefully acknowledges the Swiss National Science Foundation (SNSF) for generously funding her research with a Marie Heim-Voegtlin fellowship.

# 1 Introduction

Empirical work in asset pricing vastly relies on linear multi-factor models with either time-invariant coefficients (unconditional models) or time-varying coefficients (conditional models). The factor structure is often based on observable variables (empirical factors) and supposed to be rich enough to extract systematic risks while idiosyncratic risk is left over to the error term. Linear factor models are rooted in the Arbitrage Pricing Theory (APT, Ross (1976), Chamberlain and Rothschild (1983)) or come from a loglinearization of nonlinear consumption-based models (Campbell (1993)). Conditional linear factor models aim at capturing the time-varying influence of financial and macroeconomic variables in a simple setting (see e.g. Shanken (1990), Cochrane (1996), Ferson and Schadt (1996), Ferson and Harvey (1991, 1999), Lettau and Ludvigson (2001), Petkova and Zhang (2005)). Time variation in risk biases time-invariant estimates of alphas and betas, and therefore asset pricing test conclusions (Jagannathan and Wang (1996), Lewellen and Nagel (2006), Boguth et al. (2011)). Ghysels (1998) discusses the pros and cons of modeling time-varying betas.

A central and practical issue is to determine whether there are one or more factors omitted in the chosen specification. Approximate factor structures with nondiagonal error covariance matrices (Chamberlain and Rothschild (1983)) answer the potential empirical mismatch of exact factor structures with diagonal error covariance matrices underlying the original APT of Ross (1976). If the set of observable factors is correctly specified, the errors are weakly cross-sectionally correlated. Given the large menu of factors available in the literature (the factor zoo of Cochrane (2011), see also Harvey, Liu and Zhu (2013)), we need a simple diagnostic criterion to decide whether we can feel comfortable with the chosen set of observable factors.

For models with unobservable (latent) factors, Connor and Korajczyk (1993) were the first to develop a test for the number of factors for large balanced panels of individual stock returns in time-invariant models under covariance stationarity and homoskedasticity. Unobservable factors are estimated by the method of asymptotic principal components developed by Connor and Korajczyk (1986) (see also Stock and Watson (2002b)). For heteroskedastic settings, the recent literature on large panels with static factors has extended the toolkit available to researchers. Bai and Ng (2002a) introduce a penalized least-squares strategy to estimate the number of factors, at least one, without restrictions on the relation between the cross-sectional dimension ($n$) and the time-series dimension ($T$). Caner and Han (forthcoming, 2014) propose an estimator with a group bridge penalization to determine the number of unobservable factors. Onatski (2009, 2010) looks at the behavior of the adjacent eigenvalues to determine the number of factors when $n$ and $T$ are comparable. Ahn and Horenstein (2013) opt for the same strategy and cover the possibility of zero factors. Kapetanios (2010) uses subsampling to estimate the limit distribution of the adjacent eigenvalues. In the spirit of Lehmann and Modest (1988) and Connor and Korajczyk (1988), Bai and Ng (2006) analyze statistics to test whether the observable factors in time-invariant models span the space of unobservable factors. They do not impose any restriction on $n$ and $T$. They find that the three factor model of Fama and French (1993, FF) is the most satisfactory proxy for the unobservable factors estimated from balanced panels of portfolio and individual stock returns. Ahn, Horenstein and Wang (2013) study a rank estimation method to also check whether time-invariant factor models are compatible with a number of unobservable factors. For portfolio returns, they find that the FF model exhibits a full rank beta (factor loading) matrix.

In this paper, we build a simple diagnostic criterion for approximate factor structure in large cross-sectional datasets. The criterion checks whether the error terms in a given model with observable factors are weakly cross-sectionally correlated or share at least one common factor. It only requires computing the largest eigenvalue of the empirical cross-sectional covariance matrix of the residuals of a large unbalanced panel and subtracting a penalization term vanishing to zero for large $n$ and $T$. The steps of the diagnostic are easy: 1) compute the largest eigenvalue, 2) subtract a penalty, 3) conclude to validity of the proposed approximate factor structure if the difference is negative, or conclude to at least one omitted factor if the difference is positive. Our theoretical contribution shows that step 3) yields asymptotically the correct model selection. We also propose a general version of the diagnostic criterion that determine the number of omitted common factors. We derive all properties for unbalanced panels in the setting of Connor and Korajczyk (1987) to avoid the survivorship bias inherent to studies restricted to balanced subsets of available stock return databases (Brown, Goetzmann, and Ross (1995)). The panel data model is sufficiently general to accommodate both time-invariant and time-varying factor structures (Gagliardini, Ossola, and Scaillet (2011), GOS). We develop the theory for large cross-section and time-series dimensions. No restriction is imposed on the relation between both dimensions. As shown below, the criterion is related to the penalized least-squares approach of Bai and Ng (2002a) for model selection with unobservable factors.

For our empirical contribution, we consider the Center for Research in Security Prices (CRSP) database and take the Compustat database to match firm characteristics. The merged dataset comprises about ten thousands stocks with monthly returns from January 1968 to December 2011. We look at fifteen empirical factors and we build thirteen factor models popular in the empirical finance literature to explain monthly equity returns. They differ by the choice of the observable factors. We analyze monthly returns using the three factors of FF; the five factors of Chen, Roll and Ross (1986, CRR); the three factor of (Jagannathan and Wang, 1996, JW); the three liquidity related factors of Pastor and Stambaugh (2002, LIQ), plus the momentum (MOM) factor and the two return reversal (REV) factors (short-term and long-term). We study time-invariant and time-varying versions of the factor models (Shanken (1990), Cochrane (1996), Ferson and Schadt (1996), Ferson and Harvey (1999)). For the latter, we use both macrovariables and firm characteristics as instruments (Avramov and Chordia (2006)). Among the time-invariant multi-factor models, we cannot select a model with zero factors in the errors. However, we select an approximate factor structure in the errors for some time-varying specifications.

The outline of the paper is as follows. In Section 2, we consider a general framework of conditional linear factor model for asset returns. In Section 3, we present our diagnostic criterion for approximate factor structure. In Section 4, we provide the diagnostic criterion to determine the number of omitted factors. Section 5 contains the empirical results. In the Appendices 1 and 2, we gather the theoretical assumptions and some proofs. We use high-level assumptions to get our results. In Appendix 3, we provide the link of our approach to the expectation-maximization (EM) algorithm proposed by Stock and Watson (2002b). Appendix 4 includes the Monte Carlo simulation results. We place all omitted proofs in the online supplementary materials. There, we also include some additional empirical results and robustness checks.

## 2 Conditional factor model of asset returns

In this section, we consider a conditional linear factor model with time-varying coefficients. We work in a multi-period economy (Hansen and Richard (1987)) under an approximate factor structure (Chamberlain and Rothschild (1983)) with a continuum of assets as in GOS. Such a construction is close to the setting advocated by Al-Najjar (1995, 1998, 1999a) in a static framework with an exact factor structure. He discusses several key advantages of using a continuum economy in arbitrage pricing and risk decomposition. A key advantage is robustness of factor structures to asset repackaging (Al-Najjar (1999b); see GOS for a proof).

Let $\mathcal{F}_t$, with $t = 1, 2, ...$, be the information available to investors. Without loss of generality, the continuum of assets is represented by the interval $[0,1]$. The excess returns $R_t(\gamma)$ of asset $\gamma \in [0,1]$ at dates $t = 1, 2, ...$ satisfy the conditional linear factor model:

$$R_t(\gamma) = a_t(\gamma) + b_t(\gamma)' f_t + \varepsilon_t(\gamma), \tag{1}$$

where vector $f_t$ gathers the values of $K$ observable factors at date $t$. The intercept $a_t(\gamma)$ and factor sensitivities $b_t(\gamma)$ are $\mathcal{F}_{t-1}$-measurable. The error terms $\varepsilon_t(\gamma)$ have mean zero and are uncorrelated with the factors conditionally on information $\mathcal{F}_{t-1}$. Moreover, we exclude asymptotic arbitrage opportunities in the economy: there are no portfolios that approximate arbitrage opportunities when the number of assets increases. In this setting, GOS show that the following asset pricing restriction holds:

$$a_t(\gamma) = b_t(\gamma)'\nu_t, \text{ for almost all } \gamma \in [0,1], \tag{2}$$

almost surely in probability, where random vector $\nu_t \in \mathbb{R}^K$ is unique and is $\mathcal{F}_{t-1}$-measurable. The asset pricing restriction (2) is equivalent to $E[R_t(\gamma)|\mathcal{F}_{t-1}] = b_t(\gamma)'\lambda_t$, where $\lambda_t = \nu_t + E[f_t|\mathcal{F}_{t-1}]$ is the vector of the conditional risk premia.

To have a workable version of Equations (1) and (2), we define how the conditioning information is generated and how the model coefficients depend on it via simple functional specifications. The conditioning information $\mathcal{F}_{t-1}$ contains $Z_{\underline{t-1}}$ and $Z_{\underline{t-1}}(\gamma)$, for all $\gamma \in [0,1]$, where the vector of lagged instruments $Z_{t-1} \in \mathbb{R}^p$ is common to all stocks, the vector of lagged instruments $Z_{t-1}(\gamma) \in \mathbb{R}^q$ is specific to stock $\gamma$, and $Z_{\underline{t}} = \{Z_t, Z_{t-1}, ...\}$. Vector $Z_{t-1}$ may include the constant and past observations of the factors and some additional variables such as macroeconomic variables. Vector $Z_{t-1}(\gamma)$ may include past observations of firm characteristics and stock returns. To end up with a linear regression model, we assume that: (i) the vector of factor loadings $b_t(\gamma)$ is a linear function of lagged instruments $Z_{t-1}$ (Shanken (1990), Ferson and Harvey (1991)) and $Z_{t-1}(\gamma)$ (Avramov and Chordia (2006)); (ii) the vector of risk premia $\lambda_t$ is a linear function of lagged instruments $Z_{t-1}$ (Cochrane (1996), Jagannathan and Wang (1996)); (iii) the conditional expectation of $f_t$ given the information $\mathcal{F}_{t-1}$ depends on $Z_{t-1}$ only and is linear (as e.g. if $Z_t$ follows a Vector Autoregressive (VAR) model of order 1).

To ensure that cross-sectional limits exist and are invariant to reordering of the assets, we introduce a sampling scheme as in GOS. We formalize it so that observable assets are random draws from an underlying population (Andrews (2005)). In particular, we rely on a sample of $n$ assets by randomly drawing i.i.d.

indices $\gamma_i$ from the population according to a probability distribution $G$ on $[0, 1]$. For any $n, T \in \mathbb{N}$, the excess returns are $R_{i,t} = R_t(\gamma_i)$. Similarly, let $a_{i,t} = a_t(\gamma_i)$ and $b_{i,t} = b_t(\gamma_i)$ be the characteristics, and $\varepsilon_{i,t} = \varepsilon_t(\gamma_i)$ be the error terms. By random sampling, we get a random coefficient panel model (e.g. Hsiao (2003), Chapter 6). In available datasets, we do not observe asset returns for all firms at all dates. Thus, we account for the unbalanced nature of the panel through a collection of indicator variables $I_{i,t}$, for any asset $i$ at time $t$. We define $I_{i,t} = 1$ if the return of asset $i$ is observable at date $t$, and 0 otherwise (Connor and Korajczyk (1987)).

Through appropriate redefinitions of the regressors and coefficients, GOS show that we can rewrite the model for Equations (1) and (2) as follows:

$$R_{i,t} = x_{i,t}'\beta_i + \varepsilon_{i,t}, \tag{3}$$

where the regressor $x_{i,t} = \left(x_{1,i,t}', x_{2,i,t}'\right)'$ has dimension $d = d_1 + d_2$ and includes vectors $x_{1,i,t} = \left(vech\,[X_t]', Z_{t-1}' \otimes Z_{i,t-1}'\right)' \in \mathbb{R}^{d_1}$ and $x_{2,i,t} = \left(f_t' \otimes Z_{t-1}', f_t' \otimes Z_{i,t-1}'\right)' \in \mathbb{R}^{d_2}$ with $d_1 = p(p+1)/2 + pq$ and $d_2 = K(p+q)$. The symmetric matrix $X_t = [X_{t,k,l}] \in \mathbb{R}^{p \times p}$ is such that $X_{t,k,l} = Z_{t-1,k}^2$, if $k = l$, and $X_{t,k,l} = 2Z_{t-1,k}Z_{t-1,l}$, otherwise, $k, l = 1, \ldots, p$. The vector-half operator $vech\,[\cdot]$ stacks the elements of the lower triangular part of a $p \times p$ matrix as a $p\,(p+1)\,/2 \times 1$ vector (see Chapter 2 in Magnus and Neudecker (2007) for properties of this matrix tool). In matrix notation, for any asset $i$, we have

$$R_i = X_i\beta_i + \varepsilon_i, \tag{4}$$

where $R_i$ and $\varepsilon_i$ are $T \times 1$ vectors. Regression (3) contains both explanatory variables that are common across assets (scaled factors) and asset-specific regressors. It includes models with time-invariant coefficients as a particular case. In such a case, the regressor reduces to $x_t = (1, f_t')'$ and is common across assets.

In order to build the diagnostic criterion for the set of observable factors, we consider the following rival models:

$\mathcal{M}_1$ : the linear regression model (3), where the errors $(\varepsilon_{i,t})$ follow an approximate factor structure,

and

$\mathcal{M}_2$ : the linear regression model (3), where the errors $(\varepsilon_{i,t})$ satisfy a factor structure.

Thus, the error terms $\varepsilon_{i,t}$ are weakly cross-sectionally dependent under model $\mathcal{M}_1$. On the other hand, under model $\mathcal{M}_2$, the following error factor structure holds

$$\varepsilon_{i,t} = \theta_i' h_t + u_{i,t}, \tag{5}$$

where the $m \times 1$ vector $h_t$ includes unobservable (i.e., latent or hidden) factors. The $m \times 1$ vector $\theta_i$ corresponds to the factor loadings, and the number $m$ of common factors is assumed unknown. In vector notation, we have:

$$\varepsilon_i = H\theta_i + u_i, \tag{6}$$

where $H$ is the $T \times m$ matrix of unobservable factor values, and $u_i$ is a $T \times 1$ vector.

Model $\mathcal{M}_2$ can be distinguished from model $\mathcal{M}_1$ only if the systematic component $H\theta_i$ in the error vector $\varepsilon_i$ is not spanned by the columns of matrix $X_i$ for most assets. Otherwise, the common component $H\theta_i$ can be absorbed in the observable regressors, and we face an identification issue. Therefore, we introduce the next assumption under model $\mathcal{M}_2$.

**Assumption 1** *Under model $\mathcal{M}_2$, the factor structure in the error term is such that*

$$\mu_1 \left( \frac{1}{nT} \sum_i M_{\tilde{X}_i} \tilde{H}_i \theta_i \theta_i' \tilde{H}_i' M_{\tilde{X}_i} \right) \geq c, \text{ with probability approaching } 1,$$

*for a constant $c > 0$, where $M_{\tilde{X}_i} = I_T - \tilde{X}_i \left( \tilde{X}_i' \tilde{X}_i \right)^{-1} \tilde{X}_i'$ for any $i$, $I_T$ denotes the identity matrix of dimension $T$, $\tilde{X}_i = \boldsymbol{I}_i \odot X_i$, $\tilde{H}_i = \boldsymbol{I}_i \odot H$, with $\boldsymbol{I}_i$ the $T \times 1$ vector of observability indicators of asset $i$, and $\mu_1(\cdot)$ and $\odot$ denote the largest eigenvalue of a symmetric matrix and the Hadamard product between matrices, respectively.*

In Assumption Assumption 1, vector $\hat{\eta}_i = M_{\tilde{X}_i} \tilde{H}_i \theta_i$ is the residual vector in the regression of $\tilde{H}_i \theta_i$ on the explanatory variables $\tilde{X}_i$. Thus, $\hat{\eta}_i$ is the part of the systematic component of the error vector, which is not spanned by the observable regressors. Assumption Assumption 1 requires that the largest eigenvalue of the cross-sectional second-order moments matrix of vectors $\hat{\eta}_i$, standardized by their dimension $T$, does not vanish asymptotically. By the literature on unobservable factor models (e.g., Bai and Ng (2002a)) this condition amounts to the presence of some common factors in the $\hat{\eta}_i$. For a factor model with time invariant coefficients and a balanced panel, we have $\tilde{X}_i = X_i = X$, and $\tilde{H}_i = H$ (see Appendix A.2.1):

$$\mu_1 \left( \frac{1}{nT} \sum_i M_X H \theta_i \theta_i' H' M_X \right) \geq \mu_m \left( \frac{\Theta'\Theta}{n} \right) \mu_1 \left( \frac{H' M_X H}{T} \right), \tag{7}$$

where $\Theta$ is the $n \times m$ matrix of factor loadings and $\mu_m(\cdot)$ is the smallest eigenvalue of a $m \times m$ symmetric matrix. Without loss of generality, the unobservable factors can be selected orthogonal to the observable regressors. Thus, Assumption Assumption 1 is satisfied, if matrices $\frac{\Theta'\Theta}{n}$ and $\frac{H'H}{T}$, i.e., the second-moments matrices of the loadings and of the factors converge to positive definite matrices (see Assumptions A and B in Bai and Ng (2002a)).

## 3 Diagnostic criterion

In this section, we provide the diagnostic criterion that checks whether the error terms are weakly cross-sectionally correlated or share at least one common factor. To compute the criterion, we estimate model (3) by ordinary least square (OLS) asset by asset, and we get estimators $\hat{\beta}_i = \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_t I_{i,t} x_{i,t} R_{i,t}$, for $i = 1, ..., n$, where $\hat{Q}_{x,i} = \frac{1}{T_i} \sum_t I_{i,t} x_{i,t} x_{i,t}'$. We get the residuals $\hat{\varepsilon}_{i,t} = R_{i,t} - x_{i,t}' \hat{\beta}_i$, where $\hat{\varepsilon}_{i,t}$ is observable only if $I_{i,t} = 1$. In available panels, the random sample size $T_i$ for asset $i$ can be small, and the inversion of matrix $\hat{Q}_{x,i}$ can be numerically unstable. To avoid unreliable estimates of $\beta_i$, we apply a trimming approach as in GOS. We define $\mathbf{1}_i^{\chi} = \mathbf{1} \left\{ CN \left( \hat{Q}_{x,i} \right) \leq \chi_{1,T}, \tau_{i,T} \leq \chi_{2,T} \right\}$, where $CN \left( \hat{Q}_{x,i} \right) = \sqrt{\mu_1 \left( \hat{Q}_{x,i} \right) / \mu_d \left( \hat{Q}_{x,i} \right)}$ is the condition number of matrix $\hat{Q}_{x,i}$, and $\tau_{i,T} = T/T_i$. The two sequences $\chi_{1,T} > 0$ and $\chi_{2,T} > 0$ diverge asymptotically. The first trimming condition $\{CN \left( \hat{Q}_{x,i} \right) \leq \chi_{1,T}\}$ keeps in the cross-section only assets for which the time series regression is not too badly conditioned. A too large value of $CN \left( \hat{Q}_{x,i} \right)$ indicates multicollinearity problems and ill-conditioning (Belsley et al. (2004), Greene (2008)). The second trimming condition $\{\tau_{i,T} \leq \chi_{2,T}\}$ keeps in the cross-section only assets for which the time series is not too short. We also use both trimming conditions in the proofs of the asymptotic results.

We consider the following diagnostic criterion:

$$\xi = \mu_1 \left( \frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right) - g(n, T), \tag{8}$$

where the vector $\bar{\varepsilon}_i$ of dimension $T$ gathers the values $\bar{\varepsilon}_{i,t} = I_{i,t}\hat{\varepsilon}_{i,t}$, the penalty $g(n,T)$ is such that $g(n,T) \to 0$ and $C_{n,T}^2 g(n,T) \to \infty$, when $n,T \to \infty$, for $C_{n,T}^2 = \min\{n,T\}$. Bai and Ng (2002a) consider several simple potential candidates for the penalty $g(n,T)$. We list and implement them in Section 4. In vector $\bar{\varepsilon}_i$, the unavailable residuals are replaced by zeros. The following model selection rule explains our choice of the diagnostic criterion (8) for approximate factor structure in large unbalanced cross-sectional datasets.

**Proposition 1** *Model selection rule: Under Assumption Assumption 1 and Assumptions A.1-A.5, (a) we select* $\mathcal{M}_1$ *if* $\xi < 0$, *since* $Pr\left(\xi < 0 \mid \mathcal{M}_1\right) \to 1$, *when* $n,T \to \infty$; *(b) we select* $\mathcal{M}_2$ *if* $\xi > 0$, *since* $Pr\left(\xi > 0 \mid \mathcal{M}_2\right) \to 1$, *when* $n,T \to \infty$.

Proposition Proposition 1 characterizes an asymptotically valid model selection rule, which treats both models symmetrically. This is not a testing procedure since we do not use a critical region based an asymptotic distribution and a chosen significance level. The proof of Proposition Proposition 1 shows that the largest eigenvalue in (8) vanishes at a faster rate than the penalization term under $\mathcal{M}_1$ when $n$ and $T$ go to infinity. This explains why we select the first model when $\xi$ is negative. On the contrary, the largest eigenvalue remains bounded from below away from zero under $\mathcal{M}_2$ when $n$ and $T$ go to infinity. This explains why we select the second model when $\xi$ is positive. The criterion (8) can be interpreted as the adjusted gain in fit including a single additional (unobservable) factor in model $\mathcal{M}_1$. In the balanced case, where $I_{i,t} = 1$ for all $i$ and $t$, we can rewrite (8) as $\xi = SS_0 - SS_1 - g\left(n,T\right)$, where $SS_0 = \frac{1}{nT}\sum_i\sum_t \hat{\varepsilon}_{i,t}^2$ is the sum of squared errors and $SS_1 = \min \frac{1}{nT}\sum_i\sum_t \left(\hat{\varepsilon}_{i,t} - \theta_i h_t\right)^2$, where the minimization is w.r.t. the vectors $H \in \mathbb{R}^T$ of factor values and $\Theta \in \mathbb{R}^n$ of factor loadings in a one-factor model, subject to the normalization constraint $\frac{H'H}{T} = 1$. Indeed, the largest eigenvalue $\mu_1\left(\frac{1}{nT}\sum_i \hat{\varepsilon}_i\hat{\varepsilon}_i'\right)$ corresponds to the difference between $SS_0$ and $SS_1$. Furthermore, the criterion $\xi$ is equal to the difference of the penalized criteria for zero- and one-factor models defined in Bai and Ng (2002a) applied on the residuals. Indeed, $\xi = PC\left(0\right) - PC\left(1\right)$, where $PC\left(0\right) = SS_0$, and $PC\left(1\right) = SS_1 + g\left(n,T\right)$. Given such an interpretation in terms of sums of squared errors, we can suggest another diagnostic criterion based on a logarithmic transform as in Corollary 2 of Bai and Ng (2002a). The second diagnostic criterion is

$$\check{\xi} = \ln\left(\frac{1}{nT}\sum_i\sum_t \mathbf{1}_i^{\chi}\hat{\varepsilon}_{i,t}^2\right) - \ln\left(\frac{1}{nT}\sum_i\sum_t \mathbf{1}_i^{\chi}\hat{\varepsilon}_{i,t}^2 - \mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\hat{\varepsilon}_i\hat{\varepsilon}_i'\right)\right) - g(n,T). \qquad (9)$$

In the balanced case, we get $\check{\xi} = \ln(SS_0/SS_1) - g(n,T)$ and it is equal to the difference of $IC\,(0)$ and $IC\,(1)$ criteria in Bai and Ng (2002a). The following proposition states the model selection rule based on $\check{\xi}$.

**Proposition 2** *The model selection rule is the same as in Proposition 1 with $\check{\xi}$ substituted for $\xi$.*

The recent literature on the properties of the two-pass regressions for fixed $n$ and large $T$ shows that the presence of useless factors (Kan and Zhang (1999a,b), Gospodinov, Kan and Robotti (2014)) or weak factor loadings (Kleibergen (2009)) does not affect the asymptotic distributional properties of factor loading estimates, but alters the ones of the risk premia estimates. Useless factors have zero loadings, and weak loadings drift to zero at rate $1/\sqrt{T}$. The vanishing rate of the largest eigenvalue of the empirical cross-sectional covariance matrix of the residuals does not change if we face useless factors or weak factor loadings in the observable factors under $\mathcal{M}_1$. The same remark applies under $\mathcal{M}_2$. Hence the selection rule remains the same since the probability of taking the right decision still approaches 1. If we have a number of useless factors or weak factor loadings strictly lower than the number $m$ of the omitted factors under $\mathcal{M}_2$, this does not impact the asymptotic rate of the diagnostic criterion if Assumption Assumption 1 holds. If we only have useless factors in the omitted factors under $\mathcal{M}_2$, we face an identification issue. Assumption Assumption 1 is not satisfied. We cannot distinguish such a specification from $\mathcal{M}_1$ since it corresponds to a particular approximate factor structure. Again the selection rule remains the same since the probability of taking the right decision still approaches 1. Finally, let us study the case of only weak factor loadings under $\mathcal{M}_2$. We consider a simplified setting:

$$R_{i,t} = x_{i,t}'\beta_i + \epsilon_{i,t}$$

where $\epsilon_{i,t} = \theta_i h_t + u_{i,t}$ has only one factor with a weak factor loading, namely $m = 1$ and $\theta_i = \bar{\theta}_i/T^{\gamma}$ with $\gamma > 0$. Let us assume that $\mu_1\left(\frac{1}{nT}\sum_i M_{\tilde{X}_i}\tilde{H}_i\tilde{H}_i'M_{\tilde{X}_i}\bar{\theta}_i^2\right)$ is bounded from below away from zero (see Assumption Assumption 1) and bounded from above. By the properties of the eigenvalues of a scalar multiple of a matrix, we deduce that $c_1/T^{2\gamma} \leq \mu_1\left(\frac{1}{nT}\sum_i M_{\tilde{X}_i}\tilde{H}_i\tilde{H}_i'M_{\tilde{X}_i}\theta_i^2\right) \leq c_2/T^{2\gamma}$, for some constants

$c_1, c_2$ such that $c_2 \geq c_1 > 0$. Hence, by similar arguments as in the proof of Proposition Proposition 1, we get:

$$c_1 T^{-2\gamma} - g(n,T) + O_p\left(C_{nT}^{-2} + \bar{\chi}_T T^{-1}\right) \leq \xi \leq c_2 T^{-2\gamma} - g(n,T) + O_p\left(C_{nT}^{-2} + \bar{\chi}_T T^{-1}\right),$$

where we define $\bar{\chi}_T = \chi_{1,T}^4 \chi_{2,T}^2$. To conclude $\mathcal{M}_2$, we need that $C_{nT}^{-2} + \bar{\chi}_T T^{-1}$ and the penalty $g(n,T)$ vanish at a faster rate than $T^{-2\gamma}$, namely $C_{nT}^{-2} + \bar{\chi}_T T^{-1} = o\left(T^{-2\gamma}\right)$ and $g(n,T) = o\left(T^{-2\gamma}\right)$. To conclude $\mathcal{M}_1$, we need that $g(n,T)$ is the dominant term, namely $T^{-2\gamma} = o\left(g(n,T)\right)$ and $C_{nT}^{-2} + \bar{\chi}_T T^{-1} = o\left(g(n,T)\right)$. As an example, let us take $g(n,T) = T^{-1}\log T$ and $n = T^{\bar{\gamma}}$ with $\bar{\gamma} > 1$, and assume that the trimming is such that $\bar{\chi}_T = o(\log T)$. Then, we conclude $\mathcal{M}_2$ if $\gamma < 1/2$ and $\mathcal{M}_1$ if $\gamma > 1/2$. This means that detecting a weak factor loading structure is difficult if gamma is not sufficiently small. The factor loading should drift to zero not too fast to conclude $\mathcal{M}_2$. Otherwise, we cannot distinguish it asymptotically from an approximate factor structure.

# 4 Determining the number of factors

In the previous section, we have studied a diagnostic criterion to check whether the error terms are weakly cross-sectionally correlated or share at least one unobservable common factor. This section aims at answering: do we have one, two, or more omitted factors? The design of the diagnostic criterion to check whether the error terms share exactly $k$ unobservable common factors or share at least $k+1$ unobservable common factors follows the same mechanics. We consider the following rival models:

$\mathcal{M}_1(k)$ : the linear regression model (3), where the errors $(\varepsilon_{i,t})$ satisfy a factor structure with exactly $k$ unobservable factors,

and

$\mathcal{M}_2(k)$ : the linear regression model (3), where the errors $(\varepsilon_{i,t})$ satisfy a factor structure with at least $k+1$ unobservable factors.

The following assumption consists of identification assumptions similar to Assumption Assumption 1, but for exactly $k$ unobservable factors and at least $k+1$ unobservable factors in the error terms.

**Assumption 2** *a) Under model $\mathcal{M}_1(k)$, the factor structure in the error term is such that*

$$\mu_k\left(\frac{1}{nT}\sum_i M_{\tilde{X}_i}\tilde{H}_i\theta_i\theta_i'\tilde{H}_i'M_{\tilde{X}_i}\right) \geq c, \quad \textit{with probability approaching 1},$$

*for a constant $c > 0$. b) Under model $\mathcal{M}_2(k)$, the factor structure in the error term is such that*

$$\mu_{k+1}\left(\frac{1}{nT}\sum_i M_{\tilde{X}_i}\tilde{H}_i\theta_i\theta_i'\tilde{H}_i'M_{\tilde{X}_i}\right) \geq c, \quad \textit{with probability approaching 1},$$

*for a constant $c > 0$.*

For a factor model with time invariant coefficients and a balanced panel, Assumption Assumption 2 is satisfied if matrices $\frac{\Theta'\Theta}{n}$ and $\frac{H'M_XH}{T}$ converge to positive definite matrices since we deduce as in Section 2 from the inequalities in Wang and Zhang (1992):

$$\mu_k\left(\frac{1}{nT}\sum_1 M_XH\theta_i\theta_i'H'M_X\right) \geq \mu_m\left(\frac{\Theta'\Theta}{n}\right)\mu_k\left(\frac{H'M_XH}{T}\right).$$

The diagnostic criterion exploits the $k$th largest eigenvalue of the empirical cross-sectional covariance matrix of the residuals:

$$\xi(k) = \mu_{k+1}\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) - g(n,T). \tag{10}$$

As discussed in Ahn and Horenstein (2013) (see also Onatski (2013)), we can rewrite (10) in the balanced case as $\xi(k) = SS_k - SS_{k+1} - g(n,T)$ where $SS_k$ equals the sample mean of the squared residuals from the time series regressions of individual response variables $(\hat{\varepsilon}_{i,t})$ on the first $k$ principal components of $\frac{1}{nT}\sum_i \hat{\varepsilon}_i\hat{\varepsilon}_i'$. The criterion $\xi(k)$ is equal to the difference of the penalized criteria for $k$ and $(k+1)$- factor models defined in Bai and Ng (2002a) applied on the residuals. Indeed, $\xi(k) = PC(k) - PC(k+1)$, where $PC(k) = SS_k + kg(n,T)$ , and $PC(k+1) = SS_{k+1} + (k+1)g(n,T)$. To determine the number of unobservable factors, we choose the minimum $k$ such that $\xi(k) < 0$. Graphically, we can build a penalized scree plot where we display the penalized eigenvalues associated with each factor in descending order versus the number of the factor, and use the $x$-axis for the cut-off point. The following model selection rule extends Proposition Proposition 1 to determine the number of factors.

**Proposition 3** *Model selection rule: under Assumptions..., (a) we select $\mathcal{M}_1(k)$ if $\xi(k) < 0$, since $Pr[\xi(k) < 0|\mathcal{M}_1(k)] \to 1$, when $n, T \to \infty$ such that $n = O(T^2)$; (b) we select $\mathcal{M}_2(k)$ if $\xi(k) > 0$, since $Pr[\xi(k) > 0|\mathcal{M}_2(k)] \to 1$, when $n, T \to \infty$.*

In Proposition Proposition 3 part a), we need the additional constraint $n = O(T^2)$ on the relative rate of the cross-sectional dimension w.r.t. the time series dimension. The contribution $\mu_{k+1}\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right) = O_p(1/\sqrt{\max\{n,T\}})$ coming from the $k$ omitted factors (Lemma ... in the appendix) does not dominate asymptotically the contribution $\mu_1\left(\frac{1}{nT}\sum_i \tilde{u}_i\tilde{u}_i'\right) = C_{n,T}^{-2}$ under $\mathcal{M}_1(k)$ when $n = O(T^2)$. We do not need such an additional restriction in a balanced panel since $\mu_{k+1}\left(H\left(\frac{1}{nT}\sum_i \theta_i\theta_i'\right)H'\right) = 0$ if we have exactly $k$ factors. This exemplifies a key difference between the asymptotics for balanced and unbalanced panels, and the proportional asymptotics used in Onatski (2009, 2010) or Ahn and Horenstein (2013). Those papers rely on the asymptotic distribution of the eigenvalues of large dimensional sample covariances matrices when $n/T(n) \to c > 0$ as $n \to \infty$. The condition $n = O(T^2)$ agrees with the "large $n$, small $T$" case that we face in the empirical application (ten thousands individual stocks monitored over forty-five years of monthly returns). The proof of Proposition Proposition 3 is also more complicated than the proof of Proposition Proposition 1. The proof of the latter directly exploits the equality between the largest value of a symmetric matrix and its operator norm, the triangular inequality of the matrix norm, and its upper bound given by the Frobenius norm. We need additional arguments based on Weyl inequalities (Theorem 4.3.1 in Horn and Johnson (1985)) when we look at the $k+1$th eigenvalue.

# 5 Empirical results

## 5.1 Factor models and data description

We consider fifteen non-repetitive empirical factors as in Ahn, Horenstein and Wang (2013). The three factor of Fama and French (1993) are the monthly excess return on CRSP NYSE/AMEX/Nasdaq value-weighted market portfolio over the risk free rate $r_{m,t}$, the monthly returns on zero-investment factor-mimicking portfolios for size, book-to-market, denoted by $r_{smb,t}$ and $r_{hml,t}$ respectively. The monthly returns on portfolio for momentum is denoted by $r_{mom,t}$. Two reversal factors are monthly returns on portfolio for short $r_{str}$, and long term $r_{ltr}$. We have downloaded the time series of these factors from the website of Kenneth French. We consider the five factors of Chen, Roll and Ross (1986) available from Laura Xiaolei Liu's webpage. The monthly CRR factors are the growth rate of industrial production $mp_t$, the unexpected inflation $ui_t$, the term spread $uts_t$, proxied by the difference between yields on 10-year Treasury and 3-month T-bill, and the default premia $upr_t$, proxied by the yield difference between Moody's Baa-rated and Aaa-rated corporate bonds. Moreover, we consider the three liquidity-related factors of Pastor and Stambaugh (2002) that concern of the monthly liquidity level $al_t$, traded liquidity $tl_t$ and the innovation in aggregate liquidity $il_t$. We have downloaded the LIQ factors from the website of Lubos Pastor. Finally, we build the monthly growth rate of labor income $lab_t$ from the Bureau of Economic Analysis's webpage. We proxy the risk free rate with the monthly 30-day T-bill beginning-of-month yield. To account for time-varying coefficients, we use two conditional specifications based on two common variables and a firm-level variable. We take the instruments $Z_t = (1,\ Z_t^{*\prime})'$, where bivariate vector $Z_t^*$ includes either (i) the term spread and the default spread, or (ii) the monthly 30-day T-bill and the dividend yields. We take a scalar $Z_{i,t}$ corresponding to the book-to-market equity of firm $i$. We refer to Avramov and Chordia (2006) for convincing theoretical and empirical arguments in favor of the chosen conditional specification. The parsimony explains why we have not included e.g. the size of firm $i$ as an additional stock specific instrument.

Table 1 reports the thirteen linear factor models that we estimate in order to computed the diagnostic criteria. For each model, we specify the empirical factors involved and the number $K$ of observable factors. We look at factor models popular in the empirical finance. We also consider nested models built from the fifteen empirical factors.

We compute the firm characteristics from Compustat as in the appendix of Fama and French (2008). The CRSP database provides the monthly stock returns data and we exclude financial firms (Standard Industrial Classification Codes between 6000 and 6999) as in Fama and French (2008). The dataset after matching CRSP and Compustat contents comprises $n = 10,442$ stocks, and covers the period from January 1968 to December 2011 with $T = 528$ months.

### 5.2 Diagnostic results

In this section, we compute the diagnostic criteria in Equations (8) and (9) assuming time-invariant and time-varying specifications of the linear factor models in Table 1. In order to compute the criteria we need to define the specification for the penalty $g\left(n,T\right)$. Bai and Ng (2002a) propose three choices for the penalty function in Equation (8), leading to the following criteria:

1. $\xi_1 = \mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) - \hat{\sigma}^2\left(\frac{n+T}{nT}\right)\ln\left(\frac{nT}{n+T}\right)$;

2. $\xi_2 = \mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) - \hat{\sigma}^2\left(\frac{n+T}{nT}\right)\ln C_{nT}^2$;

3. $\xi_3 = \mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) - \hat{\sigma}^2\left(\frac{\ln C_{nT}^2}{C_{nT}^2}\right)$,

where $\hat{\sigma}^2 = \frac{1}{nT}\sum_i\sum_t \mathbf{1}_i^{\chi}\bar{\varepsilon}_{1,i,t}^2$, and $\bar{\varepsilon}_{1,i,t}$ is the fitted residual of the time-varying linear factor model built on the FF, MOM, REV observable factors and a latent factor. Furthermore, we define the specification for the penalty $g(n,T)$ for the logarithmic diagnostic criterion in Equation (9). We get the following logarithmic criteria:

1. $\check{\xi}_1 = \ln\left(\frac{1}{nT}\sum_i\sum_t \mathbf{1}_i^{\chi}\bar{\varepsilon}_{i,t}^2\right) - \ln\left(\frac{1}{nT}\sum_i\sum_t \mathbf{1}_i^{\chi}\bar{\varepsilon}_{i,t}^2 - \mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right)\right) - \left(\frac{n+T}{nT}\right)\ln\left(\frac{nT}{n+T}\right)$;

2. $\check{\xi}_2 = \ln\left(\frac{1}{nT}\sum_i\sum_t \mathbf{1}_i^{\chi}\bar{\varepsilon}_{i,t}^2\right) - \ln\left(\frac{1}{nT}\sum_i\sum_t \mathbf{1}_i^{\chi}\bar{\varepsilon}_{i,t}^2 - \mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right)\right) - \left(\frac{n+T}{nT}\right)\ln C_{nT}^2$;

3. $\check{\xi}_3 = \ln\left(\frac{1}{nT}\sum_i\sum_t \mathbf{1}_i^{\chi}\bar{\varepsilon}_{i,t}^2\right) - \ln\left(\frac{1}{nT}\sum_i\sum_t \mathbf{1}_i^{\chi}\bar{\varepsilon}_{i,t}^2 - \mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right)\right) - \left(\frac{\ln C_{nT}^2}{C_{nT}^2}\right)$,

In order to ensure that all series have a common scale of measurement, each time-series is demeaned and standardized to have unit variance before to computing the eigenvalues (see Pena and Poncela (2006)). In order to compute the diagnostic criteria, we estimate time-invariant and time-varying factor models. We fix $\chi_{1,T} = 15$ as advocated by Greene (2008), and $\chi_{2,T} = 546/12$ for the time-invariant estimation and $\chi_{1,T} = 20$ and $\chi_{2,T} = 546/60$ for the time-varying estimation. In Table 2, we report the size of trimmed cross-sectional dimension $n^{\chi}$ that comes from the trimming procedure applied in the estimation approach.

In some time-varying specifications, we incur in empirical multicollinearity problems due to the correlations within the vector of regressors $x_{i,t}$, that involves cross product of factors $f_t$ and instruments $Z_{t-1}$ (e.g., in the JW and CRR models), and the large dimension of vector $x_{i,t}$ (e.g., the number of parameter to estimate is larger than 40 in models 11-13).

For the time-invariant specifications of (1)-(13) models, we plot the values of the diagnostic criteria $\xi_1, \xi_2$ and $\xi_3$ in Figure 1, and $\check{\xi}_1, \check{\xi}_2$ and $\check{\xi}_3$ in Figure 2. For the time-varying specifications, Figures 3 and 4 plot the values of the diagnostic criteria computed by using the common instruments (i). Figures 5 and 6 plot the results by using the second set of common instruments. Since the penalty function is proportional to $\frac{1}{T} \ln T$, the numerical value of criteria $\xi_s$ and $\check{\xi}_s$, with $s = 1, 2, 3$, are not too much different from each other. For the majority of the models, the conclusion about the selection model is the same both for the diagnostic criterion in equation (8), that for the logarithmic diagnostic criterion in equation (9). In particular, we cannot select a time-invariant model with zero factors in the errors. We conclude for an approximate factor structure in the error terms when we estimate the time-varying linear factor models based on FF and REV factors. In general, focusing on nested models, when the number of factor increases the diagnostic criteria decreases. Finally, in many cases, the diagnostic criteria is smaller for the time-varying specifications than for the time-invariant models.

In Tables 3-6, we compare the descriptive statistics of four measures of missing factor impact: (i) the estimated time-series coefficient of determination $\hat{\rho}_i^2 = \frac{ESS_i}{TSS_i}$,- where $ESS_i = \sum_t I_{i,t} \left( \hat{R}_{i,t} - \bar{\hat{R}}_i \right)^2$, with $\hat{R}_{i,t} = \hat{\beta}_i' x_{i,t}$ and $\bar{\hat{R}}_i = \frac{1}{T_i} \sum_t I_{i,t} \hat{R}_{i,t}$, and $TSS_i = \sum_t I_{i,t} \left( R_{i,t} - \bar{R}_i \right)^2$, with $\bar{R}_i = \frac{1}{T_i} \sum_t I_{i,t} R_{i,t}$; (ii) the estimated adjusted $R^2$ defined by $\hat{\rho}_{ad,i}^2 = 1 - \frac{(T_i - 1)}{(T_i - d)} \left( 1 - \hat{\rho}_i^2 \right)$; (iii) the idiosyncratic risk $IdiVol_i = \sqrt{\frac{RSS_i}{T_i}}$, with $RSS_i = \sum_t I_{i,t} \hat{\varepsilon}_{i,t}^2$; (iv) the systematic risk $SysRisk_i = \sqrt{\frac{ESS_i}{T_i}}$, for the time-invariant and time-varying specifications. We consider those estimates as measures of missing factor impact (see Ang, Liu and Schwarz (2008)). The time-series (adjusted) coefficient of determination tend to be a bit larger in the time-varying model than in the time-invariant specifications. The $\hat{\rho}_i^2$, $\hat{\rho}_{ad,i}^2$ and $SysRisk_i$ admit large values for the models that introduced the FF, MOM and/or REV factors in their specification. For these linear specifications, we observe that the diagnostic criteria $\xi$ and $\check{\xi}$ admits small values.

### 5.3 The number of factors

In this section, we compute the diagnostic criteria (9) that exploit the $k$-th largest eigenvalue of the empirical cross-sectional covariance matrix of the errors. We compute the diagnostic criteria for the first five eigenvalues, and we use the penalty function $g(n,T)$ defined in the previous section. In particular, we compute the following criteria:

1. $\xi_1(k) = \mu_{k+1}\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i \bar{\varepsilon}_i'\right) - \hat{\sigma}^2 \left(\frac{n+T}{nT}\right) \ln\left(\frac{nT}{n+T}\right)$;
2. $\xi_2(k) = \mu_{k+1}\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i \bar{\varepsilon}_i'\right) - \hat{\sigma}^2 \left(\frac{n+T}{nT}\right) \ln C_{nT}^2$;
3. $\xi_3(k) = \mu_{k+1}\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i \bar{\varepsilon}_i'\right) - \hat{\sigma}^2 \left(\frac{\ln C_{nT}^2}{C_{nT}^2}\right)$,

with $k = 1, ..., 5$. For each linear factor specification, we build a penalized scree plot. Figures 7 and 8 show the results for the time-invariant specifications. We observe that diagnostic criteria change signs when we consider the time-invariant specifications based on the FF factors. In particular, the diagnostic criteria becomes negative when $k = 4$ for the FF and Carhart (1997) models. The number of unobservable common factors $k$ is 3 for the time-invariant model that accounts for more than $8$ observable factors (e.g., models (11)-(13)). However, the three FF factors alone do not explain the excess returns for stocks. Let us consider the results for the time-varying specifications in Figures 9 and 10. In both the figures, the cut-off point is smaller than for the time-invariant specifications. Thus, the time-varying specifications capture more properties of excess returns than the corresponding time-invariant models. Indeed, the number of omitted factors is smaller for the time-varying models than for the time-invariant cases. Moreover, the set of common instruments involving the monthly 30-day T-bill and the dividend yields seems to capture in a better way the characteristics of returns of individual stocks.

**Table 1: Linear factor models**

| Model | Empirical factors | $K$ |
|---|---|---|
| (1) CAPM | $r_{m,t}$ | 1 |
| (2) FF model | $r_{m,t}, r_{smb,t}, r_{hml,t}$ | 3 |
| (3) LIQ model | $al_t, tl_t, il_t$ | 3 |
| (4) JW model | $r_{m,t}, lab_{t,} upr_t$ | 3 |
| (5) MOM and REV factors | $r_{mom,t}, r_{str}, r_{ltr}$ | 3 |
| (6) Carhart (1997) model | $r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}$ | 4 |
| (7) CRR model | $mp_t, ui_t, dei_t uts_t upr_t$ | 5 |
| (8) FF and REV factors | $r_{m,t}, r_{smb,t}, r_{hml,t}, r_{str}, r_{ltr}$ | 5 |
| (9) FF and JW factors | $r_{m,t}, r_{smb,t}, r_{hml,t}, lab_{t,} upr_t$ | 5 |
| (10) FF, MOM and REV factors | $r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}, r_{str}, r_{ltr}$ | 6 |
| (11) FF and CRR factors | $r_{m,t}, r_{smb,t}, r_{hml,t}, mp_t, ui_t, dei_t$ $uts_t upr_t$ | 8 |
| (12) FF, CRR and JW factors | $r_{m,t}, r_{smb,t}, r_{hml,t}, mp_t, ui_t, dei_t,$ $uts_t upr_t, lab_t$ | 9 |
| (13) FF, MOM, REV, CRR and JW factors | $r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}, r_{str}, r_{ltr},$ $mp_t, ui_t, dei_t uts_t upr_t, lab_t$ | 12 |

The table lists the linear factor models that we estimate in order to compute the diagnostic criteria. For each model, the empirical factors involved are specified. $K$ is the number of observable factors. FF, CRR, MOM, REV, LIQ and JW, respectively, refer to the three Fama-French factors, the five Chen-Roll-Ross macroeconomic factors, the momentum factor, the reversal factors, the three liquidity factors of Pastor and Stambaugh (2002), and the three Jagannathan and Wang (1996) factors.

**Table 2: Trimmed cross-sectional dimensions $n^{\chi}$ and number of parameter to estimate $d$**

| Model | time-invariant spec. | time-varying spec. | (i) | (ii) |
|---|---|---|---|---|
| | $n^{\chi}$ | $d$ | $n^{\chi}$ | $n^{\chi}$ |
| (1) CAPM | 10,410 | 13 | 5,046 | 1,661 |
| (2) FF model | 10,410 | 21 | 4,476 | 1,476 |
| (3) LIQ model | 10,410 | 21 | 3,393 | 1,008 |
| (4) JW model | 7,578 | 21 | - | - |
| (5) MOM and REV factors | 10,410 | 21 | 4,568 | 1,471 |
| (6) Carhart (1997) model | 10,410 | 25 | 4,020 | 1,354 |
| (7) CRR model | 7,171 | 29 | - | - |
| (8) FF and REV factors | 10,396 | 29 | 3,828 | 1,076 |
| (9) FF and JW factors | 5,271 | 29 | - | - |
| (10) FF, MOM and REV factors | 7,461 | 33 | 3,217 | 960 |
| (11) FF and CRR factors | 6,786 | 41 | - | - |
| (12) FF, CRR and JW factors | 6,110 | 45 | - | - |
| (13) FF, MOM, REV, CRR and JW factors | 5,572 | 57 | - | - |

For each linear factor model, the table reports the trimmed cross-sectional dimension $n^{\chi}$ that comes from the estimation procedure. For the time-varying specifications, the dimension of vector $x_{i,t}$, denoted by $d$, is also specified. For the time-invariant specifications, the number of regressors corresponds to the number of observable factors $K$ (see Table 1).

**Table 3: Summary statistics of $\hat{\rho}_i^2$, $\hat{\rho}_{ad,i}^2$, $IdiVol_i$ and $SysRisk_i$ for the time-invariant specifications**

| Model | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | $\hat{\rho}_i$ | | | | | | |
| Min | 0.0000 | 0.0013 | 0.0012 | 0.0009 | 0.0005 | 0.0050 | 0.0019 | 0.0025 | 0.0060 | 0.0090 | 0.0083 | 0.0130 | 0.0305 |
| Quantile 25% | 0.0948 | 0.1475 | 0.0429 | 0.1071 | 0.0521 | 0.1618 | 0.0338 | 0.1746 | 0.1534 | 0.1857 | 0.1730 | 0.1754 | 0.1972 |
| Median | 0.1872 | 0.2509 | 0.0889 | 0.1882 | 0.1107 | 0.2671 | 0.0652 | 0.2803 | 0.2454 | 0.2950 | 0.2596 | 0.2617 | 0.2823 |
| Mean | 0.2399 | 0.2948 | 0.1761 | 0.2198 | 0.1974 | 0.3111 | 0.1069 | 0.3239 | 0.2678 | 0.3383 | 0.2774 | 0.2782 | 0.2995 |
| Quantile 75% | 0.3172 | 0.3856 | 0.2110 | 0.2954 | 0.2576 | 0.4051 | 0.1274 | 0.4181 | 0.3525 | 0.4374 | 0.3623 | 0.3617 | 0.3827 |
| Max | 0.9828 | 0.9849 | 0.9863 | 0.9514 | 0.9868 | 0.9916 | 0.9535 | 0.9933 | 0.9574 | 0.9971 | 0.9582 | 0.9473 | 0.8934 |
| Std | 0.2003 | 0.2020 | 0.2105 | 0.1541 | 0.2143 | 0.2043 | 0.1232 | 0.2044 | 0.1561 | 0.2072 | 0.1420 | 0.1379 | 0.1389 |
| | | | | | | | $\hat{\rho}_{ad,i}$ | | | | | | |
| Min | -0.0689 | -0.2114 | -0.2900 | -0.1845 | -0.2223 | -0.3304 | -0.1898 | -0.3737 | -0.3639 | -0.5507 | -0.1401 | -0.1761 | -0.2287 |
| Quantile 25% | 0.0845 | 0.1164 | 0.0188 | 0.0841 | 0.0282 | 0.1219 | 0.005 | 0.1253 | 0.1140 | 0.1274 | 0.1158 | 0.1203 | 0.1302 |
| Median | 0.1778 | 0.2220 | 0.0571 | 0.1664 | 0.0781 | 0.2276 | 0.0275 | 0.2319 | 0.2093 | 0.2379 | 0.2103 | 0.2115 | 0.2219 |
| Mean | 0.2285 | 0.2621 | 0.1388 | 0.1955 | 0.1614 | 0.2679 | 0.0627 | 0.2701 | 0.2298 | 0.2744 | 0.2234 | 0.2239 | 0.2332 |
| Quantile 75% | 0.3067 | 0.3549 | 0.1587 | 0.2735 | 0.2074 | 0.3618 | 0.0740 | 0.3652 | 0.3175 | 0.3723 | 0.3117 | 0.3106 | 0.3232 |
| Max | 0.9815 | 0.9808 | 0.9811 | 0.9417 | 0.9826 | 0.9878 | 0.9324 | 0.9884 | 0.9440 | 0.9937 | 0.9344 | 0.9176 | 0.8182 |
| Std | 0.2004 | 0.2026 | 0.2053 | 0.1534 | 0.2095 | 0.2049 | 0.1122 | 0.2044 | 0.1561 | 0.2071 | 0.1429 | 0.1383 | 0.1395 |

The table contains the descriptive statistics (cross-sectional minimum, $25\%$ and $75\%$ quantiles, median, mean, maximum and standard deviation) of the estimated coefficient of determination ($\hat{\rho}_i^2$), the estimated adjusted coefficients of determination ($\hat{\rho}_{ad,i}^2$) for the time-invariant linear factor models.

**Table 4: Summary statistics of $\hat{\rho}_i^2$, $\hat{\rho}_{ad,i}^2$, $IdiVol_i$ and $SysRisk_i$ for the time-invariant specifications**

| Model | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | $IdiVol_i$ | | | | | | |
| Min | 0.0109 | 0.0104 | 0.0103 | 0.0105 | 0.0106 | 0.0104 | 0.0085 | 0.0101 | 0.0102 | 0.0101 | 0.0081 | 0.0081 | 0.0260 |
| Quantile 25% | 0.1121 | 0.1069 | 0.1183 | 0.1046 | 0.1165 | 0.1057 | 0.1128 | 0.1042 | 0.0997 | 0.1029 | 0.0964 | 0.0947 | 0.0912 |
| Median | 0.1558 | 0.1488 | 0.1616 | 0.1432 | 0.1591 | 0.1466 | 0.1521 | 0.1453 | 0.1376 | 0.1432 | 0.1322 | 0.1283 | 0.1229 |
| Mean | 0.1781 | 0.1707 | 0.1829 | 0.1613 | 0.1797 | 0.1679 | 0.1684 | 0.1656 | 0.1555 | 0.1630 | 0.1482 | 0.1426 | 0.1366 |
| Quantile 75% | 0.2158 | 0.2072 | 0.2196 | 0.1951 | 0.2166 | 0.2037 | 0.2016 | 0.2015 | 0.1888 | 0.1986 | 0.1797 | 0.1731 | 0.1661 |
| Max | 3.0700 | 3.0206 | 3.2080 | 2.6677 | 2.8717 | 2.7810 | 3.2114 | 2.7925 | 2.3288 | 2.5306 | 2.0782 | 1.3228 | 1.3010 |
| Std | 0.1097 | 0.1045 | 0.1071 | 0.0904 | 0.1029 | 0.1013 | 0.0923 | 0.0996 | 0.0867 | 0.0965 | 0.0804 | 0.0711 | 0.0680 |
| | | | | | | | $SysRisk_i$ | | | | | | |
| Min | 0.001 | 0.0039 | 0.0022 | 0.0041 | 0.0022 | 0.0056 | 0.0045 | 0.0065 | 0.0059 | 0.0075 | 0.0083 | 0.0084 | 0.0100 |
| Quantile 25% | 0.0471 | 0.0572 | 0.0308 | 0.0469 | 0.0338 | 0.0589 | 0.0258 | 0.0608 | 0.0548 | 0.0623 | 0.0566 | 0.0560 | 0.0575 |
| Median | 0.0702 | 0.0825 | 0.0516 | 0.0660 | 0.0585 | 0.0854 | 0.0410 | 0.0880 | 0.0757 | 0.0904 | 0.0772 | 0.0758 | 0.0777 |
| Mean | 0.1030 | 0.1162 | 0.0904 | 0.0831 | 0.0970 | 0.1202 | 0.0570 | 0.1232 | 0.0924 | 0.1264 | 0.0883 | 0.0846 | 0.0857 |
| Quantile 75% | 0.1108 | 0.1268 | 0.0960 | 0.0949 | 0.1071 | 0.1312 | 0.0663 | 0.1359 | 0.1064 | 0.1396 | 0.1062 | 0.1028 | 0.1046 |
| Max | 3.4809 | 3.5239 | 3.3542 | 3.7981 | 3.6462 | 3.7159 | 3.3509 | 3.7073 | 4.0148 | 3.8908 | 4.1501 | 1.0911 | 1.1171 |
| Std | 0.1190 | 0.1223 | 0.1241 | 0.0812 | 0.1274 | 0.1249 | 0.0702 | 0.1257 | 0.0837 | 0.1283 | 0.0700 | 0.0454 | 0.0437 |

The table contains the descriptive statistics (cross-sectional minimum, $25\%$ and $75\%$ quantiles, median, mean, maximum and standard deviation) of the idiosyncratic risks ($IdiVol_i$), and the systematic risks ($SysRisk_i$) for the time-invariant linear factor models.

**Table 5: Summary statistics of $\hat{\rho}_i^2$, $\hat{\rho}_{ad,i}^2$, $IdiVol_i$ and $SysRisk_i$ for the time-varying specifications (i)**

| Model | 1 | 2 | 3 | 5 | 6 | 8 | 10 |
|---|---|---|---|---|---|---|---|
| | | | | $\hat{\rho}_i$ | | | |
| Min | 0.0218 | 0.0465 | 0.0242 | 0.0325 | 0.0514 | 0.0547 | 0.0559 |
| Quantile 25% | 0.1569 | 0.2294 | 0.1037 | 0.1414 | 0.2517 | 0.2673 | 0.2790 |
| Median | 0.2305 | 0.3112 | 0.1434 | 0.2024 | 0.3331 | 0.3471 | 0.3598 |
| Mean | 0.2440 | 0.3225 | 0.1612 | 0.2316 | 0.3435 | 0.3566 | 0.3680 |
| Quantile 75% | 0.3191 | 0.4069 | 0.1984 | 0.2938 | 0.4261 | 0.4370 | 0.4468 |
| Max | 0.7052 | 0.8512 | 0.7069 | 0.8822 | 0.8628 | 0.9032 | 0.9051 |
| Std | 0.1147 | 0.1265 | 0.0832 | 0.1228 | 0.1272 | 0.1229 | 0.1221 |
| | | | | $\hat{\rho}_{ad,i}$ | | | |
| Min | -0.1268 | -0.2463 | -0.1287 | -0.2671 | -0.3089 | -0.2981 | -0.3427 |
| Quantile 25% | 0.0884 | 0.1306 | 0.0214 | 0.0532 | 0.1435 | 0.1461 | 0.1519 |
| Median | 0.1666 | 0.2240 | 0.0563 | 0.1052 | 0.2358 | 0.2429 | 0.2497 |
| Mean | 0.1794 | 0.2296 | 0.0656 | 0.1270 | 0.2404 | 0.2433 | 0.2479 |
| Quantile 75% | 0.2575 | 0.3223 | 0.0983 | 0.1764 | 0.3335 | 0.3374 | 0.3423 |
| Max | 0.6764 | 0.7835 | 0.5897 | 0.8287 | 0.7804 | 0.8279 | 0.8101 |
| Std | 0.1201 | 0.1362 | 0.0708 | 0.1132 | 0.1380 | 0.1365 | 0.1369 |
| | | | | $IdiVol_i$ | | | |
| Min | 0.0358 | 0.0315 | 0.0387 | 0.0334 | 0.0312 | 0.0322 | 0.0311 |
| Quantile 25% | 0.0948 | 0.0868 | 0.0980 | 0.0965 | 0.0847 | 0.0838 | 0.0809 |
| Median | 0.1283 | 0.1171 | 0.1338 | 0.1274 | 0.1134 | 0.1125 | 0.1083 |
| Mean | 0.1421 | 0.1310 | 0.1460 | 0.1393 | 0.1268 | 0.1255 | 0.1209 |
| Quantile 75% | 0.1730 | 0.1603 | 0.1774 | 0.1679 | 0.1545 | 0.1537 | 0.1472 |
| Max | 0.7487 | 0.6984 | 0.7015 | 0.6770 | 0.6842 | 0.6529 | 0.6236 |
| Std | 0.0683 | 0.0630 | 0.0682 | 0.0627 | 0.0611 | 0.0601 | 0.0574 |
| | | | | $SysRisk_i$ | | | |
| Min | 0.0091 | 0.0131 | 0.0102 | 0.0106 | 0.0155 | 0.0150 | 0.0185 |
| Quantile 25% | 0.0523 | 0.0597 | 0.0387 | 0.0463 | 0.0609 | 0.0623 | 0.0617 |
| Median | 0.0698 | 0.0805 | 0.0555 | 0.0667 | 0.0820 | 0.0834 | 0.0823 |
| Mean | 0.0766 | 0.0874 | 0.0638 | 0.0765 | 0.0894 | 0.0917 | 0.0909 |
| Quantile 75% | 0.0931 | 0.1068 | 0.0803 | 0.0954 | 0.1096 | 0.1115 | 0.1096 |
| Max | 0.4210 | 0.6166 | 0.5218 | 0.6278 | 0.6208 | 0.6352 | 0.6359 |
| Std | 0.0360 | 0.0407 | 0.0379 | 0.0442 | 0.0422 | 0.0438 | 0.0442 |

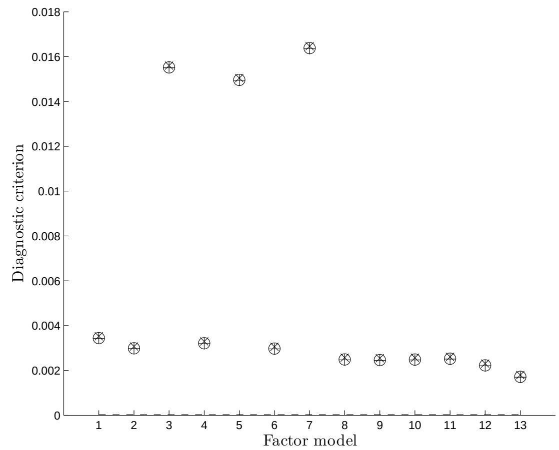
The table contains the descriptive statistics (cross-sectional minimum, 25% and 75% quantiles, median, mean, maximum and standard deviation) of the estimated coefficient of determination ($\hat{\rho}_i^2$), the estimated adjusted coefficients of determination ($\hat{\rho}_{ad,i}^2$), the idiosyncratic risks ($IdiVol_i$), and the systematic risks ($SysRisk_i$) for the time-varying linear factor models estimated by using the term spread and the default spread as common instruments.

**Table 6: Summary statistics of $\hat{\rho}_i^2$, $\hat{\rho}_{ad,i}^2$, $IdiVol_i$ and $SysRisk_i$ for the time-varying specifications (ii)**

| Model | 1 | 2 | 3 | 5 | 6 | 8 | 10 |
|---|---|---|---|---|---|---|---|
| | | | | $\hat{\rho}_i$ | | | |
| Min | 0.0210 | 0.0490 | 0.0200 | 0.0306 | 0.0562 | 0.0688 | 0.0730 |
| Quantile 25% | 0.1368 | 0.2034 | 0.0827 | 0.1096 | 0.2240 | 0.2366 | 0.2508 |
| Median | 0.2012 | 0.2795 | 0.1128 | 0.1545 | 0.3044 | 0.3013 | 0.3162 |
| Mean | 0.2200 | 0.3010 | 0.1349 | 0.1908 | 0.3237 | 0.3100 | 0.3225 |
| Quantile 75% | 0.2893 | 0.3778 | 0.1638 | 0.2423 | 0.4034 | 0.3761 | 0.3876 |
| Max | 0.7134 | 0.7780 | 0.5545 | 0.7180 | 0.7885 | 0.8491 | 0.7868 |
| Std | 0.1094 | 0.1306 | 0.0783 | 0.1157 | 0.1336 | 0.1070 | 0.1049 |
| | | | | $\hat{\rho}_{ad,i}$ | | | |
| Min | -0.0934 | -0.1017 | -0.1479 | -0.2031 | -0.1336 | -0.1400 | -0.1738 |
| Quantile 25% | 0.0765 | 0.1174 | 0.0170 | 0.0425 | 0.1292 | 0.1263 | 0.1313 |
| Median | 0.1487 | 0.2095 | 0.0472 | 0.0784 | 0.2190 | 0.2111 | 0.2188 |
| Mean | 0.1633 | 0.2176 | 0.0544 | 0.0994 | 0.2284 | 0.2142 | 0.2199 |
| Quantile 75% | 0.2396 | 0.3080 | 0.0811 | 0.1333 | 0.3263 | 0.3016 | 0.3042 |
| Max | 0.6497 | 0.7218 | 0.3925 | 0.5954 | 0.7208 | 0.7573 | 0.6318 |
| Std | 0.1135 | 0.1333 | 0.0592 | 0.0949 | 0.1358 | 0.1221 | 0.1239 |
| | | | | $IdiVol_i$ | | | |
| Min | 0.0377 | 0.0347 | 0.0385 | 0.0313 | 0.0316 | 0.0311 | 0.0289 |
| Quantile 25% | 0.0878 | 0.0808 | 0.0885 | 0.0899 | 0.0787 | 0.0768 | 0.0749 |
| Median | 0.1200 | 0.1098 | 0.1219 | 0.1180 | 0.1073 | 0.1091 | 0.1048 |
| Mean | 0.1357 | 0.1260 | 0.1379 | 0.1333 | 0.1233 | 0.1247 | 0.1222 |
| Quantile 75% | 0.1651 | 0.1535 | 0.1680 | 0.1612 | 0.1506 | 0.1552 | 0.1537 |
| Max | 0.7620 | 0.7186 | 0.7430 | 0.6825 | 0.7141 | 0.6540 | 0.6430 |
| Std | 0.0718 | 0.0659 | 0.0713 | 0.0645 | 0.0652 | 0.0670 | 0.0660 |
| | | | | $SysRisk_i$ | | | |
| Min | 0.0111 | 0.0164 | 0.0088 | 0.0101 | 0.0167 | 0.0188 | 0.0190 |
| Quantile 25% | 0.0468 | 0.0541 | 0.0310 | 0.0365 | 0.0557 | 0.0544 | 0.0550 |
| Median | 0.0618 | 0.0727 | 0.0454 | 0.0565 | 0.0756 | 0.0728 | 0.0733 |
| Mean | 0.0666 | 0.0786 | 0.0548 | 0.0640 | 0.0817 | 0.0809 | 0.0817 |
| Quantile 75% | 0.0813 | 0.0964 | 0.0707 | 0.0805 | 0.1002 | 0.0989 | 0.0999 |
| Max | 0.3502 | 0.3626 | 0.3189 | 0.4540 | 0.3713 | 0.5322 | 0.5439 |
| Std | 0.0306 | 0.0361 | 0.0358 | 0.0386 | 0.0386 | 0.0419 | 0.0421 |

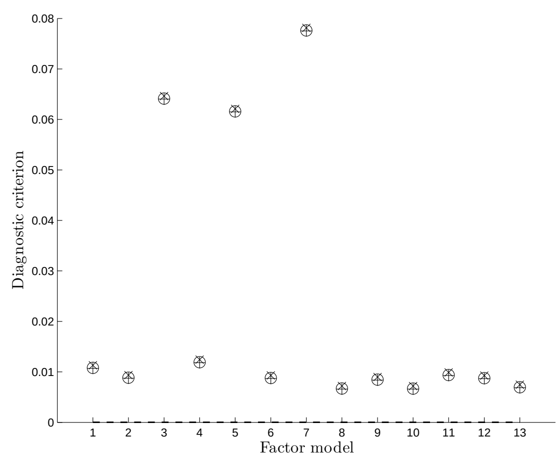
The table contains the descriptive statistics (cross-sectional minimum, 25% and 75% quantiles, median, mean, maximum and standard deviation) of the estimated coefficient of determination ($\hat{\rho}_i^2$), the estimated adjusted coefficients of determination ($\hat{\rho}_{ad,i}^2$), the idiosyncratic risks ($IdiVol_i$), and the systematic risks ($SysRisk_i$) for the time-varying linear factor models estimated by using the monthly 30-day T-bill and the dividend yields as common instruments.

**Figure 1: Values of the diagnostic criteria $\xi_1, \xi_2$ and $\xi_3$ for the time-invariant models**



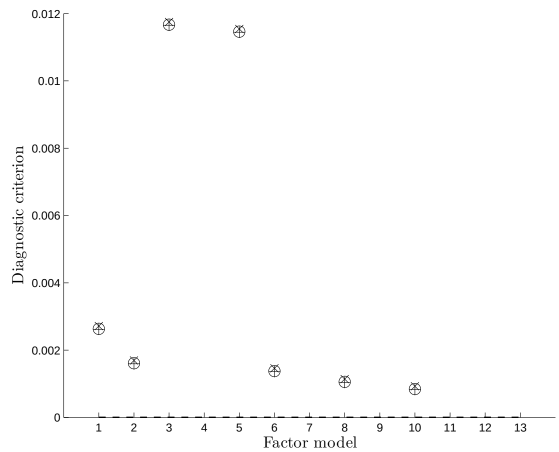

The figure plots the values of the diagnostic criteria $\xi_1$(red circle), $\xi_2$ (green plus sign) and $\xi_3$ (blue cross) for the time-invariant specifications. We also report the zero axis (red dashed horizontal line).

**Figure 2: Estimated values of the diagnostic criteria $\check{\xi}_1, \check{\xi}_2$ and $\check{\xi}_3$ for the time-invariant models**



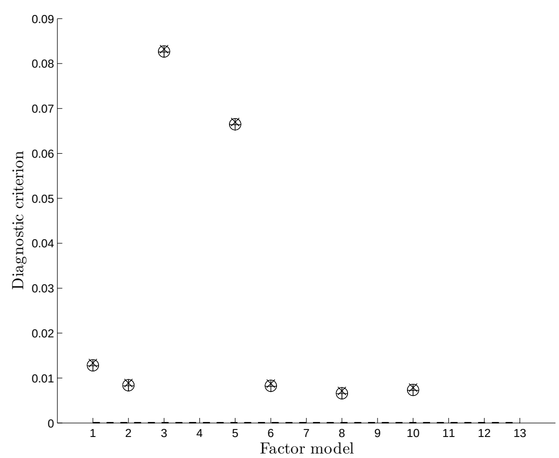

The figure plots the values of the logarithmic diagnostic criteria $\check{\xi}_1$(red circle), $\check{\xi}_2$ (green plus sign) and $\check{\xi}_3$ (blue cross) for the time-invariant specifications. We also report the zero axis (red dashed horizontal line).

**Figure 3: Values of the diagnostic criteria $\xi_1, \xi_2$ and $\xi_3$ for the time-varying models (i)**



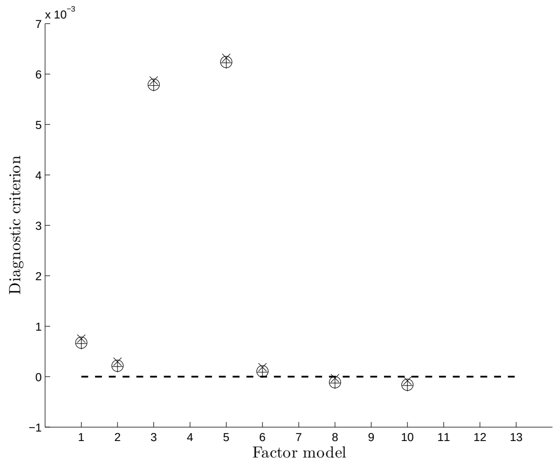

The figure plots the values of the diagnostic criteria $\xi_1$(red circle), $\xi_2$ (green plus sign) and $\xi_3$ (blue cross) for the time-varying specifications when $Z_t^*$ includes default and term spreads. The diagnostic criteria cannot be computed for the JW, CRR, (9), (11)-(13) models due to the multicollinearity problems. We also report the zero axis (red dashed horizontal line).

**Figure 4: Values of the diagnostic criteria $\check{\xi}_1, \check{\xi}_2$ and $\check{\xi}_3$ for the time-varying models (i)**



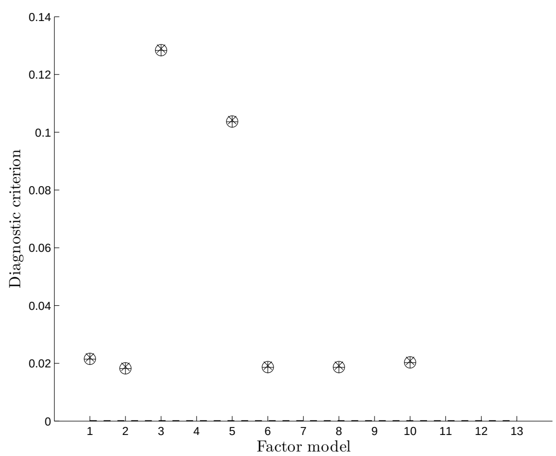

The figure plots the values of the logarithmic diagnostic criteria $\check{\xi}_1$(red circle), $\check{\xi}_2$ (green plus sign) and $\check{\xi}_3$ (blue cross) for the time-varying specifications when $Z_t^*$ includes default and term spreads. The logarithmic diagnostic criteria cannot be computed for the JW, CRR, (9), (11)-(13) models due to the multicollinearity problems. We also report the zero axis (red dashed horizontal line).

**Figure 5: Values of the diagnostic criteria $\xi_1, \xi_2$ and $\xi_3$ for the time-varying models (ii)**



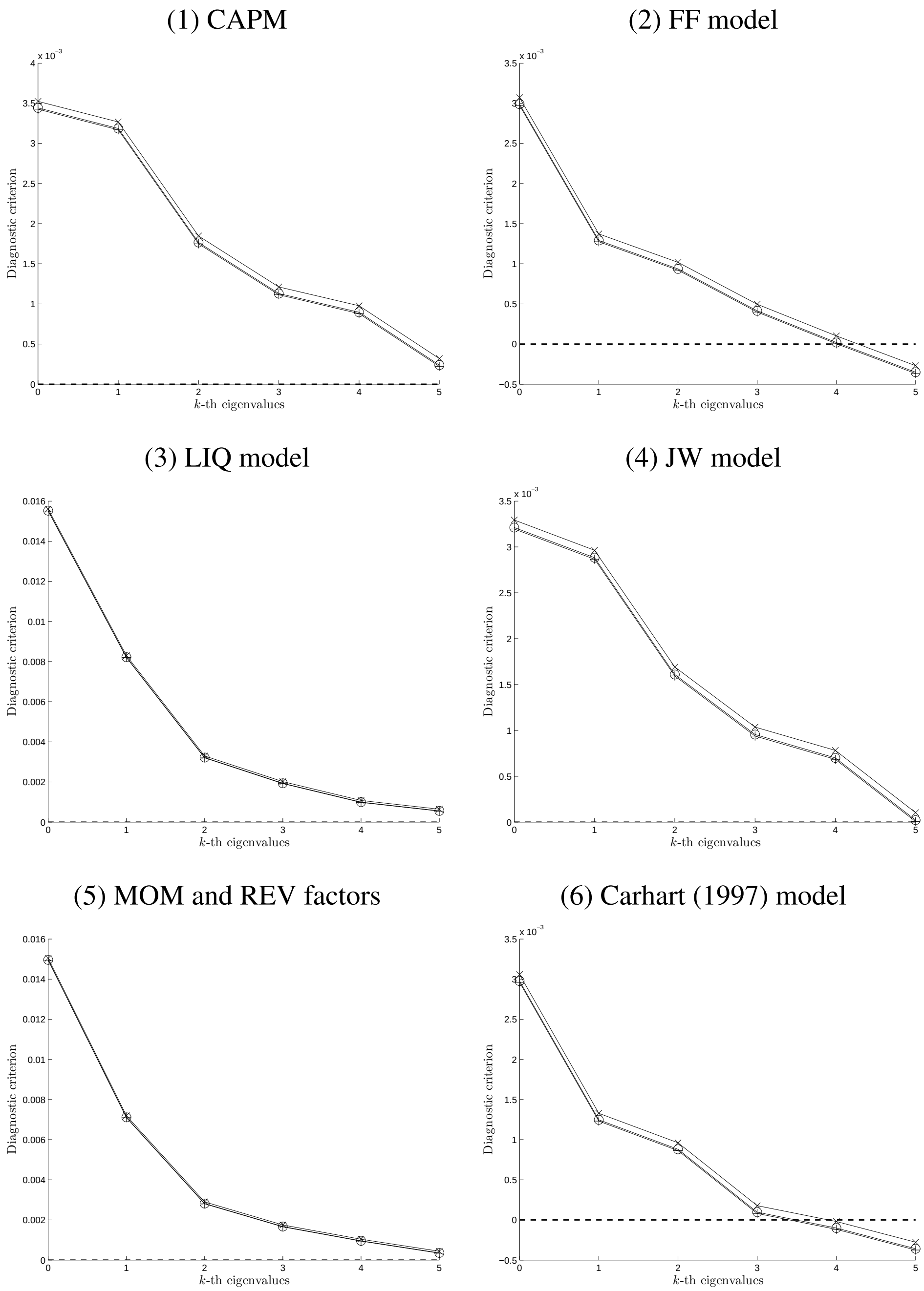

The figure plots the values of the diagnostic criteria $\xi_1$(red circle), $\xi_2$ (green plus sign) and $\xi_3$ (blue cross) for the time-varying specifications when $Z_t^*$ includes one-month T-Bill and dividend yield. The diagnostic criteria cannot be computed for the JW, CRR, (9), (11)-(13) models due to the multicollinearity problems. We also report the zero axis (red dashed horizontal line).

**Figure 6: Values of the diagnostic criteria $\check{\xi}_1, \check{\xi}_2$ and $\check{\xi}_3$ for the time-varying models (ii)**



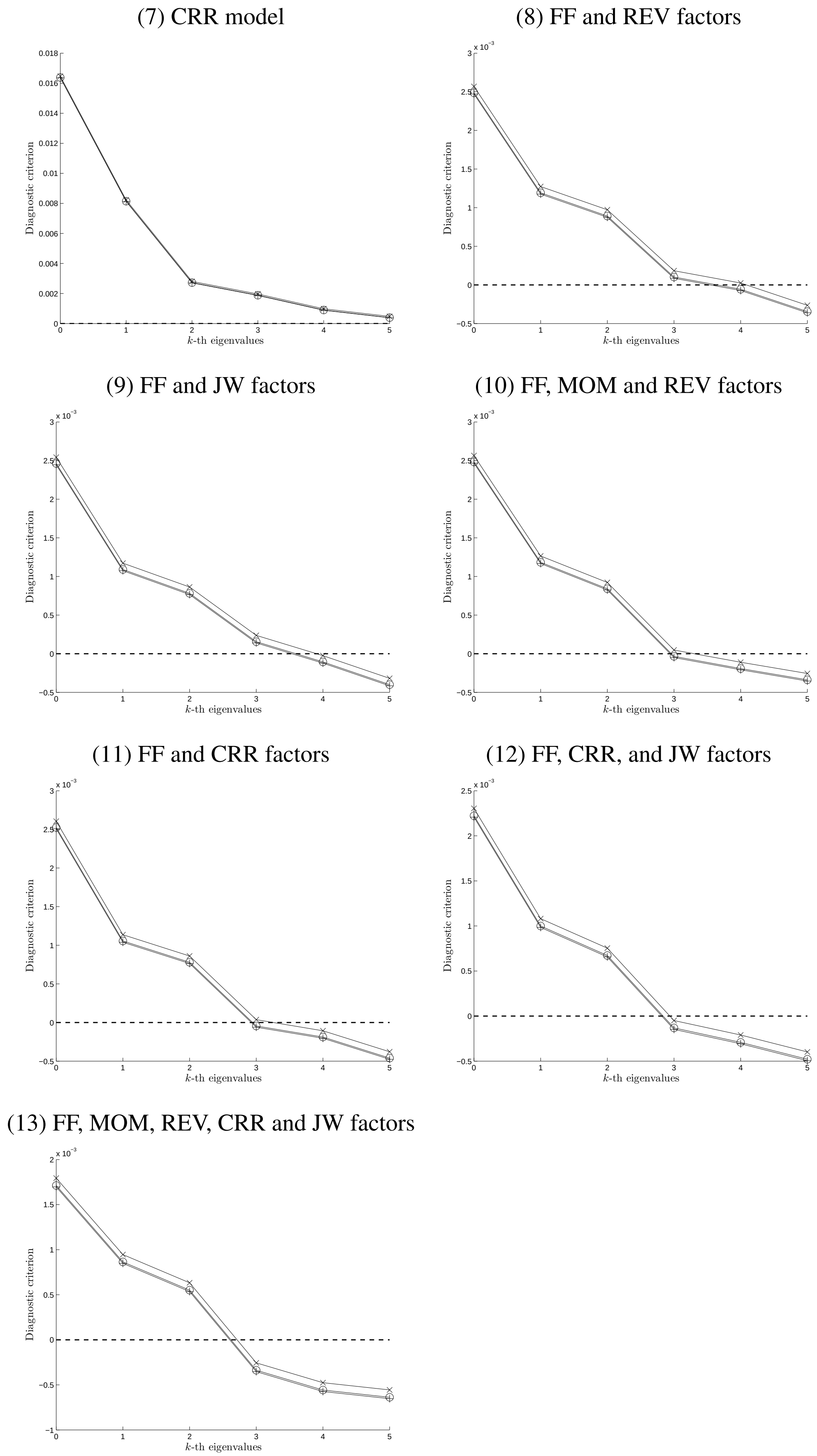

The figure plots the values of the logarithmic diagnostic criteria $\check{\xi}_1$(red circle), $\check{\xi}_2$ (green plus sign) and $\check{\xi}_3$ (blue cross) for the time-varying specifications when $Z_t^*$ includes one-month T-Bill and dividend yield. The logarithmic diagnostic criteria cannot be computed for the JW, CRR, (9), (11)-(13) models due to the multicollinearity problems. We also report the zero axis (red dashed horizontal line).

**Figure 7: Values of criteria $\xi(k)$ for the time-invariant models**



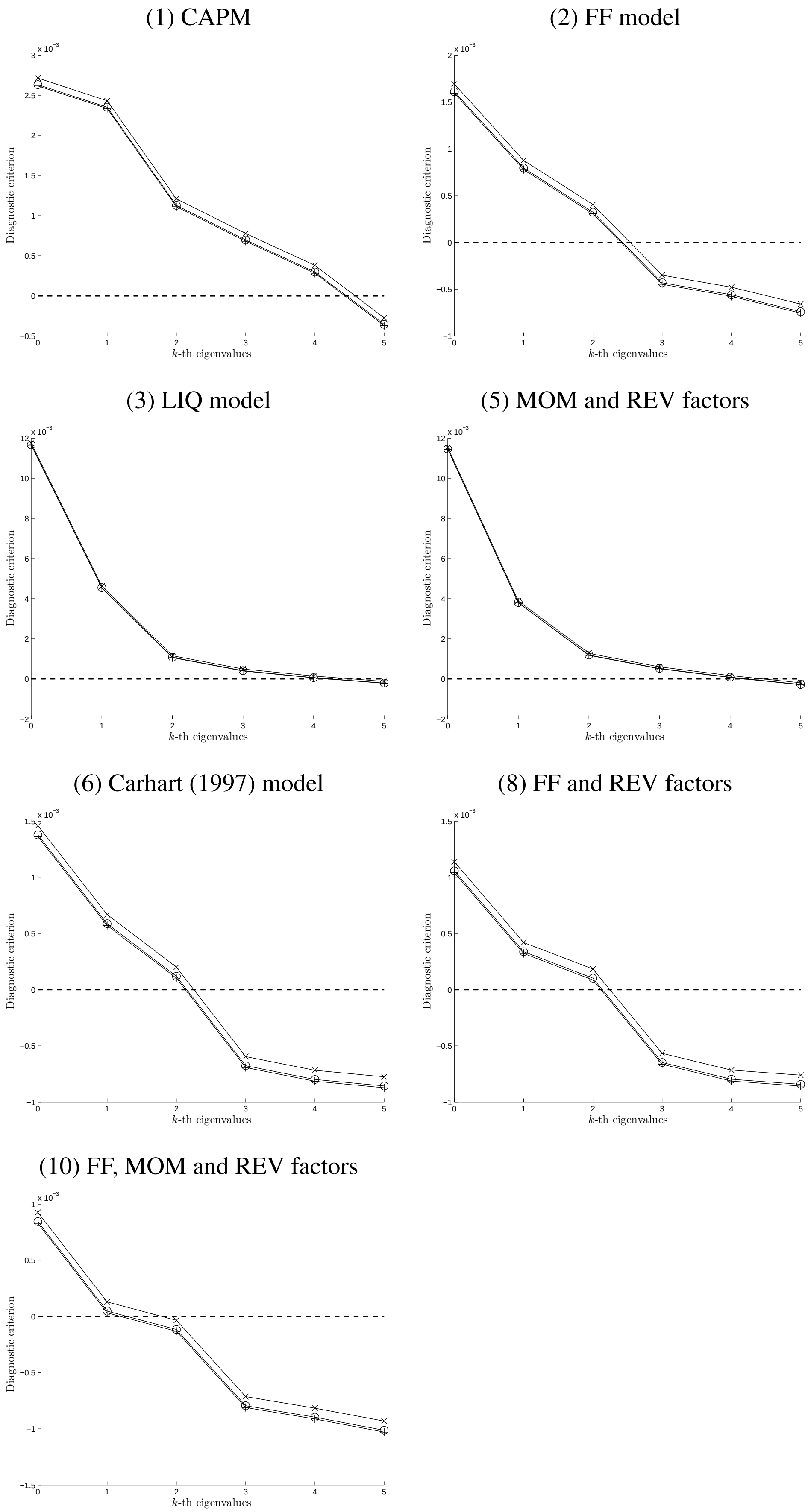

The figure plots the values of the diagnostic criteria $\xi_1(k)$(red circle), $\xi_2(k)$ (green plus sign) and $\xi_3(k)$ (blue cross) with $k = 1, ..., 5$, for the time-invariant specifications (1)-(6). We also report the zero axis (red dashed horizontal line).

**Figure 8: Values of criteria $\xi(k)$ for the time-invariant models**



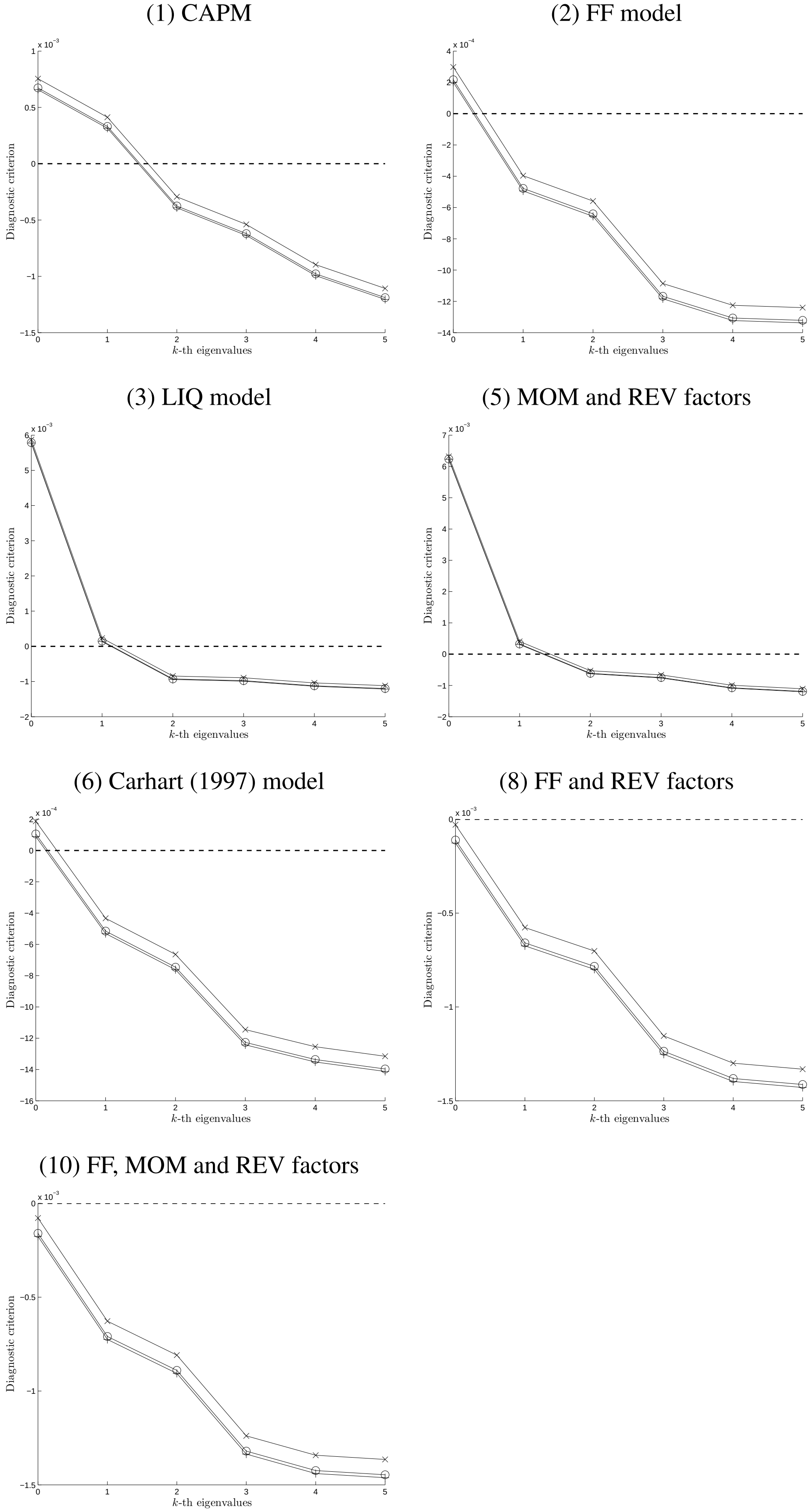

The figure plots the values of the diagnostic criteria $\xi_1(k)$(red circle), $\xi_2(k)$ (green plus sign) and $\xi_3(k)$ (blue cross) with $k = 1, ..., 5$, for the time-invariant specifications (7)-(13). We also report the zero axis (red dashed horizontal line).

**Figure 9: Values of criteria $\xi(k)$ for the time-varying models (i)**




The figure plots the values of the diagnostic criteria $\xi_1(k)$(red circle), $\xi_2(k)$ (green plus sign) and $\xi_3(k)$ (blue cross) with $k = 1, ..., 5$, for the time-varying specifications when $Z_t^*$ includes default and term spreads. We also report the zero axis (red dashed horizontal line).

**Figure 10: Values of criteria $\xi\left(k\right)$ for the time-varying models (ii)**




The figure plots the values of the diagnostic criteria $\xi_1\left(k\right)$(red circle), $\xi_2\left(k\right)$ (green plus sign) and $\xi_3\left(k\right)$ (blue cross) with $k = 1, ..., 5$, for the time-varying specifications when $Z_t^*$ includes one-month T-Bill and dividend yield. We also report the zero axis (red dashed horizontal line).

## References

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

S. C. Ahn, A. R. Horenstein, and N. Wang. Beta matrix and common factors in stock returns. *Working paper*, 2013.

N. Al-Najjar. Decomposition and characterization of risk with a continuum of random variables. *Econometrica*, 63(5):1195–1224, 1995.

N. Al-Najjar. Factor analysis and arbitrage pricing in large asset economies. *Journal of Economic Theory*, 78(2):231–262, 1998.

N. Al-Najjar. Decomposition and characterization of risk with a continuum of random variables: Corrigendum. *Econometrica*, 67(4):919–920, 1999a.

N. Al-Najjar. On the robustness of factor structures to asset repackaging. *Journal of Mathematical Economics*, 31(3):309–320, 1999b.

D. W. K. Andrews. Cross-section regression with common shocks. *Econometrica*, 73(5):1551–1585, 2005.

A. Ang, J. Liu, and K. Schwarz. Using individual stocks or portfolios in tests of factor models. 2008.

D. Avramov and T. Chordia. Asset pricing models and financial market anomalies. *The Review of Financial Studies*, 19(3):1000–1040, 2006.

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

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

J. Bai and S. Ng. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150, 2006.

D.A. Belsley, E. Kuh, and R.E. Welsch. *Regression diagnostics - Identifying influential data and sources of collinearity*. John Wiley & Sons, New York, 2004.

O. Boguth, M. Carlson, A. Fisher, and M. Simutin. Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. *Journal of Financial Economics*, 102 (2):363–389, 2011.

S. J. Brown, W. N. Goetzmann, and S. A. Ross. Survival. *The Journal of Finance*, 50(3):853–873, 1995.

J. Y. Campbell. Understanding risk and return. *NBER Working Papers*, (4554), 1993.

M. Caner and X. Han. Selecting the correct number of factors in approximate factor models: The large panel case with group bridge estimator. *Journal of Business and Economics Statistics*, forthcoming, 2014.

M. Carhart. On persistence of mutual fund performance. *Journal of Finance*, 52(1):57–82, 1997.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, 1983.

N. Chen, R. Roll, and S. Ross. Economic forces and the stock market. *Journal of Business*, 59:563–594, 1986.

J. H. Cochrane. A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy*, 104(3):572–621, 1996.

J. H. Cochrane. Presidential address: Discount rates. *Journal of Finance*, 66(4):1047–1108, 2011.

G. Connor and R. A. Korajczyk. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, 15(3):373–394, 1986.

G. Connor and R. A. Korajczyk. Estimating pervasive economic factors with missing observations. *Working Paper No. 34, Department of Finance, Northwestern University*, 1987.

G. Connor and R. A. Korajczyk. Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, 21:255–289, 1988.

G. Connor and R. A. Korajczyk. A test for the number of factors in an approximate factor model. *Journal of Finance*, 48(4):1263–1291, 1993.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

E. F. Fama and K. R. French. Dissecting anomalies. *Journal of Finance*, 63(4):1653–1678, 2008.

W. E. Ferson and C. R. Harvey. The variation of economic risk premiums. *Journal of Political Economy*, 99(2):385–415, 1991.

W. E. Ferson and C. R. Harvey. Conditioning variables and the cross section of stock returns. *Journal of Finance*, 54(4):1325–1360, 1999.

W. E. Ferson and R. W. Schadt. Measuring fund strategy and performance in changing economic conditions. *Journal of Finance*, 51(2):425–61, 1996.

P. Gagliardini, E. Ossola, and O. Scaillet. Time-varying risk premium in large cross-sectional equity datasets. *Working paper*, 2011.

E. Ghysels. On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *Journal of Finance*, 53(2):549–573, 1998.

N. Gospodinov, R. Kan, and C. Robotti. Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. *Review of Financial Studies*, 27(7):2139–2170, 2014.

W. Greene. *Econometric Analysis*. Prentice Hall, 6th Edition, 2008.

L. P. Hansen and S. F. Richard. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica*, 55(3):587–613, 1987.

C. R. Harvey, Y. Liu, and H. Zhu. ... and the cross-section of expected returns. *Working Paper*, 2013.

R. Horn and C. Johnson. *Matrix analysis*. Cambridge University Press, 1985.

C. Hsiao. *Analysis of Panel Data*. Econometric Society Monographs, 2nd Edition, Cambridge University Press, 2003.

R. Jagannathan and Z. Wang. The conditional CAPM and the cross-section of expected returns. *Journal of Finance*, 51(1):3–53, 1996.

R. Kan and C. Zhang. Two-pass tests of asset pricing models with useless factions. *Journal of Finance*, 54 (1):203–235, 1999a.

R. Kan and C. Zhang. GMM tests of stochastic discount factor models with useless factors. *Journal of Financial Economics*, 54(1):103–127, 1999b.

G. Kapetanios. A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business and Economic Statistics*, 28(3):397–409, 2010.

F. Kleibergen. Test of risk premia in linear factor models. *Journal of econometrics*, 149(2):149–173, 2009.

B. N. Lehmann and D. M. Modest. The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics*, 21:213–254, 1988.

M. Lettau and S. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3):815–849, 2001.

J. Lewellen and S. Nagel. The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82(2):289–314, 2006.

J. R. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, 2007.

C. D. Meyer. *Matrix Analysis and Applied Linear Algebra Book*. Siam, 2000.

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

A. Onatski. Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics*, 92(4):1004–1016, 2010.

A. Onatski. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258, 2013.

L. Pastor and R.F. Stambaugh. Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics*, 63:315–349, 2002.

D. Pena and P. Poncela. *Advances in distribution theory, order statistics, and inference*, chapter Dimension reduction in multivariate time series, pages 433–458. Birkhauser, 2006.

R. Petkova and L. Zhang. Is value riskier than growth? *Journal of Financial Economics*, 78(1):187–202, 2005.

S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976.

J. Shanken. Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, 45(1-2): 99–120, 1990.

J. H. Stock and M. W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business* *and Economic Statistics*, 20(2):147–62, 2002b.

B. Wang and F. Zhang. Some inequalities for the eigenvalues of the product of positive semidefinite hermitian matrices. *Linear algebra and its applications*, 160:113–118, 1992.

# Appendix 1 Regularity conditions

In this Appendix, we list and comment additional assumptions used to derive the proofs in Appendix 2.

**Assumption A.1** *There exists a constant $M > 0$ such that, for all $n, T \in \mathbb{N}$, we have:*
*a)* $\frac{1}{n^2 T} \sum_{i,j} \sum_{t_1,t_2,t_3} \left| E\left[\varepsilon_{i,t_1}\varepsilon_{j,t_2}\varepsilon_{i,t_3}\varepsilon_{j,t_3} \left| x_{i,\underline{T}}, x_{j,\underline{T}}, \gamma_i, \gamma_j \right.\right]\right| \leq M$;
*b)* $\frac{1}{n^2 T^2} \sum_{i,j} \sum_{t_1,t_2,t_3,t_4} \left| E\left[\varepsilon_{i,t_1}\varepsilon_{i,t_2}\varepsilon_{j,t_3}\varepsilon_{j,t_4} \left| x_{i,\underline{T}}, x_{j,\underline{T}}, \gamma_i, \gamma_j \right.\right]\right| \leq M$.

**Assumption A.2** *The error terms $(\varepsilon_{i,t})$ are $\varepsilon_{i,t} = u_{i,t}$ under model $\mathcal{M}_1$, and $\varepsilon_{i,t} = \theta_i' h_t + u_{i,t}$ under model $\mathcal{M}_2$, where the $(u_{i,t})$ are such that for a constant $M > 0$ and for all $n, T \in \mathbb{N}$ we have:*
*a)* $\frac{1}{nT} \sum_{i,j} \sum_{t} \left| E\left[u_{i,t}u_{j,t} \left| h_{\underline{T}}, \gamma_i, \gamma_j \right.\right]\right| \leq M$;
*b)* $\frac{1}{nT} \sum_{i,j} \sum_{t_1,t_2} \left| E\left[u_{i,t_1}u_{j,t_2} \left| h_{\underline{T}}, x_{i,\underline{T}}, x_{j,\underline{T}}, \gamma_i, \gamma_j \right.\right]\right| \leq M$;
*c)* $\frac{1}{n^2 T} \sum_{i,j} \sum_{t_1,t_2} \left| E\left[u_{i,t_1}u_{j,t_1}u_{i,t_2}u_{j,t_2}\right]\right| \leq M$;
*d)* $\frac{1}{n^2 T} \sum_{i,j} \sum_{t_1,t_2,t_3} \left| E\left[u_{i,t_1}u_{j,t_2}u_{i,t_3}u_{j,t_3} \left| h_{\underline{T}}, x_{i,\underline{T}}, x_{j,\underline{T}}, \gamma_i, \gamma_j \right.\right]\right| \leq M$;
*e)* $\frac{1}{n^2 T^2} \sum_{i,j} \sum_{t_1,t_2,t_3,t_4} \left| E\left[u_{i,t_1}u_{i,t_2}u_{j,t_3}u_{j,t_4} \left| h_{\underline{T}}, x_{i,\underline{T}}, x_{j,\underline{T}}, \gamma_i, \gamma_j \right.\right]\right| \leq M$.
*f)* $E\left[\varepsilon_{i,t}^4\right] \leq M$, *for all $i \leq n$ and $t \leq T$.*

**Assumption A.3** *There exists a constant $M > 0$ such that $\|x_{i,t}\| \leq M$, $P$-a.s., for any $i$ and $t$.*

**Assumption A.4** *a) There exists a constant $M > 0$ such that $\|h_t\| \leq M$, $P$-a.s., for all $t$. Moreover, b) $\|\theta_i\| < M$, for all $i$.*

**Assumption A.5** *The trimming constants $\chi_{1,T}$ and $\chi_{2,T}$ are such that $\chi_{1,T}^2 \chi_{2,T} = o\left(Tg\left(n,T\right)\right)$.*

**Assumption A.6** *Under model $\mathcal{M}_1$, $\tilde{\mathcal{E}} = (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_i, \ldots, \tilde{\varepsilon}_N) = E_{1T}^{1/2} E_2 E_{3N}^{1/2}$, where $\tilde{\varepsilon}_i = I_i \odot \varepsilon_i$, and under model $\mathcal{M}_2$, $\tilde{\mathcal{U}} = (\tilde{u}_1, \ldots, \tilde{u}_i, \ldots, \tilde{u}_N)$, where $\tilde{u}_i = I_i \odot u_i$, such that (i) $E_2' = (e_{it})$, and $E_{1T}^{1/2}$ and $E_{3N}^{1/2}$ are the symmetric square roots of $T \times T$ and $N \times N$ positive semidefinite matrices $E_{1T}$ and $E_{3N}$, respectively, (ii) the $e_{it}$ are independent and identically distributed (i.i.d.) random variables with zero mean* *and uniformly bounded moments up to the fourth order, (iii) $\mu_1(E_{1T})$ and $\mu_1(E_{3N})$ are bounded from above uniformly in $T$ and $N$, respectively.*

Assumption A.6 is the same as in Ahn and Horenstein (2013) (see also Onatski (2010))). It allows for correlation and heteroskedasticity in both dimensions of $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{U}}$. Bai and Ng (2002b) show that those assumptions of weak cross-section and serial correlations ensure that $\mu_1\left(\frac{1}{nT}\sum_i \tilde{\varepsilon}_i\tilde{\varepsilon}_i'\right) = O_p\left(C_{nT}^{-2}\right)$ under $\mathcal{M}_1$, and $\mu_1\left(\frac{1}{nT}\sum_i \tilde{u}_i\tilde{u}_i'\right) = O_p\left(C_{nT}^{-2}\right)$ under $\mathcal{M}_2$.

# Appendix 2 Proofs

### A.2.1 Derivation of Equation (7)

We have:

$$\begin{aligned}\mu_1\left(\frac{1}{nT}\sum_i M_X H\theta_i\theta_i' H' M_X\right) &= \mu_1\left(\frac{1}{T}M_X H\left(\frac{1}{n}\sum_i \theta_i\theta_i'\right)H' M_X\right)\\ &= \mu_1\left(\frac{1}{T}M_X H\left(\frac{\Theta'\Theta}{n}\right)H' M_X\right)\\ &\geq \mu_m\left(\frac{\Theta'\Theta}{n}\right)\mu_1\left(\frac{1}{T}M_X H H' M_X\right).\end{aligned}$$

Now, we use that matrices $AA'$ and $A'A$ share the same non-zero eigenvalues, and $M_X$ is idempotent. Thus,

$$\mu_1\left(\frac{1}{T}M_X H H' M_X\right) = \mu_1\left(\frac{H' M_X H}{T}\right).$$

### A.2.2 Proof of Proposition Proposition 1

**a)** The OLS estimator of $\beta_i$ in matrix notation is $\hat{\beta}_i = \left(\tilde{X}_i'\tilde{X}_i\right)^{-1}\tilde{X}_i'\tilde{R}_i$, with $\tilde{R}_i = \boldsymbol{I}_i \odot R_i$. We get the vector of residuals $\hat{\varepsilon}_i = R_i - X_i\left(\tilde{X}_i'\tilde{X}_i\right)^{-1}\tilde{X}_i'\tilde{R}_i$. Then, we have $\bar{\varepsilon}_i = \boldsymbol{I}_i \odot \hat{\varepsilon}_i = M_{\tilde{X}_i}\tilde{R}_i = M_{\tilde{X}_i}\tilde{\varepsilon}_i$, where $\tilde{\varepsilon}_i = \boldsymbol{I}_i \odot \varepsilon_i$ and $M_{\tilde{X}_i} = I_T - P_{\tilde{X}_i}$, with $P_{\tilde{X}_i} = \tilde{X}_i\left(\tilde{X}_i'\tilde{X}_i\right)^{-1}\tilde{X}_i'$. Let us decompose $\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'$ as:

$$\begin{aligned}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i' &= \sum_i \tilde{\varepsilon}_i\tilde{\varepsilon}_i' + \sum_i\left(\mathbf{1}_i^{\chi} - 1\right)\tilde{\varepsilon}_i\tilde{\varepsilon}_i'\\ &\quad -\sum_i \mathbf{1}_i^{\chi}\left(P_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i' + \tilde{\varepsilon}_i\tilde{\varepsilon}_i' P_{\tilde{X}_i} - P_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i' P_{\tilde{X}_i}\right).\end{aligned}$$

The second term in the r.h.s. is a negative semi-definite matrix. Hence:

$$\mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) \leq \mu_1\left(\frac{1}{nT}\sum_i \tilde{\varepsilon}_i\tilde{\varepsilon}_i' - \frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\left(P_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i' + \tilde{\varepsilon}_i\tilde{\varepsilon}_i'P_{\tilde{X}_i} - P_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i'P_{\tilde{X}_i}\right)\right).$$

Let us now use that the largest eigenvalue $\mu_1(A)$ of a symmetric matrix $A$ corresponds to its operator norm $\|A\|_{op} = \sup_{x:\|x\|=1} \|Ax\|$ , and that the operator norm (as any matrix norm) is such that $\|A+B\|_{op} \leq \|A\|_{op} + \|B\|_{op}$ for two matrices $A$ and $B$. We have

$$\begin{aligned}
\mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) &\leq \left\|\frac{1}{nT}\sum_i \tilde{\varepsilon}_i\tilde{\varepsilon}_i'\right\|_{op} + \left\|\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}P_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i'\right\|_{op} \\
&\quad + \left\|\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\tilde{\varepsilon}_i\tilde{\varepsilon}_i'P_{\tilde{X}_i}\right\|_{op} + \left\|\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}P_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i'P_{\tilde{X}_i}\right\|_{op} \\
&\leq \mu_1\left(\frac{1}{nT}\sum_i \tilde{\varepsilon}_i\tilde{\varepsilon}_i'\right) + \|I_1\| + \|I_2\| + \|I_3\|, \qquad (11)
\end{aligned}$$

where $\|\cdot\|$ is the Frobenius norm and it is such that $\|A\|_{op} \leq \|A\|$, for any $T\times T$ matrix $A$ (see, for example, Meyer (2000)). To control terms $\|I_j\|$, $j = 1,2,3$ in the r.h.s. of (11) we use the next Lemma, which is proved in Section A.2.2.

**Lemma 1** *Under Assumptions A.1, A.3 and A.5, (i)* $\|I_1\| = O_p\left(\frac{\chi_{1,T}^2\chi_{2,T}}{T}\right)$*, (ii)* $\|I_2\| = O_p\left(\frac{\chi_{1,T}^2\chi_{2,T}}{T}\right)$ *and (iii)* $\|I_3\| = o_p\left(\frac{1}{T}\right)$*, when* $n,T\to\infty$.

From Equation (11) and Lemma 1, we get $\mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) \leq \mu_1\left(\frac{1}{nT}\sum_i \tilde{\varepsilon}_i\tilde{\varepsilon}_i'\right) + O_p\left(\frac{\chi_{1,T}^2\chi_{2,T}}{T}\right)$. Under model $\mathcal{M}_1$, we have $\mu_1\left(\frac{1}{nT}\sum_i \tilde{\varepsilon}_i\tilde{\varepsilon}_i'\right) = O_p\left(C_{nT}^{-2}\right)$, if the $\tilde{\varepsilon}_{i,t} = I_{i,t}\varepsilon_{i,t}$ satisfy the weak time-series and cross-sectional correlation assumptions in Bai and Ng (2002b) and Ahn and Horenstein (2013). Then,

$$\xi = O_p\left(C_{nT}^{-2}\right) + O_p\left(\frac{\chi_{1,T}^2\chi_{2,T}}{T}\right) - g(n,T).$$

Since $g(n,T)\,C_{nT}^2 \to \infty$ and using Assumption A.5, we get $\xi = -g(n,T)(1+o(1))$. Proposition Proposition 1a) follows.

**b)** Under model $\mathcal{M}_2$, by definition of $\bar{\varepsilon}_i$, we have:

$$\sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' = \sum_i M_{\tilde{X}_i} \tilde{\varepsilon}_i \tilde{\varepsilon}_i' M_{\tilde{X}_i} + \sum_i \left(\mathbf{1}_i^{\chi} - 1\right) M_{\tilde{X}_i} \tilde{\varepsilon}_i \tilde{\varepsilon}_i' M_{\tilde{X}_i},$$

where $\tilde{\varepsilon}_i = \tilde{H}_i \theta_i + \tilde{u}_i$, with $\tilde{u}_i = \boldsymbol{I}_i \odot u_i$. By using $\|A+B\|_{op} \geq \|A\|_{op} - \|B\|_{op}$ and $\|A\|_{op} \leq \|A\|$, we have

$$\begin{aligned}
\mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) \geq\ & \left\|\frac{1}{nT}\sum_i M_{\tilde{X}_i}\tilde{H}_i\theta_i\theta_i'\tilde{H}_i' M_{\tilde{X}_i}\right\|_{op} - \left(\left\|\frac{1}{nT}\sum_i \left(1-\mathbf{1}_i^{\chi}\right) M_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i' M_{\tilde{X}_i}\right\|_{op}\right. \\
& + \left\|\frac{1}{nT}\sum_i M_{\tilde{X}_i}\tilde{H}_i\theta_i\tilde{u}_i' M_{\tilde{X}_i}\right\|_{op} + \left\|\frac{1}{nT}\sum_i M_{\tilde{X}_i}\tilde{u}_i\theta_i'\tilde{H}_i' M_{\tilde{X}_i}\right\|_{op} \\
& \left. + \left\|\frac{1}{nT}\sum_i M_{\tilde{X}_i}\tilde{u}_i\tilde{u}_i' M_{\tilde{X}_i}\right\|_{op}\right) \\
\geq\ & \mu_1\left(\frac{1}{nT}\sum_i M_{\tilde{X}_i}\tilde{H}_i\theta_i\theta_i'\tilde{H}_i' M_{\tilde{X}_i}\right) - \left(\|I_4\| + \|I_5\| + \|I_6\| + \|I_7\|_{op}\right). \quad (12)
\end{aligned}$$

Proposition Proposition 1(b) follows from the next Lemma, which is proved in Section A.2.4.

**Lemma 2** *Under Assumptions A.2-A.5, (i)* $\|I_4\| = O_p(T^{-\bar{b}})$*, for any* $\bar{b} > 0.$*, (ii)* $\|I_5\| = O_p\left(\frac{1}{\sqrt{n}}\right)$*, (iii)* $\|I_6\| = O_p\left(\frac{1}{\sqrt{n}}\right)$ *and (iv)* $\|I_7\|_{op} = O_p\left(C_{n,T}^{-2} + \frac{\chi_{1,T}^4\chi_{2,T}^2}{T}\right)$*, when* $n, T \to \infty$*.*

From Equation (12) and Lemma 2, we get $\mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) \geq \mu_1\left(\frac{1}{nT}\sum_i M_{\tilde{X}_i}\tilde{H}_i\theta_i\theta_i'\tilde{H}_i' M_{\tilde{X}_i}\right)$ $+o_p(1)$. Under Assumption Assumption 1, the result follows.

### A.2.3 Proof of Lemma 1

*Term depending on $I_1$.* Let us define the information sets $\mathcal{J}_i = \left\{ x_{i,\underline{T}}, I_{i,\underline{T}}, \gamma_i \right\}$ for asset $i$ and $\mathcal{J} = \bigcup_{i=1}^{n} \mathcal{J}_i$. By using the properties of the trace operator, we have:

$$
\begin{aligned}
E\left[ \|I_1\|^2 \,|\mathcal{J} \right] &= E\left[ \frac{1}{n^2T^2} Tr\left[ \sum_{i,j} \mathbf{1}_i^{\chi} \mathbf{1}_j^{\chi} P_{\tilde{X}_i} \tilde{\varepsilon}_i \tilde{\varepsilon}_i' \tilde{\varepsilon}_j \tilde{\varepsilon}_j' P_{\tilde{X}_j} \right] |\mathcal{J} \right] \\
&= E\left[ \frac{1}{n^2T^2} \sum_{i,j} \mathbf{1}_i^{\chi} \mathbf{1}_j^{\chi} \tilde{\varepsilon}_j' P_{\tilde{X}_j} P_{\tilde{X}_i} \tilde{\varepsilon}_i \tilde{\varepsilon}_i' \tilde{\varepsilon}_j \,|\mathcal{J} \right] \\
&= \frac{1}{n^2T^2} \sum_{i,j} \mathbf{1}_i^{\chi} \mathbf{1}_j^{\chi} E\left[ \tilde{\varepsilon}_j' P_{\tilde{X}_j} P_{\tilde{X}_i} \tilde{\varepsilon}_i \left( \tilde{\varepsilon}_i' \tilde{\varepsilon}_j \right) |\mathcal{J}_i, \mathcal{J}_j \right].
\end{aligned}
$$

To ease notation $x_{i,t}$ is now treated as a scalar. By using that $\tilde{\varepsilon}_i' \tilde{\varepsilon}_j = \sum_t I_{ij,t} \varepsilon_{i,t} \varepsilon_{j,t}$ where $I_{ij,t} = I_{i,t} I_{j,t}$ for $i, j = 1, ..., n$, and $\tilde{\varepsilon}_j' P_{\tilde{X}_j} P_{\tilde{X}_i} \tilde{\varepsilon}_i = \hat{Q}_{x,i}^{-1} \hat{Q}_{x,j}^{-1} \left( \sum_t I_{ij,t} x_{i,t} x_{j,t} \right) \left( \frac{1}{T_i} \sum_t I_{i,t} x_{i,t} \varepsilon_{i,t} \right) \left( \frac{1}{T_j} \sum_t I_{j,t} x_{j,t} \varepsilon_{j,t} \right)$, we get

$$
\begin{aligned}
E\left[ \|I_1\|^2 \,|\mathcal{J} \right] \leq\; & \frac{1}{n^2T^3} \sum_{i,j} \sum_{t_1,t_2,t_3} \mathbf{1}_i^{\chi} \mathbf{1}_j^{\chi} \tau_{i,T} \tau_{j,T} I_{i,t_1} I_{j,t_2} I_{ij,t_3} \hat{Q}_{x,i}^{-1} \hat{Q}_{x,j}^{-1} \\
& \left( \frac{1}{T} \sum_t I_{ij,t} x_{i,t} x_{j,t} \right) x_{i,t_1} x_{j,t_2} E\left[ \varepsilon_{i,t_1} \varepsilon_{j,t_2} \varepsilon_{i,t_3} \varepsilon_{j,t_3} \,|\mathcal{J}_i, \mathcal{J}_j \right].
\end{aligned}
$$

From the definition of condition number $CN$, $\|\hat{Q}_{x,i}^{-1}\|^2 = Tr\left( \hat{Q}_{x,i}^{-2} \right) = \sum_{k=1}^{d} \mu_{k,i}^{-2} \leq dCN\left( \hat{Q}_{x,i} \right)^4$, where the $\mu_{k,i}$ are the eigenvalues of matrix $\hat{Q}_{x,i}$, we use that $\mu_{1,i}\left( \hat{Q}_{x,i} \right) \geq 1$, which implies $\mathbf{1}_i^{\chi} \|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}^2$. Moreover, by $\mathbf{1}_i^{\chi} \tau_{i,T} \leq \chi_{2,T}$ and Assumption A.3, we deduce that

$$
E\left[ \|I_1\|^2 \,|\mathcal{J} \right] \leq \frac{1}{n^2T^3} C\chi_{1,T}^4 \chi_{2,T}^2 \sum_{i,j} \sum_{t_1,t_2,t_3} \left| E\left[ \varepsilon_{i,t_1} \varepsilon_{j,t_2} \varepsilon_{i,t_3} \varepsilon_{j,t_3} \left| x_{i,\underline{T}}, x_{j,\underline{T}}, \gamma_i, \gamma_j \right. \right] \right|.
$$

From Assumption A.1a), we have $E\left[ \|I_1\|^2 \,|\mathcal{J} \right] \leq \frac{1}{T^2} C\chi_{1,T}^4 \chi_{2,T}^2$, which implies $\|I_1\| = O_p\left( \frac{\chi_{1,T}^2 \chi_{2,T}}{T} \right)$.

*Term depending on $I_2$.* The result follows by similar arguments used to prove that $\|I_1\| = O_p\left( \frac{\chi_{1,T}^2 \chi_{2,T}}{T} \right)$.

*Term depending on $I_3$*. By the properties of trace operator, we have

$$\begin{aligned} E\left[\|I_3\|^2 \,|\mathcal{J}\right] &= E\left[\frac{1}{n^2T^2}Tr\left[\sum_{i,j}\mathbf{1}_i^{\chi}\mathbf{1}_j^{\chi}P_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i'P_{\tilde{X}_i}P_{\tilde{X}_j}\tilde{\varepsilon}_j\tilde{\varepsilon}_j'P_{\tilde{X}_j}\right]|\mathcal{J}\right] \\ &= E\left[\frac{1}{n^2T^2}\sum_{i,j}\mathbf{1}_i^{\chi}\mathbf{1}_j^{\chi}\left(\tilde{\varepsilon}_j'P_{\tilde{X}_j}P_{\tilde{X}_i}\tilde{\varepsilon}_i\right)^2|\mathcal{J}\right] \\ &= \frac{1}{n^2T^2}\sum_{i,j}\mathbf{1}_i^{\chi}\mathbf{1}_j^{\chi}E\left[\left(\tilde{\varepsilon}_j'P_{\tilde{X}_j}P_{\tilde{X}_i}\tilde{\varepsilon}_i\right)^2|\mathcal{J}_i,\mathcal{J}_j\right]. \end{aligned}$$

By using that $\tilde{\varepsilon}_j'P_{\tilde{X}_j}P_{\tilde{X}_i}\tilde{\varepsilon}_i = \hat{Q}_{x,i}^{-1}\hat{Q}_{x,j}^{-1}\left(\sum_t I_{ij,t}x_{i,t}x_{j,t}\right)\left(\frac{1}{T_i}\sum_t I_{i,t}x_{i,t}\varepsilon_{i,t}\right)\left(\frac{1}{T_j}\sum_t I_{j,t}x_{j,t}\varepsilon_{j,t}\right)$, we get

$$\begin{aligned} E\left[\|I_3\|^2 \,|\mathcal{J}\right] &\leq \frac{1}{n^2T^4}\sum_{i,j}\sum_{t_1,t_2,t_3,t_4}\mathbf{1}_i^{\chi}\mathbf{1}_j^{\chi}\tau_{i,T}^2\tau_{j,T}^2I_{i,t_1}I_{i,t_2}I_{j,t_3}I_{j,t_4}\hat{Q}_{x,i}^{-2}\hat{Q}_{x,j}^{-2} \\ &\quad \left(\frac{1}{T}\sum_t I_{ij,t}x_{i,t}x_{j,t}\right)^2 x_{i,t_1}x_{i,t_2}x_{j,t_3}x_{j,t_4}E\left[\varepsilon_{i,t_1}\varepsilon_{i,t_2}\varepsilon_{j,t_3}\varepsilon_{j,t_4}\,|\mathcal{J}_i,\mathcal{J}_j\right]. \end{aligned}$$

From $\mathbf{1}_i^{\chi}\|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}^2$, $\mathbf{1}_i^{\chi}\tau_{i,T} \leq \chi_{2,T}$ and Assumption A.3, we deduce that

$$E\left[\|I_3\|^2 \,|\mathcal{J}_i,\mathcal{J}_j\right] = \frac{1}{n^2T^4}C\chi_{1,T}^8\chi_{2,T}^4\sum_{i,j}\sum_{t_1,t_2,t_3,t_4}\left|E\left[\varepsilon_{i,t_1}\varepsilon_{i,t_2}\varepsilon_{j,t_3}\varepsilon_{j,t_4}\left|x_{i,\underline{T}},x_{j,\underline{T}},\gamma_i,\gamma_j\right.\right]\right|.$$

Then, by Assumption A.1b) the result follows.

### A.2.3 Proof of Lemma Lemma 2

*Term depending on $I_4$*. We have:

$$\|I_4\|_{op} \leq \frac{1}{nT}\sum_i(1-\mathbf{1}_i^{\chi})\left\|M_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i'M_{\tilde{X}_i}\right\|_{op} \leq \frac{1}{nT}\sum_i(1-\mathbf{1}_i^{\chi})\left\|\tilde{\varepsilon}_i\tilde{\varepsilon}_i'\right\|_{op},$$

where the second inequality is because matrix $M_{\tilde{X}_i}$ is idempotent. Moreover, by using $\left\|\tilde{\varepsilon}_i\tilde{\varepsilon}_i'\right\|_{op} \leq \left\|\tilde{\varepsilon}_i\tilde{\varepsilon}_i'\right\| = \tilde{\varepsilon}_i'\tilde{\varepsilon}_i$, we get:

$$\|I_4\|_{op} \leq \frac{1}{n}\sum_i(1-\mathbf{1}_i^{\chi})\left(\frac{1}{T}\sum_t I_{i,t}\varepsilon_{i,t}^2\right).$$

By the Cauchy-Schwarz inequality we have:

$$E\left[\|I_4\|_{op}\right] \leq \frac{1}{n}\sum_i P\left[\mathbf{1}_i^{\chi}=0\right]^{1/2}\left(E\left[\left(\frac{1}{T}\sum_t I_{i,t}\varepsilon_{i,t}^2\right)^2\right]\right)^{1/2}.$$

Lemma 7 of GOS states that $P\left[\mathbf{1}_i^{\chi}=0\right]=O(T^{-\bar{b}})$, for any $\bar{b}>0$. This unconditional probability is independent of $i$ since the indices $(\gamma_i)$ are i.i.d.. From Assumption A.2 f), we have $E\left[\left(\frac{1}{T}\sum_t I_{i,t}\varepsilon_{i,t}^2\right)^2\right] \leq M$ for all $i$ and a constant $M$. Thus, $E\left[\|I_4\|_{op}\right]=O(T^{-\bar{b}})$, for any $\bar{b}>0$.

*Term depending on* $I_5$. By the definition of $M_{\tilde{X}_i}$, we have $M_{\tilde{X}_i}\tilde{H}_i\theta_i\tilde{u}_i' M_{\tilde{X}_i} = \tilde{H}\theta_i\tilde{u}_i' - \tilde{H}_i\theta_i\tilde{u}_i' P_{\tilde{X}_i} - P_{\tilde{X}_i}\tilde{H}_i\theta_i\tilde{u}_i' + P_{\tilde{X}_i}\tilde{H}_i\theta_i\tilde{u}_i' P_{\tilde{X}_i}$. From the properties of the norm operator, we deduce

$$\begin{aligned}\|I_5\| &\leq \left\|\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\tilde{u}_i'\right\| + \left\|\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\tilde{u}_i' P_{\tilde{X}_i}\right\| + \left\|\frac{1}{nT}\sum_i P_{\tilde{X}_i}\tilde{H}_i\theta_i\tilde{u}_i'\right\| + \left\|\frac{1}{nT}\sum_i P_{\tilde{X}_i}\tilde{H}_i\theta_i\tilde{u}_i' P_{\tilde{X}_i}\right\| \\ &= \|I_{51}\| + \|I_{52}\| + \|I_{53}\| + \|I_{54}\|.\end{aligned}$$

By the properties of the trace operator, we have:

$$\begin{aligned}E\left[\|I_{51}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] &= E\left[\frac{1}{n^2T^2}Tr\left[\sum_{i,j}\tilde{H}_i\theta_i\tilde{u}_i'\tilde{u}_j\theta_j'\tilde{H}_j'\right]\left| h_{\underline{T}}, \mathcal{J}\right.\right] \\ &= \frac{1}{n^2T^2}\sum_{i,j}E\left[\theta_j'\tilde{H}_i'\tilde{H}_i\theta_i\tilde{u}_i'\tilde{u}_j \left| h_{\underline{T}}, \mathcal{J}_i, \mathcal{J}_j\right.\right],\end{aligned}$$

where $\theta_j'\tilde{H}_j{}'\tilde{H}_i\theta_i = \sum_t I_{ij,t}h_t'\theta_i\theta_j' h_t$ and $\tilde{u}_i'\tilde{u}_j = \sum_t I_{ij,t}u_{i,t}u_{j,t}$. To ease notation $\theta_i$ is now treated as a scalar. Then,

$$E\left[\|I_{51}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] \leq \frac{1}{n^2T}\sum_{i,j}\sum_t I_{ij,t}\theta_i\theta_j\left(\frac{1}{T}\sum_t h_t' h_t\right)E\left[u_{i,t}u_{j,t}\left| h_{\underline{T}}, \mathcal{J}_i, \mathcal{J}_j\right.\right].$$

By Assumptions A.2a) and A.4, we have $E\left[\|I_{51}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] \leq \frac{1}{n}C$, which implies $\|I_{51}\| = O_p\left(\frac{1}{\sqrt{n}}\right)$. Similarly, by applying the properties of the trace operator and expectation, we have:

$$\begin{aligned}E\left[\|I_{52}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] &= \frac{1}{n^2T^2}\sum_{i,j}E\left[\theta_j'\tilde{H}_j'\tilde{H}_i\theta_i\tilde{u}_i' P_{\tilde{X}_i}P_{\tilde{X}_j}\tilde{u}_j \left| h_{\underline{T}}, \mathcal{J}_i, \mathcal{J}_j\right.\right] \\ &\leq \frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1,t_2}\tau_{i,T}\tau_{j,T}I_{i,t_1}I_{j,t_2}\theta_i\theta_j\hat{Q}_{x,i}^{-1}\hat{Q}_{x,j}^{-1}\left(\frac{1}{T}\sum_t h_t' h_t\right)\left(\frac{1}{T}\sum_t I_{ij,t}x_{i,t}x_{j,t}\right) \\ &\quad x_{i,t_1}x_{j,t_2}E\left[u_{i,t_1}u_{j,t_2}\left| h_{\underline{T}}, \mathcal{J}_i, \mathcal{J}_j\right.\right].\end{aligned}$$

From $\tau_{i,T} \leq \chi_{2,T}$ and $\left\|\hat{Q}_{x,i}^{-1}\right\| \leq C\chi_{1,T}^2$, and Assumptions A.3, A.4 and A.2b), $E\left[\|I_{52}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] \leq \frac{1}{nT}C\chi_{1,T}^4\chi_{2,T}^2$, which implies $\|I_{52}\| = O_p\left(\frac{\chi_{1,T}^2\chi_{2,T}}{\sqrt{nT}}\right)$. Moreover, we get

$$
\begin{aligned}
E\left[\|I_{53}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] &= \frac{1}{n^2T^2}\sum_{i,j} E\left[\theta_j'\tilde{H}_j'P_{\tilde{X}_j}P_{\tilde{X}_i}\tilde{H}_i\theta_i\tilde{u}_i'\tilde{u}_j \left| h_{\underline{T}}, \mathcal{J}_i, \mathcal{J}_j\right.\right] \\
&\leq \frac{1}{n^2T}\sum_{i,j}\sum_t \tau_{i,T}\tau_{j,t}I_{ij,t}\theta_i\theta_j\hat{Q}_{x,i}^{-1}\hat{Q}_{x,j}^{-1}\left(\frac{1}{T}\sum_t I_{ij,t}x_{i,t}x_{j,t}\right) \\
&\quad \left(\frac{1}{T}\sum_t I_{i,t}x_{i,t}h_t\right)\left(\frac{1}{T}\sum_t I_{j,t}x_{j,t}h_t\right)E\left[u_{i,t}u_{j,t}\left| h_{\underline{T}}, \mathcal{J}_i, \mathcal{J}_j\right.\right].
\end{aligned}
$$

By using that $\tau_{i,T} \leq \chi_{2,T}$, $\left\|\hat{Q}_{x,i}^{-1}\right\| \leq C\chi_{1,T}^2$, and Assumptions A.2a), A.3, and A.4, $E\left[\|I_{53}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] \leq \frac{1}{n}C\chi_{1,T}^4\chi_{2,T}^2$, and $\|I_{53}\| = O_p\left(\frac{\chi_{1,T}^2\chi_{2,T}}{\sqrt{n}}\right)$. Finally, we have:

$$
\begin{aligned}
E\left[\|I_{54}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] &= \frac{1}{n^2T^2}\sum_{i,j} E\left[\theta_j'\tilde{H}_j'P_{\tilde{X}_j}P_{\tilde{X}_i}\tilde{H}_i\theta_i\tilde{u}_i'P_{\tilde{X}_i}P_{\tilde{X}_j}\tilde{u}_j \left| h_{\underline{T}}, \mathcal{J}_i, \mathcal{J}_j\right.\right] \\
&\leq \frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1,t_2}\tau_{i,T}^2\tau_{j,t}^2 I_{i,t_1}I_{j,t_2}\theta_i\theta_j\hat{Q}_{x,i}^{-2}\hat{Q}_{x,j}^{-2}\left(\frac{1}{T}\sum_t I_{ij,t}x_{i,t}x_{j,t}\right)^2 \\
&\quad \left(\frac{1}{T}\sum_t I_{i,t}x_{i,t}h_t\right)\left(\frac{1}{T}\sum_t I_{j,t}x_{j,t}h_t\right)x_{i,t_1}x_{j,t_2}E\left[u_{i,t_1}u_{j,t_2}\left| h_{\underline{T}}, \mathcal{J}_i, \mathcal{J}_j\right.\right],
\end{aligned}
$$

then, by $\tau_{i,T} \leq \chi_{2,T}$, $\left\|\hat{Q}_{x,i}^{-1}\right\| \leq C\chi_{1,T}^2$, and Assumptions A.2b), A.3, and A.4, $E\left[\|I_{54}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] \leq \frac{1}{nT}C\chi_{1,T}^8\chi_{2,T}^4$, and $\|I_{54}\| = O_p\left(\frac{\chi_{1,T}^4\chi_{2,T}^2}{\sqrt{nT}}\right)$. The results obtained for terms $I_{51}, I_{52}, I_{53}$ and $I_{54}$ imply that $\|I_5\| = O_p\left(\frac{1}{\sqrt{n}}\right)$.

*Term depending on* $I_6$. The result follows by similar arguments used to prove that $\|I_5\| = O_p\left(\frac{1}{\sqrt{n}}\right)$.

*Term depending on* $I_7$. By the definition of $M_{\tilde{X}_i}$, we have $M_{\tilde{X}_i}\tilde{u}_i\tilde{u}_i'M_{\tilde{X}_i} = \tilde{u}_i\tilde{u}_i' - \tilde{u}_i\tilde{u}_i'P_{\tilde{X}_i} - P_{\tilde{X}_i}\tilde{u}_i\tilde{u}_i' + P_{\tilde{X}_i}\tilde{u}_i\tilde{u}_i'P_{\tilde{X}_i}$. From the properties of the norm operator, we deduce

$$
\begin{aligned}
\|I_7\|_{op} &\leq \left\|\frac{1}{nT}\sum_i \tilde{u}_i\tilde{u}_i'\right\|_{op} + \left\|\frac{1}{nT}\sum_i \tilde{u}_i\tilde{u}_i'P_{\tilde{X}_i}\right\|_{op} + \left\|\frac{1}{nT}\sum_i P_{\tilde{X}_i}\tilde{u}_i\tilde{u}_i'\right\|_{op} + \left\|\frac{1}{nT}\sum_i P_{\tilde{X}_i}\tilde{u}_i\tilde{u}_i'P_{\tilde{X}_i}\right\|_{op} \\
&\leq \|I_{71}\|_{op} + \|I_{72}\| + \|I_{73}\| + \|I_{74}\|.
\end{aligned}
$$

If the $\tilde{u}_{i,t} = I_{i,t}u_{i,t}$ satisfy the weak time-series and cross-sectional correlation assumptions in Bai and Ng (2002b) and Ahn and Horenstein (2013), we have $\|I_{71}\|_{op} \leq \mu_1 (I_{71}) = O_p \left(C_{n,T}^{-2}\right)$. Moreover,

$$\begin{aligned} E\left[\|I_{72}\|^2 \left| h_{\underline{T}}, \mathcal{J} \right.\right] &= \frac{1}{n^2T^2} \sum_{i,j} E\left[\tilde{u}_j' \tilde{u}_i \tilde{u}_i' P_{\tilde{X}_i} P_{\tilde{X}_j} \tilde{u}_j \left| h_{\underline{T}}, \mathcal{J}_i, \mathcal{J}_j \right.\right] \\ &\leq \frac{1}{n^2T^3} \sum_{i,j} \sum_{t_1,t_2,t_3} \tau_{i,T}\tau_{j,T} I_{i,t} I_{j,t} I_{ij,t} \hat{Q}_{x,i}^{-1} \hat{Q}_{x,j}^{-1} \\ &\quad \left(\frac{1}{T}\sum_t I_{ij,t} x_{i,t} x_{j,t}\right) x_{i,t_1} x_{j,t_2} E\left[u_{i,t_1} u_{j,t_2} u_{i,t_3} u_{j,t_3} \left| h_{\underline{T}}, \mathcal{J}_i, \mathcal{J}_j \right.\right]. \end{aligned}$$

The bound $E\left[\|I_{72}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] \leq \frac{1}{T^2} C\chi_{1,T}^4 \chi_{2,T}^2$ follows by $\tau_{i,T} \leq \chi_{2,T}, \left\|\hat{Q}_{x,i}^{-1}\right\| \leq C\chi_{1,T}^2$, and Assumptions A.2d) and A.3. Thus, $\|I_{72}\| = O_p \left(\frac{\chi_{1,T}^2 \chi_{2,T}}{T}\right)$. By applying similar arguments, we get $E\left[\|I_{73}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] \leq \frac{1}{T^2} C\chi_{1,T}^4 \chi_{2,T}^2$ and $\|I_{73}\| = O_p \left(\frac{\chi_{1,T}^2 \chi_{2,T}}{T}\right)$. Furthermore, by using similar arguments for $I_3$, from $\tau_{i,T} \leq \chi_{2,T}$, $\left\|\hat{Q}_{x,i}^{-1}\right\| \leq C\chi_{1,T}^2$, and Assumptions A.2d), A.3, we get $E\left[\|I_{74}\|^2 \left| h_{\underline{T}}, \mathcal{J}\right.\right] \leq \frac{1}{T^2} C\chi_{1,T}^8 \chi_{2,T}^4$ and $\|I_{74}\| = O_p \left(\frac{\chi_{1,T}^4 \chi_{2,T}^2}{T}\right)$. Finally, we deduce that $\|I_7\|_{op} = O_p \left(C_{n,T}^{-2} + \frac{\chi_{1,T}^4 \chi_{2,T}^2}{T}\right)$.

# Appendix 3 Link with Stock and Watson (2002b)

We consider the EM algorithm proposed by Stock and Watson (2002b):

$$\tilde{\tilde{\varepsilon}}_{i,t} = \begin{cases} \hat{\varepsilon}_{i,t}, & \text{if } I_{i,t} = 1, \\ \hat{\theta}_i \hat{h}_t, & \text{if } I_{i,t} = 0. \end{cases}$$

The statistic is $\xi = \mu_1 \left(\frac{\tilde{\tilde{\varepsilon}}\tilde{\tilde{\varepsilon}}'}{nT}\right) - \frac{1}{nT}\sum_i \sum_t (1 - I_{i,t}) \left(\hat{\theta}_i \hat{h}_t\right)^2 - g(n,T)$. Below we show that $\xi$ is the difference of the EM criteria under the two models. Comparing the two test statistics gives the following link: $\frac{1}{nT}\sum_i \sum_t (1 - I_{i,t}) \left(\hat{\theta}_i \hat{h}_t\right)^2 = \frac{1}{nT}\left\|\tilde{\varepsilon} - \tilde{\tilde{\varepsilon}}\right\|^2$.

To study the EM algorithm, we work as if the true error terms $\varepsilon_{i,t}$ are observed when $I_{i,t} = 1$. This error is replaced by the residual $\hat{\varepsilon}_{i,t}$. We consider the $j$th iteration of the algorithm. Let $\tilde{\zeta} = \left(\tilde{\Theta}, \tilde{H}\right)$ denotes the estimates of $\Theta$ and $H$ obtained from the $(j-1)$th iteration, and let $Q\left(\zeta, \tilde{\zeta}\right) = E_{\tilde{\zeta}}\left[\mathcal{L}(\zeta) \left|\varepsilon\right.\right]$, where $\mathcal{L}(\zeta) = \frac{1}{nT}\sum_i\sum_t \left(\varepsilon^*_{i,t} - \theta_i h_t\right)^2$, and $E_{\tilde{\zeta}}[\cdot|\varepsilon]$ denotes conditional expectation given the panel of observations under parameter $\tilde{\zeta}$. We study $Q\left(\zeta, \tilde{\zeta}\right)$ under the two models. Under both $\mathcal{M}_1$ and $\mathcal{M}_2$, we consider a pseudo model for the innovations such that $u_{i,t} \sim i.i.d.\left(0, \sigma^2_{i,t}\right)$

- Under $\mathcal{M}_1$: we get

$$Q_0\left(\zeta, \tilde{\zeta}\right) = E\left[\frac{1}{nT}\sum_i\sum_t \left(\varepsilon^*_{i,t}\right)^2 |\varepsilon\right] = \frac{1}{nT}\sum_i\sum_t E\left[\left(\varepsilon^*_{i,t}\right)^2 |\varepsilon\right].$$

We have

$$E\left[\varepsilon^*_{i,t}|\varepsilon\right] = \begin{cases} \varepsilon_{i,t}, & \text{if } I_{i,t} = 1, \\ 0, & \text{if } I_{i,t} = 0, \end{cases} \text{ and } V\left[\varepsilon^*_{i,t}|\varepsilon\right] = \begin{cases} 0, & \text{if } I_{i,t} = 1, \\ \sigma^2_{i,t}, & \text{if } I_{i,t} = 0. \end{cases}$$

and $E\left[\left(\varepsilon^*_{i,t}\right)^2 |\varepsilon\right] = I_{i,t}\varepsilon^2_{i,t} + \left(1 - I_{i,t}\right)\sigma^2_{i,t}$. Thus,

$$Q_0 = Q_0\left(\zeta, \tilde{\zeta}\right) = \frac{1}{nT}\sum_i\sum_t I_{i,t}\varepsilon^2_{i,t} + \frac{1}{nT}\sum_i\sum_t \left(1 - I_{i,t}\right)\sigma^2_{i,t}.$$

- Under $\mathcal{M}_2$: we get

$$\begin{aligned} Q_1\left(\zeta, \tilde{\zeta}\right) &= E_{\tilde{\zeta}}\left[\frac{1}{nT}\sum_i\sum_t \left(\varepsilon^*_{i,t} - \theta_i h_t\right)^2 |\varepsilon\right] \\ &= \frac{1}{nT}\sum_i\sum_t E_{\tilde{\zeta}}\left[\left(\varepsilon^*_{i,t} - \theta_i h_t\right)^2 |\varepsilon\right] \\ &= \frac{1}{nT}\sum_i\sum_t E_{\tilde{\zeta}}\left[\left(\varepsilon^*_{i,t} - E_{\tilde{\zeta}}\left[\varepsilon^*_{i,t}|\varepsilon\right] + E_{\tilde{\zeta}}\left[\varepsilon^*_{i,t}|\varepsilon\right] - \theta_i h_t\right)^2 |\varepsilon\right] \\ &= \frac{1}{nT}\sum_i\sum_t V_{\tilde{\zeta}}\left[\varepsilon^*_{i,t}|\varepsilon\right] + \frac{1}{nT}\sum_i\sum_t \left(E_{\tilde{\zeta}}\left[\varepsilon^*_{i,t}|\varepsilon\right] - \theta_i h_t\right)^2. \end{aligned}$$

We have

$$\tilde{\varepsilon}_{i,t} := E_{\tilde{\zeta}}\left[\varepsilon^*_{i,t}|\varepsilon\right] = \begin{cases} \varepsilon_{i,t}, & \text{if } I_{i,t} = 1, \\ \tilde{\theta}_i\tilde{h}_t, & \text{if } I_{i,t} = 0, \end{cases} \text{ and } V\left[\varepsilon^*_{i,t}|\varepsilon\right] = \begin{cases} 0, & \text{if } I_{i,t} = 1, \\ \sigma^2_{i,t}, & \text{if } I_{i,t} = 0. \end{cases}$$

Thus, $Q_1\left(\zeta, \tilde{\zeta}\right) = \frac{1}{nT}\sum_i\sum_t \left(\tilde{\tilde{\varepsilon}}_{i,t} - \theta_i h_t\right)^2 + \frac{1}{nT}\sum_i\sum_t \left(1 - I_{i,t}\right)\sigma^2_{i,t}$, and the values of $\zeta$ that minimize $Q_1\left(\zeta, \tilde{\zeta}\right)$ can be calculated by $\min_\zeta \frac{1}{nT}\sum_i\sum_t \left(\tilde{\tilde{\varepsilon}}_{i,t} - \theta_i h_t\right)^2$. This minimization problem reduces to the usual PCA on data $\tilde{\tilde{\varepsilon}}$: $\min_{\zeta} \frac{1}{nT} \sum_i \sum_t \left(\tilde{\tilde{\varepsilon}}_{i,t} - \theta_i h_t\right)^2 = \frac{1}{nT} \sum_i \sum_t \tilde{\tilde{\varepsilon}}_{i,t}^2 - \mu_1 \left(\frac{\tilde{\tilde{\varepsilon}}\tilde{\tilde{\varepsilon}}'}{nT}\right)$. Therefore, at convergence with $\hat{\zeta} = \tilde{\zeta}$, we have

$$
\begin{aligned}
Q_1\left(\hat{\zeta}, \tilde{\zeta}\right) &= \frac{1}{nT} \sum_i \sum_t \tilde{\tilde{\varepsilon}}_{i,t}^2 - \mu_1 \left(\frac{\tilde{\tilde{\varepsilon}}\tilde{\tilde{\varepsilon}}'}{nT}\right) + \frac{1}{nT} \sum_i \sum_t \left(1 - I_{i,t}\right) \sigma_{i,t}^2 \\
&= \frac{1}{nT} \sum_i \sum_t I_{i,t} \varepsilon_{i,t}^2 + \frac{1}{nT} \sum_i \sum_t \left(1 - I_{i,t}\right) \left(\hat{\theta}_i \hat{h}_t\right)^2 \\
&\quad - \mu_1 \left(\frac{\tilde{\tilde{\varepsilon}}\tilde{\tilde{\varepsilon}}'}{nT}\right) + \frac{1}{nT} \sum_i \sum_t \left(1 - I_{i,t}\right) \sigma_{i,t}^2.
\end{aligned}
$$

Finally, the difference of the two EM criterias is

$$
Q_0 - Q_1\left(\hat{\zeta}, \hat{\zeta}\right) = \mu_1 \left(\frac{\tilde{\tilde{\varepsilon}}\tilde{\tilde{\varepsilon}}'}{nT}\right) - \frac{1}{nT} \sum_i \sum_t \left(1 - I_{i,t}\right) \left(\hat{\theta}_i \hat{h}_t\right)^2,
$$

which gives the interpretation of the test statistic.

## Appendix 4 Monte-Carlo experiments

In this section, we perform simulation exercises on balanced and unbalanced panels in order to study the properties of our diagnostic criterion. We pay particular attention to the probability of diagnosing the correct model and its interaction with $n$ and $T$ in finite samples. The simulation design mimics the empirical features of our data. The balanced case serves as benchmark to understand when $T$ and $n$ are sufficiently large to apply theory. The unbalanced case shows that we can exploit the guidelines found for the balanced case when we substitute the average of the sample sizes of the individual assets, i.e., a kind of operative sample size, for $T$. To summarize our Monte Carlo findings, we do not face any finite sample distortions for the selection rule under $\mathcal{M}_1$ when $n =$ .... and $T =$ ... since we get estimates of $Pr(\xi < 0|\mathcal{M}_1)$ close to ..., and under $\mathcal{M}_2$ when $n =$ .... and $T =$ ... since we get estimates of $Pr(\xi > 0|\mathcal{M}_2)$ close to .... In light of these results, we do not expect to face significant diagnostic bias in our empirical application.

### A.4.1 Balanced panel

Under $\mathcal{M}_1$, we simulate $S$ datasets of excess returns from a one-factor model (CAPM). A simulated dataset includes: a vector of factor loadings $b^s \in \mathbb{R}^n$, and a variance-covariance matrix $\Omega^s \in \mathbb{R}^{n \times n}$. At each simulation $s = 1, ..., S$, we randomly draw $n \leq 10,410$ assets from the sample of our empirical analysis that comprises 10,410 individual stocks with $T_i \geq 12$. The assets are listed by industrial sectors. We use the classification proposed by Ferson and Harvey (1999). The vector $b^s$ is composed by the estimated factor loadings for the $n$ randomly chosen assets. At each simulation, we build a block diagonal matrix $\Omega^s$ with blocks matching industrial sectors. The $n$ elements of the main diagonal of $\Omega^s$ correspond to the variances of the estimated residuals of the individual assets. The off-diagonal elements of $\Omega^s$ are covariances computed by fixing correlations within block equal to the average correlation of the industrial sector computed from the $10,410 \times 10,410$ thresholded variance-covariance matrix of estimated residuals. Hence we get a setting in line with the weak block dependence case shown in GOS to exhibit an approximate factor structure.

Let us define $R^s_{i,t}$ the simulated excess returns of asset $i$ at time $t$ as follows

$$R^s_{i,t} = b^s_i f_t + \varepsilon^s_{i,t}, \text{ for } i = 1, ..., n, \text{ and } t = 1, ..., T, \tag{13}$$

where $f_t$ is the market excess returns and $\varepsilon^s_{i,t}$ is the error term. In (13), we impose the intercepts to be zero to satisfy the no-arbitrage restrictions for tradable factors. The $n \times 1$ error vectors $\boldsymbol{\varepsilon}^s_t$ are independent across time and Gaussian with mean zero and variance-covariance matrix $\Omega^s_B$. We apply our diagnostic criterion on every simulated dataset of excess returns. Since the panel is balanced, we do not need to fix $\chi_{2,T}$. We only use $\chi_{1,T} = 15$. However, this trimming level does not affect the number of assets $n$ in the simulations.

In order to study the properties under $\mathcal{M}_2$, we generate data under a three-factor alternative hypothesis, i.e., two omitted factors, and then we estimate a one-factor model to get the residuals. We build the simulated dataset as above except that we use estimated loadings, variance, and covariances for the Fama-French model on the CRSP dataset instead of the CAPM estimates.

In order to understand how our diagnostic criterion works for different finite samples, we perform exercises combining different values of the cross-sectional dimension $n$ and the time dimension $T$. Table ... reports estimates of $Pr(\xi < 0|\mathcal{M}_1)$ and $Pr(\xi > 0|\mathcal{M}_2)$, i.e., selection probabilities of the correct model estimated from the simulated datasets.

**Table 7: Selection probabilities, balanced case**

| $T$ | 150 | | | | 500 | | | |
|---|---|---|---|---|---|---|---|---|
| $n$ | 150 | 500 | 1,000 | 1,500 | 150 | 500 | 1,000 | 1,500 |
| $Pr(\xi_1 < 0 \vert \mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_2 < 0 \vert \mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_3 < 0 \vert \mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_1 < 0 \vert \mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_2 < 0 \vert \mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_3 < 0 \vert \mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_1 > 0 \vert \mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_2 > 0 \vert \mathcal{M}_2)$ | 0.9580 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_3 > 0 \vert \mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_1 > 0 \vert \mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_2 > 0 \vert \mathcal{M}_2)$ | 0.9640 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_3 > 0 \vert \mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

### A.4.2 Unbalanced panel

Let us repeat similar exercises as in the previous section, but with unbalanced characteristics for the simulated datasets. We introduce these characteristics through a matrix of observability indicators $I^s \in \mathbb{R}^{n \times T}$. The matrix gathers the indicator vectors for the $n$ randomly chosen assets. We fix the maximal sample size $T = 528$ as in the empirical application.

In the unbalanced setting, the excess returns $R^s_{i,t}$ of asset $i$ at time $t$ under $\mathcal{M}_1$ is:

$$R^s_{i,t} = b^s_i f_t + \varepsilon^s_{i,t}, \text{ if } I^s_{i,t} = 1, \text{ for } i = 1, ..., n, \text{ and } t = 1, ..., T, \tag{14}$$

where $I^s_{i,t}$ is the observability indicator of asset $i$ at time $t$ in simulation $s$.

Under $\mathcal{M}_2$, we again replace CAPM estimates with estimates for the Fama-French model to get a three-factor alternative.

In Tables 8 and 9, we provide the operative cross-sectional and time-series sample sizes in the Monte-Carlo repetitions for trimming $\chi_{1,T} = 15$ and four different levels of trimming $\chi_{2,T}$. More precisely, in Table 8, we report the average number $\bar{n}^{\chi}$ of retained assets across simulations, as well as the minimum $\min(n^{\chi})$ and the maximum $\max(n^{\chi})$ across simulations (rounded). For the lowest level of trimming $\chi_{2,T} = T/12$, all assets are kept in all simulations, while for the level of trimming $\chi_{2,T} = T/60$ on average we keep about two thirds of the assets. In Table 9, we report the average across assets of the $\bar{T}_i$, that are the average time-series size $T_i$ for asset $i$ across simulations, as well as the min and the max of the $\bar{T}_i$. Since the distribution of $T_i$ for an asset $i$ is right-skewed, we also report the average across assets of the median $T_i$. For trimming level $\chi_{2,T} = T/60$, the average mean time-series size is about 180 months,while the average median time-series size is 140 months.

Table 10 reports estimates of $Pr(\xi < 0|\mathcal{M}_1)$ and $Pr(\xi > 0|\mathcal{M}_2)$.

**Table 8: Operative cross-sectional sample size**

| trimming level | $\chi_{2,T} = \frac{T}{12}$ | | | | | $\chi_{2,T} = \frac{T}{60}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 500 | 1,000 | 3,000 | 6,000 | 9,000 | 500 | 1,000 | 3,000 | 6,000 | 9,000 |
| $\bar{n}^{\chi}$ | 500 | 1,000 | 3,000 | 6,000 | 9,000 | 326 | 651 | 1,955 | 3,905 | 5,857 |
| $\min(n^{\chi})$ | 500 | 1,000 | 3,000 | 6,000 | 9,000 | 299 | 613 | 1,890 | 3,820 | 5,823 |
| $\max(n^{\chi})$ | 500 | 1,000 | 3,000 | 6,000 | 9,000 | 359 | 694 | 2,018 | 3,977 | 5,903 |
| trimming level | $\chi_{2,T} = \frac{T}{120}$ | | | | | $\chi_{2,T} = \frac{T}{240}$ | | | | |
| $n$ | 500 | 1,000 | 3,000 | 6,000 | 9,000 | 500 | 1,000 | 3,000 | 6,000 | 9,000 |
| $\bar{n}^{\chi}$ | 194 | 388 | 1,161 | 2,325 | 3,488 | 65 | 128 | 386 | 772 | 1,158 |
| $\min(n^{\chi})$ | 162 | 348 | 1,080 | 2,245 | 3,437 | 44 | 97 | 338 | 712 | 1,123 |
| $\max(n^{\chi})$ | 223 | 434 | 1,223 | 2,398 | 3,533 | 88 | 162 | 442 | 826 | 1,185 |

**Table 9: Operative time-series sample size**

| trimming level | $\chi_{2,T} = \frac{T}{12}$ | $\chi_{2,T} = \frac{T}{60}$ | $\chi_{2,T} = \frac{T}{120}$ | $\chi_{2,T} = \frac{T}{240}$ |
|---|---|---|---|---|
| mean$\left(\bar{T}_i\right)$ | 126 | 175 | 235 | 365 |
| $\min\left(\bar{T}_i\right)$ | 113 | 158 | 216 | 331 |
| $\max\left(\bar{T}_i\right)$ | 141 | 190 | 260 | 400 |
| mean(median $(T_i)$) | 88 | 141 | 198 | 344 |

**Table 10: Selection probabilities, unbalanced case**

| trimming level | $\chi_{2,T} = \frac{T}{12}$ | | | | | $\chi_{2,T} = \frac{T}{60}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 500 | 1,000 | 3,000 | 6,000 | 9,000 | 500 | 1,000 | 3,000 | 6,000 | 9,000 |
| $Pr(\xi_1 < 0\|\mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_2 < 0\|\mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_3 < 0\|\mathcal{M}_1)$ | 0.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_1 < 0\|\mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_2 < 0\|\mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_3 < 0\|\mathcal{M}_1)$ | 0.0020 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_1 > 0\|\mathcal{M}_2)$ | 0.9520 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9960 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_2 > 0\|\mathcal{M}_2)$ | 0.6180 | 0.9980 | 1.0000 | 1.0000 | 1.0000 | 0.9380 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_3 > 0\|\mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_1 > 0\|\mathcal{M}_2)$ | 0.9480 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9960 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_2 > 0\|\mathcal{M}_2)$ | 0.6140 | 0.9980 | 1.0000 | 1.0000 | 1.0000 | 0.9360 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_3 > 0\|\mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

| trimming level | $\chi_{2,T} = \frac{T}{120}$ | | | | | $\chi_{2,T} = \frac{T}{240}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 500 | 1,000 | 3,000 | 6,000 | 9,000 | 500 | 1,000 | 3,000 | 6,000 | 9,000 |
| $Pr(\xi_1 < 0\|\mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.0000 | 0.0040 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_2 < 0\|\mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.0000 | 0.1140 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_3 < 0\|\mathcal{M}_1)$ | 0.0000 | 0.9660 | 1.0000 | 1.0000 | 1.0000 | 0.0000 | 0.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_1 < 0\|\mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.0000 | 0.0040 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_2 < 0\|\mathcal{M}_1)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.0000 | 0.1480 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_3 < 0\|\mathcal{M}_1)$ | 0.0000 | 0.9800 | 1.0000 | 1.0000 | 1.0000 | 0.0000 | 0.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_1 > 0\|\mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_2 > 0\|\mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\xi_3 > 0\|\mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_1 > 0\|\mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_2 > 0\|\mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $Pr(\check{\xi}_3 > 0\|\mathcal{M}_2)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |