ESTIMATION OF LARGE DIMENSIONAL CONDITIONAL FACTOR MODELS IN FINANCE

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

First draft: February 2019. This version: September 2020.

Abstract

This chapter surveys recent econometric methodologies for inference in large dimensional conditional factor models in finance. Changes in the business cycle and asset characteristics induce time variation in factor loadings and risk premia to be accounted for. The growing trend in the use of disaggregated data for individual securities motivates our focus on methodologies for a large number of assets. The chapter starts with an historical perspective on conditional factor models with a small number of assets for comparison purpose. Then, it outlines the concept of approximate factor structure in the presence of conditional information, and reviews an arbitrage pricing theory for large dimensional factor models in this framework. For inference, we distinguish between two different cases depending on whether factors are observable or not. We focus on diagnosing model specification, estimating conditional risk premia, and testing asset pricing restrictions under increasing cross-sectional and time series dimensions. At the end of the chapter, we provide new empirical findings based on a broad set of factor models and contrast analysis based on individual stocks and standard sets of portfolios. We also discuss the impact on computing time-varying cost of equity for a firm, and summarize differences between results for developed and emerging markets in an international setting.

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

Keywords: large panel, factor model, conditional information, risk premium, asset pricing, emerging markets.

[a]Università della Svizzera italiana (USI Lugano) and Swiss Finance Institute, [b] European Commission, Joint Research Centre, [c]University of Geneva and Swiss Finance Institute.

*Acknowledgements: This chapter is for eventual publication in the *Handbook of Econometrics*, Vol. 7. We gratefully acknowledge the financial support of the Swiss National Science Foundation (Prodoc project PDFM11-114533, NCCR FINRISK and grant 105218-162633). We thank Rosa Matzkin, Steven Durlauf, and the referee as well as L. Barras, I. Chaieb, J. Fan, A.-P. Fortin, B. Kelly, H. Langlois, Y. Liao, M. Pelger, S. Pruitt, and M. Weidner for helpful comments.*Disclaimer: The content of this article does not reflect the official opinion of the European Commission. Responsibility for the information and views expressed therein lies entirely with the authors.

1 Introduction

The objective of this chapter is to survey recent econometric methodologies for inference in large conditional factor models in finance. We focus on diagnosing model specification, estimating conditional risk premia, and testing asset pricing restrictions under increasing cross-sectional and time series dimensions. Until recently, most theory and empirics have focused on either time-varying factor models with a small number of assets or time-invariant factor models with a large number of assets. This chapter focuses on new econometric tools targeting the combination of both large data sets and time variation.

Financial and macroeconomic variables influence the risk-return trade-off. Conditional linear factor models aim at capturing their time-varying influence 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.

Risk premia measure financial compensation asked by investors for bearing systematic risk. We expect their time variation because of changes in the economic conditions (business cycles), the investment opportunity sets (firms to invest in), and the economic agent expectations (time preference and risk aversion of investors). The workhorse to estimate equity risk premia in a linear multi-factor setting is the two-pass cross-sectional regression method developed by Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973). A series of papers address its large and finite sample properties for linear factor models with time-invariant coefficients and a fixed number n of assets, see e.g. Shanken (1985, 1992), Jagannathan and Wang (1998), Shanken and Zhou (2007), Kan, Robotti and Shanken (2013), and the review paper of Jagannathan, Skoulakis and Wang (2009). That early literature did not formally address statistical inference for equity risk premia in conditional linear factor models despite its empirical relevance. It also limits itself to the empirical analysis of a small number of assets, typically 25 or 100 equity portfolios. Such an analysis neglects the huge heterogeneity expected in individual stocks and its informational content lost in the aggregation process yielding portfolios.

In this chapter, we study how we can infer the time-varying behaviour of equity risk premia from large stock returns databases under conditional linear factor models. Such an approach is inspired by the recent trend in macro-econometrics and forecasting methods trying to extract cross-sectional and time-series information simultaneously from large panels (see e.g. Stock and Watson (2002a,b), Bai (2003, 2009), Bai and Ng (2002, 2006), Forni, Hallin, Lippi and Reichlin (2000, 2004, 2005), Pesaran (2006)). Ludvigson and Ng (2007, 2009) exemplify this promising route when studying bond risk premia. Connor, Hagmann, and Linton (2012) show that large cross-sections exploit data more efficiently in a semiparametric characteristic-based factor model of stock returns. The theoretical framework underlying the Arbitrage Pricing Theory (APT) also inspires approaches relying on individual stocks returns. In this setting, 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). Under weak cross-sectional dependence among idiosyncratic error terms, such approximate factor models generate no-arbitrage restrictions in large economies where the number of assets grows to infinity. This chapter surveys econometric methodologies tailored to the APT framework. Indeed, we let the number of assets grow to infinity mimicking the large economies of financial theory. We also give an historical perspective on conditional factor models with a small number of assets for comparison purpose.

As already mentioned, 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 or empirical factors and supposed to be rich enough to extract systematic risks while idiosyncratic risk is left over to the error term. In the core part of this chapter, we focus mostly on observable factors since those are amenable to trading strategies based on low-minus-high deciles equity portfolios and, contrary to latent factors, are easy to extend to the conditional case with an explicit account of no-arbitrage restrictions coming from financial theory. In Section 5, we also consider some recent proposals for analysing time-varying models with unobservable factors, as the available literature on large cross-sectional equity datasets has mainly focused on latent factors. Linear factor models are rooted in the Arbitrage Pricing Theory (Ross (1976), Chamberlain and Rothschild (1983)) or come from a loglinearization of nonlinear consumption-based models (Campbell (1996)). A central and practical issue is to determine whether there are one or more factors omitted in the chosen specification. If the set of observable factors is correctly specified, the errors are weakly cross-sectionally correlated, namely the covariance matrix of the error terms in the factor model has a fastly vanishing largest eigenvalue. If the set of observable factors is not correctly specified, the no-arbitrage restrictions derived from APT do not hold, and the risk premia estimated by the two-pass regression approach are meaningless. Even if the omitted factors are not priced, i.e., their associated risk premia are nil, direct computations of the limits of first pass and second pass estimators under misspecification show that second pass estimates do not converge to the risk premia of the priced factors, and that biases on betas and risk premia do not compensate each other. Besides, since the no arbitrage restrictions do not hold, we cannot simply say that the risk premia are the expected factor returns for models with traded factors. Hence detecting an omitted factor is also important in that case to produce correct expected excess returns from the no arbitrage restrictions. Given the large menu of factors available in the literature (the factor zoo of Cochrane (2011), see also Harvey et al. (2016), Harvey and Liu (2016)), we need a simple diagnostic criterion to decide whether we can feel comfortable with the chosen set of observable factors before proceeding further in the empirical analysis of large cross sectional equity data sets under the APT setting. For example, if the factor model passes the diagnostic, and we reject that alphas are zero using a GRS-type statistic (Gibbons et al. (1989)), it will not be because of an omitted factor. This chapter also aims at studying such a diagnostic criterion.

The outline of this chapter is as follows. Section 2 puts the recent development of large conditional models into an historical perspective. We review econometric methods suited to conditional factor models for small cross-sections of assets for comparaison purpose. In Section 3, we consider a general framework of conditional linear factor model for asset returns in large economies. Section 4 presents inference in models with observable factors. Those two sections are largely inspired by Gagliardini, Ossola, Scaillet (2016, GOS) and Gagliardini, Ossola, Scaillet (2019, GOS2). We want to stress that the literature predating those papers could not provide a conditional factor framework consistent from both a finance and an econometric standpoint for empirical analysis of a large number of assets with observable factors, and we explain why in those two sections. We focus on diagnosing model specification, estimating conditional risk premia, and testing asset pricing restrictions under increasing cross-sectional and time series dimensions. This sequence of econometric tools yields a unifying umbrella for empirical applications of large dimensional conditional factor models in finance. In Section 5, we investigate models with unobservable factors. We look at empirical findings in Section 6. There we contrast analysis based on individual stocks and standard sets of portfolios, the latter being the default mode of previous empirical analysis in finance with time-invariant factor models. This new empirical study is broader than what is available in the literature in terms of analysed factor models and also allows us to exemplify key differences in the empirical results delivered by the two approaches. We also discuss the impact on computing time-varying cost of equity for a firm. Finally, we summarize differences between results for developed and emerging markets in an international setting. Section 6 gathers concluding remarks.

2 Conditional factor models for small cross-sections of assets

In order to put the methodologies of this chapter into an historical perspective and to relate to a broader and earlier literature, we review inference in conditional factor models for small cross-sections of assets in this section. In these conditional models, risk premia and risk exposures (betas) vary in time as functions of market conditions and economic variables. The cross-section consists of a few dozens of test assets, which typically correspond to portfolios.

Ferson and Harvey (1991) consider the following specification for the conditional asset expected gross returns:

$$E_t[r_{i,t+1}] = \lambda_{0,t} + \lambda'_{1,t} b_{i,t}, \tag{1}$$

where $E_t[\cdot]$ denotes the conditional expectation given the investors' information at date t. In the single-factor case, $b_{i,t}$ is the conditional market beta of asset i in a small cross section $i = 1, ..., n$, and $\lambda_{1,t}$ is the conditional market risk premium (Conditional CAPM). A generalization to a conditional multiple-factor specification is readily obtained by letting $b_{i,t} = V_t[f_{t+1}]^{-1} Cov_t[f_{t+1}, r_{i,t+1}]$ and $\lambda_{1,t}$ be vectors of sensitivities and risk premia for a vector f_t of risk factors. This model is implied by a conditionally linear specification of the Stochastic Discount Factor (SDF). More specifically, the absence of arbitrage opportunities in the market implies the existence of a SDF M_{t+1} that represents asset prices at date t as the conditional expectation of cash flows in $t + 1$ times M_{t+1} (Hansen and Richard (1987)). The SDF is such that the gross returns of the fundamental assets satisfy the conditional moment restrictions:

$$E_t[r_{i,t+1} M_{t+1}] = 1, \tag{2}$$

for all assets $i = 1, ..., n$. Let us assume that one admissible SDF is conditionally linear in the vector of risk factors, namely:

$$M_{t+1} = \delta_{0,t} + \delta'_{1,t} f_{t+1}, \tag{3}$$

where $\delta_{0,t}$ and $\delta_{1,t}$ are a scalar and a vector of risk prices conditional to the investors' information at date t, respectively. Then, inserting (3) into (2) and rearranging terms yields Equation (1) with $\lambda_{0,t} = R_{f,t}$ and $\lambda_{1,t} = -R_{f,t} V_t[f_{t+1}]\delta_{1,t}$, where $R_{f,t} = E_t[M_{t+1}]^{-1}$ is the conditionally risk-free short-term rate.

The estimation of the conditional risk premia (or conditional prices of risk) in these specifications has used either "non-parametric" approaches relying on rolling window regressions, or more structural parametric specifications, mostly assuming linear relationships between conditional risk prices (or risk premia) and a set of instrumental variables.

Ferson and Harvey (1991) estimate the conditional betas by means of rolling window regressions of asset returns onto factors in the previous 60 months. These rolling window estimates of the conditional betas are then used in a second pass cross-sectional regression of excess returns $R_{i,t+1}$ to obtain estimates of the risk premia $\lambda_{1,t}$ at each date. This approach has a "non-parametric" flavor as it does not require to specify a dynamic for conditional betas and risk premia. However, it is hard to develop proper statistical inference to account for the error-in-variable bias induced by the rolling window betas used as regressors in the second pass. In asset pricing inference with test portfolios, it is the composition of the test assets that varies over time as stocks with similar characteristics are assembled into different portfolios. In such a case, we can expect that estimating betas using either the full sample or rolling windows can adequately capture each test portfolio factor loadings. For example, a size sorted portfolio with small capitalization firms will consistently have a positive loading on a size factor over time. On the contrary, when we test asset pricing models using individual stocks as in the subsequent sections, the composition of the test assets is fixed (i.e., one stock), and their characteristics vary over time. As a firm evolves and its stock characteristics change over time, we cannot expect its betas to be constant. Consequently, estimating betas over rolling windows would necessarily lag its true time-varying factor exposures as time-invariant OLS estimates average recent and more distant exposures within the rolling window. It explains why we prefer an explicit specification of the beta dynamics in Section 4.

Lettau and Ludvigson (2001) assume that time variation in the conditional prices of risk is generated by a scalar (to simplify) variable z_t with linear specifications $\delta_{0,t} = d_{00} + d_{01} z_t$ and $\delta_{1,t} = d_{10} + d_{11} z_t$. Then, the SDF is linear in the vector of "scaled factors" $F_{t+1} = (z_t, f'_{t+1}, z_t f'_{t+1})'$ (Cochrane (2005)), i.e., $M_{t+1} = d_{00} + d'_1 F_{t+1}$ with $d_1 = (d_{01}, d'_{10}, d'_{11})'$. From the law of iterated expectation, the unconditional version of the restriction in (2) is

$$E[r_{i,t+1} M_{t+1}] = 1, \tag{4}$$

for all i, and we can write it as a static multi-beta pricing equation:

$$E[r_{i,t+1}] = \lambda_0 + \beta'_i \lambda_1, \tag{5}$$

where $\beta_i = V[F_t]^{-1} Cov[F_t, r_{i,t}]$, $\lambda_0 = R_f$ and $\lambda_1 = -R_f V[F_t] d_1$, with $R_f = E[M_{t+1}]^{-1}$. Thus, we can estimate the vector $\lambda = (\lambda_0, \lambda'_1)'$ by a two-pass regression methodology. The results of Shanken (1992) provide the valid asymptotic standard errors for this estimator with n fixed and $T \to \infty$ (see also Jagannathan and Wang (1998), Jagannathan et al. (2009)). We can recover the SDF parameter estimates and their standard errors by the delta method. Estimating the conditional risk premia vector $\lambda_{1,t} = -R_{f,t} V_t[f_{t+1}] \delta_{1,t}$ requires an estimator for the conditional variance $V_t[f_{t+1}]$.

We can estimate the SDF parameters in the conditionally linear SDF specification $M_{t+1}(d) = d_{00} + d_{01} z_t + (d_{10} + d_{11} z_t)' f_{t+1}$ also by the Generalized Method of Moments (GMM) (Hansen and Singleton (1982), Cochrane (1996)). By introducing the instrument matrix Z_t, the conditional moment restrictions (2) imply the set of orthogonality conditions:

$$E[Z_t (r_{t+1} M_{t+1}(d) - \iota)] = 0,$$

where $r_{t+1} = (r_{1,t+1}, ..., r_{n,t+1})'$ is the n-vector of asset gross returns and ι is the n-vector of ones. The GMM estimator of the SDF parameters vector d is linear. For given choice of the instrument matrix, we obtain the efficient GMM estimator by using a weighting matrix that is the inverse of the sample second-moment matrix of the orthogonality vector. We can enhance the efficiency of the estimator by introducing appropriate instruments (optimal instruments) when transforming the conditional restrictions into unconditional restrictions (see e.g. Chamberlain (1987), Newey (1990), Donald et al. (2009) for the i.i.d. setting, and Hansen (1985) for the time series case). For a scalar factor (to simplify), and using z_t as a single conditioning variable, the optimal instrument is:

$$
\begin{aligned}
Z_t^* &= E[\frac{\partial h_{t+1}(d^0)'}{\partial d}|z_t]V[h_{t+1}(d^0)|z_t]^{-1} \\
&= \left(E[W_{t+1}|z_t] \otimes \tilde{z}_t\right)\left(\phi(z_t)'V[Y_{t+1}|z_t]\phi(z_t)\right)^{-1}, \qquad (6)
\end{aligned}
$$

where $h_{t+1}(d) = r_{t+1}M_{t+1}(d) - \iota$ is the conditional moment vector, $W_{t+1} = \tilde{f}_{t+1}r'_{t+1}$ and $Y_{t+1} = \tilde{f}_{t+1} \otimes r_{t+1}$ are a matrix and a vector made of the asset returns and the cross-products between factor and assets returns, with $\tilde{f}_{t+1} = (1, f_{t+1})'$, $\tilde{z}_t = (1, z_t)'$, and $\phi(z_t) = (d_{00} + d_{01}z_t, d_{10} + d_{11}z_t)' \otimes I_n$ evaluated at the true parameter vector d^0. We refer to Ludvigson (2013) for further discussion within scaled consumption-based models.

Nagel and Singleton (2011) estimate the parameters in a conditionally linear SDF specification by deploying the conditional moment restrictions in (2) and implementing the optimal instruments (6) by nonparametric regression. Specifically, we can estimate the conditional mean and variance of W_{t+1} and Y_{t+1} in (6) by kernel smoothing or series estimators. We can also implement this optimal instrument GMM estimator by using information-theoretic approaches, such as the Euclidean Likelihood, Empirical Likelihood, or Exponential Tilting estimators (see e.g. Ai and Chen (2003), Antoine et al. (2007), Kitamura et al. (2004) in the i.i.d. case and Gospodinov and Otsu (2012), Gagliardini et al. (2011) in the time series case).

3 Large dimensional factor models

In this section, we outline the finance theory with large economies on which we can build the subsequent econometric approaches. We consider a conditional linear factor model with time-varying coefficients (see Connor and Korajczyk (1989) for an intertemporal competitive equilibrium version of APT yielding time-varying risk premia). 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. It yields a relationship between the ruling out of asymptotic arbitrage opportunities and an empirically testable restriction for large economies in a conditional setting. We also formalize the sampling scheme so that observed assets are random draws from an underlying population (Andrews (2005)). 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)). Combining the constructions of Hansen and Richard (1987) and Andrews (2005) defines a multi-period economy with a continuum of assets having strictly stationary and ergodic return processes. We use such a formal construction to guarantee that (i) the economy is invariant to time shifts, so that we can establish all properties by working at $t = 1$, (ii) time series averages converge almost surely to population expectations, (iii) under a suitable sampling mechanism, cross-sectional limits exist and are invariant to reordering of the assets, (iv) the derived no-arbitrage restriction is empirically testable. This construction allows reconciling finance and econometric analysis in a coherent and unified framework.

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:

$$R_t(\gamma) = a_t(\gamma) + b_t(\gamma)' f_t + \varepsilon_t(\gamma), \tag{7}$$

where vector f_t gathers the values of K 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 $\mathcal{F}_{t-1}$, and satisfy a weak cross-sectional dependence condition in the form of an upper bound on the largest eigenvalue of the error variance-covariance matrix (Assumption APR.3 in GOS). 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, the following asset pricing restriction holds:

$$a_t(\gamma) = b_t(\gamma)'\nu_t, \text{ for almost all } \gamma \in [0, 1], \tag{8}$$

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 (8) is equivalent to

$$E[R_t(\gamma)|\mathcal{F}_{t-1}] = b_t(\gamma)'\lambda_t, \tag{9}$$

where $\lambda_t = \nu_t + E[f_t|\mathcal{F}_{t-1}]$ is the vector of the conditional risk premia. The latter form is already known in the finance literature with large economies, but for either a static setting (unconditional expectation and time-invariant coefficients) or a countable number of assets. What is new here is the characterization for a conditional setting and a continuum of assets. Clearly, it also corresponds to Equation (1) since $E_t[r_{i,t+1}] - \lambda_{0,t} = E_t[R_{i,t+1}]$, but for a continuum of assets. We derived Equation (1) with a small cross-section of assets in Section 2. Let us stress the different arguments used for the small or large cross-section cases. In the former case, the absence of arbitrage opportunities implies the existence of a SDF. Then, assuming that one such admissible SDF is conditionally linear in the risk factors, we get (1). In the latter case, we exclude asymptotic arbitrage opportunities. Then, assuming a conditionally linear approximate factor structure for asset returns, we get (9).

To have an empirically workable version of Equations (7) and (8), 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 γ, 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), Dumas and Solnik (1995)) and $Z_{t-1}(\gamma)$ (Avramov and Chordia (2006)); (ii) the vector of risk premia λ_t is a linear function of lagged instruments Z_{t-1} (Dumas and Solnik (1995), 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). Hence:

$$b_t(\gamma) = B(\gamma)Z_{t-1} + C(\gamma)Z_{t-1}(\gamma), \qquad \lambda_t = \Lambda Z_{t-1}, \qquad E(f_t|\mathcal{F}_{t-1}) = FZ_{t-1}, \tag{10}$$

for some unknown parameter matrices $B(\gamma)$, $C(\gamma)$, Λ and F.

The specification choices for factor exposures and factor risk premia in (10) combined with the asset pricing restrictions in (8) imply that a stock intercept is a function of lagged instruments, namely,

$$a(\gamma) = Z'_{t-1}B(\gamma)'\left(\Lambda - F\right)Z_{t-1} + Z_{t-1}(\gamma)'C(\gamma)'\left(\Lambda - F\right)Z_{t-1}. \tag{11}$$

The no-arbitrage condition in Equation (11) shows that, with time-varying factor exposures and risk premia, a stock intercept is a quadratic form of instruments, not a linear form. Hence, an adhoc specification of the intercept neglecting the constraint (11) might yield a factor model which is not arbitrage free. The previous literature with time-varying specifications missed that constraint.

We introduce a sampling scheme to ensure that cross-sectional limits exist and are invariant to reordering of the assets. 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 γ_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 coefficients, $\varepsilon_{i,t} = \varepsilon_t(\gamma_i)$ be the error terms, and $Z_{i,t} = Z_t(\gamma_i)$ be the stock specific instruments. The characteristics $Z_{i,t}$ play a central role in the empirical strategy for individual stocks (Freyberger et al. (2020)). For example, it is doubtful for a successful company that its loading to the size factor remains identical after a shift from its initial small size to its actual big size.

By random sampling, we get a random coefficient panel model (e.g. Hsiao (2003), Chapter 6). Such a formalisation is key to reconciling finance theory and econometric modelling. Without drawings, cross-sectional averages such as $\frac{1}{n}\sum_i b_i$ correspond to determinist sequences since the b_is are then parameters. Working with the standard arbitrage pricing theory with approximate factor models has three issues. First, cross-sectional limits depend in general on the ordering of the financial assets, and there is no natural ordering between assets (firms). In the list of firms, there is no natural ordering, such as temporal or spatial, between firm A and firm B. Second, we cannot exploit either a law of large numbers to guarantee existence of those limits, nor a central limit theorem to get distributional results. Third, the asset pricing restrictions derived under no arbitrage are not testable, the so-called Shanken critique (Shanken (1982)). Shanken (1982) criticizes the original time-invariant APT since the asset pricing restrictions writes $\sum_{i=1}^{\infty}(a_i - b_i'\nu)^2 < \infty$ as derived by Chamberlain and Rothschild (1983) and we cannot test empirically the finiteness of the sum for a given countable economy. With a random sampling, the asset pricing restrictions become a moment condition $E[(a_i - b_i'\nu)^2] = 0$, and we can test a moment condition via its empirical counterpart $\frac{1}{n}\sum_{i=1}^{n}(\hat{a}_i - \hat{b}_i'\hat{\nu})^2$ with a plugin of estimates of a_i, b_i, ν (see Section 4.3).

In available datasets, we do not observe asset returns for all firms at all dates due to entry and exit from the panel. 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)). We assume independence between the observability and return generating processes conditionally on observed variables, which amounts to a missing-at-random hypothesis (Rubin (1976)). A more general assumption would imply model nonlinearities.

Through appropriate redefinitions of the regressors and coefficients, we can rewrite the model for Equations (7) and (8) as a generic random coefficient panel model:

$$R_{i,t} = x'_{i,t}\beta_i + \varepsilon_{i,t}, \tag{12}$$

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 $x_{1,i,t} = \left(vech\left[X_t\right]', Z'_{t-1} \otimes Z'_{i,t-1}\right)' \in \mathbb{R}^{d_1}$ 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 $d_2 = K(p+q)$. In vector $x_{2,i,t}$, the first components with common instruments take the interpretation of scaled factors (Cochrane (2005)), while the second components do not since they depend on i. The symmetric matrix $X_t = [X_{t,k,l}] \in \mathbb{R}^{p\times p}$ is such that $X_{t,k,l} = Z^2_{t-1,k}$, if $k = l$, and $X_{t,k,l} = 2Z_{t-1,k}Z_{t-1,l}$, otherwise, $k, l = 1, \ldots, p$, where $Z_{t,k}$ denotes the kth component of the vector Z_t. The vector-half operator $vech\left[\cdot\right]$ stacks the elements of the lower triangular part of a $p \times p$ matrix as a $p\left(p+1\right)/2 \times 1$ vector (see Chapter 2 in Magnus and Neudecker (2007) for properties of this matrix tool). The vector of coefficients β_i is a function of asset specific and common instrument parameters defining the dynamics of $a_{i,t}$ and $b_{i,t}$ in (8) and (10). We give their explicit forms in Section 4.2 where we first need them. Those forms are compatible with restrictions from asymptotic no arbitrage. In matrix notation, for any asset i, we have

$$R_i = X_i\beta_i + \varepsilon_i, \tag{13}$$

where R_i and ε_i are $T \times 1$ vectors. Regression (12) 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, and the regression coefficient vector is $\beta_i = (a_i, b'_i)'$ of dimension $d = K + 1$.

4 Inference in models with observable factors

In Section 4.1, we first study the diagnostic criterion for omitted factors before looking at the determination of the number of omitted factors. In Section 4.2, we discuss how to estimate risk premia via the two-pass regression methodology. We dedicate Section 4.3 to testing asset pricing restrictions. Throughout this chapter we assume a joint asymptotics in which the cross-sectional dimension n and time series dimension T grow such that:

$$n, T \to \infty, \qquad n = O(T^{1/\underline{\gamma}}), \qquad T = O(n^{\bar{\gamma}}), \tag{14}$$

with $0 < \underline{\gamma} \leq \bar{\gamma} \leq \infty$. Our asymptotics accommodate, among others, schemes such that T is much smaller than n (i.e., $\bar{\gamma} < 1$), or n and T are comparable ($\underline{\gamma} = \bar{\gamma} = 1$). We omit technical details and refer the reader to GOS and GOS2 which give all required assumptions and proofs.

4.1 Model diagnostic

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 (12), where the errors $(\varepsilon_{i,t})$ are weakly cross-sectionally dependent,

and

$\mathcal{M}_2$: the linear regression model (12), where the errors $(\varepsilon_{i,t})$ satisfy a factor structure.

Under model $\mathcal{M}_1$, the observable factors fully capture the systematic risk, and the error terms do not feature pervasive forms of cross-sectional dependence. This zero-factor case in the error terms should hold when we choose factors and instruments in a time-varying setting to build the variables $x_{i,t}$, so that their explanatory power for excess returns achieves weak cross-sectional correlation in the noise terms. Working with weak cross-sectional dependence, namely an approximate factor structure, avoids the stronger assumption of zero cross-sectional correlations, namely an exact factor structure. Under model $\mathcal{M}_2$, the following error factor structure holds

$$\varepsilon_{i,t} = \theta_i' h_t + u_{i,t}, \tag{15}$$

where the $m \times 1$ vector h_t includes unobservable (i.e., latent or hidden) factors, and the $u_{i,t}$ are weakly cross-sectionally correlated. The latent factors may include scaled factors to cover latent time-varying factor loadings with common instruments. Such scaled factors may come from mispecification of the functional form of the time-varying betas. Since the factors h_t are unobservable by definition, we cannot tell from the output of the diagnostic criterion whether they are pure or scaled factors. We cannot allow for latent time-varying factor loadings with stock-specific instruments in our setting because of identification issues in disentangling time-varying loadings and latent factors. This lack of identification means that we cannot estimate a generic time-varying unobservable structure from the spectral properties of a covariance matrix alone. A recent proposal in the direction of a functional specification for a time-varying $\theta_{i,t}$ is the Instrumented Principal Components Analysis (IPCA) of Kelly et al. (2017, 2019), which we review in Section 5 together with other inference approaches for latent factor models with time-varying betas. IPCA works with linear loading specifications, with balanced panels, and without observable factors. The $m \times 1$ vector θ_i corresponds to the factor loadings, and the number m of common factors is assumed unknown. In vector notation, we have:

$$\varepsilon_i = H\theta_i + u_i, \tag{16}$$

where H is the $T \times m$ matrix of unobservable factor values, and u_i is a $T \times 1$ vector. In Equation (15), the θ_is and h_ts are also called interactive fixed effects in the panel literature (Pesaran (2006), Bai (2009), Moon and Weidner (2015)). King et al. (1994) use them to capture the correlation between the unanticipated innovations in observable descriptors of economic performance (e.g. industrial production, inflation, etc.) and stock returns. Gobillon and Magnac (2016) use them to get treatment effect estimates in regional policy evaluation and characterize the generic bias induced by the popular difference-in-differences procedure. To diagnose the absence of omitted interactive effects is clearly important when applying the difference-in-differences procedure.

To compute the diagnostic criterion that checks whether the error terms are weakly cross-sectionally correlated or share at least one common factor, we estimate the generic panel model (12) by OLS applied asset by asset, and we get estimators

$$\hat{\beta}_i = \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_t I_{i,t} x_{i,t} R_{i,t}, \quad i = 1, ..., n, \tag{17}$$

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 β_i, we apply a trimming approach. 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 the $d \times d$ matrix $\hat{Q}_{x,i}$, $\mu_1\left(\hat{Q}_{x,i}\right)$ and $\mu_d\left(\hat{Q}_{x,i}\right)$ are its largest, resp. its smallest, eigenvalue and $\tau_{i,T} = T/T_i$. We assume that 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:

$$\xi = \mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i\bar{\varepsilon}_i'\right) - g(n,T), \tag{18}$$

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 (2002) consider several simple potential candidates for the penalty $g(n,T)$. In vector $\bar{\varepsilon}_i$, the unavailable residuals are replaced by zeros. Then we use the following model selection rule: we select $\mathcal{M}_1$ if $\xi < 0$, and we select $\mathcal{M}_2$ if $\xi > 0$, since (a) $Pr\left(\xi < 0 \mid \mathcal{M}_1\right) \to 1$, and (b) $Pr\left(\xi > 0 \mid \mathcal{M}_2\right) \to 1$, when $n,T \to \infty$ under the asymptotics (14) with $\bar{\gamma} \leq 1$. It characterizes an asymptotically valid model selection rule, which treats both models symmetrically. The model selection rule is valid since (a) and (b) imply $Pr\left(\mathcal{M}_1|\xi < 0\right) = Pr\left(\xi < 0|\mathcal{M}_1\right) Pr\left(\mathcal{M}_1\right)\left[Pr\left(\xi < 0|\mathcal{M}_1\right) Pr\left(\mathcal{M}_1\right) + Pr\left(\xi < 0|\mathcal{M}_2\right) Pr\left(\mathcal{M}_2\right)\right]^{-1}$ $\to 1$, as $n,T \to \infty$, by Bayes Theorem. Similarly, we have $Pr\left(\mathcal{M}_2|\xi > 0\right) \to 1$. The diagnostic criterion (18) does not deliver a testing procedure since we do not use a critical region based on an asymptotic distribution and a chosen significance level. The zero threshold corresponds to an implicit critical value yielding a test size asymptotically equal to zero since $Pr(\xi < 0|\mathcal{M}_1) \to 1$. The selection procedure is conservative in diagnosing zero factor by construction. We do not allow type I error under $\mathcal{M}_1$ asymptotically, and really want to ensure that there is no omitted factor as required in the APT setting. It also means that we will not suffer from false discoveries related to a multiple testing problem (see e.g. Barras et al. (2010), Harvey et al. (2016)) in our empirical application where we consider a large variety of factor models on monthly and quarterly data. However, a possibility to achieve p-values is to use a randomisation procedure as in Trapani (2018) (see Bandi and Corradi (2014) and Corradi and Swanson (2006) for recent applications in econometrics). This type of procedure controls for an error of the first type, conditional on the information provided by the sample and under a randomness induced by auxiliary experiments.

The validity of the selection rule derives from the largest eigenvalue in (18) vanishing at a faster rate than the penalization term under $\mathcal{M}_1$ when n and T go to infinity. Under $\mathcal{M}_1$, we expect a vanishing largest eigenvalue because of a lack of a common signal in the error terms. The negative penalizing term $-g(n,T)$ dominates in (18), and it explains why we select the first model when ξ 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. Under $\mathcal{M}_2$, we have at least one non vanishing eigenvalue because of a common signal due to omitted factors. The largest eigenvalue dominates in (18), and it explains why we select the second model when ξ is positive. We can interpret the criterion (18) as the adjusted gain in fit including a single additional (unobservable) factor in model $\mathcal{M}_1$. We can rewrite (18) as $\xi = SS_0 - SS_1 - g(n,T)$, where $SS_0 = \frac{1}{nT}\sum_i \sum_t \mathbf{1}_i^{\chi} \bar{\varepsilon}_{i,t}^2$ is the sum of squared errors and $SS_1 = \min \frac{1}{nT}\sum_i \sum_t \mathbf{1}_i^{\chi} \left(\bar{\varepsilon}_{i,t} - \theta_i h_t\right)^2$, where the minimization is w.r.t. the vectors $H \in \mathbb{R}^T$ of factor values and $\Theta = (\theta_1, ..., \theta_n)' \in \mathbb{R}^n$ of factor loadings in a one-factor model, subject to the normalization constraint $\frac{H'H}{T} = 1$. Indeed, the largest eigenvalue $\mu_1 \left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i'\right)$ corresponds to the difference between SS_0 and SS_1. Furthermore, the criterion ξ is equal to the difference of the penalized criteria for zero- and one-factor models defined in Bai and Ng (2002) applied on the residuals. Indeed, $\xi = PC(0) - PC(1)$, where $PC(0) = SS_0$, and $PC(1) = SS_1 + g(n,T)$. It clarifies the relationship with the previous literature on latent factor selection. The diagnostic criterion builds on the same ideas but applied to the residuals, once we have filtered the impact of observable factors, instead of being directly computed on the original observed returns.

The proof of the validity of the selection rule exploits an asymptotic upper bound on the largest eigenvalue of a symmetric matrix based on similar arguments as in Geman (1980), Yin et al. (1988), and Bai and Yin (1993) without exploiting distributional results from random matrix theory valid when n is comparable with T. It exemplifies a key difference with the proportional asymptotics used in Onatski (2010) or Ahn and Horenstein (2013) for balanced panel without observable factors. The asymptotic setting accommodates the condition $T/n = o(1)$ by having $\bar{\gamma} < 1$ in (14), which agrees with the "large n, small T" case that we face in empirical applications (for example, ten thousand individual stocks monitored over forty-five years of either monthly, or quarterly, returns). Another key difference w.r.t. the rest of the literature is the handling of unbalanced panels. We need to address explicitly the presence of the observability indicators $I_{i,t}$ and the trimming devices $\mathbf{1}_i^\chi$ in the proofs of the asymptotic results.

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 et al. (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 smaller than the number m of the omitted factors under $\mathcal{M}_2$, it does not impact the asymptotic rate of the diagnostic criterion. If we only have useless factors in the omitted factors under $\mathcal{M}_2$, we face an identification issue. 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. In a "large n, large T" setting, the estimates of the risk premia are unchanged since we keep an approximate factor structure and risk remuneration is only attached to the strong factors in an APT framework. Here the presence of weak factors affects the pattern of the weak cross-sectional dependence and it only impacts variance estimator obtained by thresholding in the next section. On the contrary, if we have weak factors among the observable factors, Anatolyev and Mikusheva (2018) show that the conventional two-pass estimation procedure delivers inconsistent estimates of the risk premia. In the time-invariant case, they propose a modified procedure based on sample-splitting instrumental variables estimation at the second pass, and examine its asymptotic distribution.

Several papers in the empirical asset pricing literature focus on distinguishing between useful, useless and redundant factors starting from different points of view. Bryzgalova (2016) develops a shinkrage-based estimator that identify the weak factors (i.e., factors that do not correlate with the assets) and ensure consistent and normality to the estimates of the risk premia. Feng et al. (2020) propose a model-selection method to evaluate the risk prices of observable factors. Freyberger et al. (2020) propose a nonparametric method to determine which firm characteristics provide incremental information for the cross section of expected returns. Kozak et al. (2018) use model selection techniques to identify characteristics portfolios with a good explanatory power for returns. These papers do not deal with the identification of systematic factors for which the errors are weakly cross-sectionally correlated. The model selection procedure is not able to answer our key question on the presence of omitted factors in the chosen specification.

In the previous lines, 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. Hereafter we aim 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 (12), where the errors $(\varepsilon_{i,t})$ satisfy a factor structure with exactly k unobservable factors,

and

$\mathcal{M}_2(k)$: the linear regression model (12), where the errors $(\varepsilon_{i,t})$ satisfy a factor structure with at least $k+1$ unobservable factors.

The above definitions yield $\mathcal{M}_1 = \mathcal{M}_1(0)$ and $\mathcal{M}_2 = \mathcal{M}_2(0)$. The diagnostic criterion exploits the $(k+1)$th largest eigenvalue of the empirical cross-sectional covariance matrix of the residuals:

$$\xi(k) = \mu_{k+1}\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i \bar{\varepsilon}_i'\right) - g(n,T). \tag{19}$$

As discussed in Ahn and Horenstein (2013) (see also Onatski (2013)) for balanced panels, we can rewrite (19) as $\xi(k) = SS_k - SS_{k+1} - g(n,T)$ where $SS_k = \min \frac{1}{nT}\sum_i \sum_t \mathbf{1}_i^\chi \left(\bar{\varepsilon}_{i,t} - \theta_i' h_t\right)^2$ and the minimization is w.r.t. $H \in \mathbb{R}^{T\times k}$ and $\Theta = (\theta_1, ..., \theta_n)' \in \mathbb{R}^{n\times k}$. 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 (2002) applied on the residuals. Indeed, $\xi(k) = PC(k) - PC(k+1)$, where $PC(k) = SS_k + kg(n,T)$ and $PC(k+1) = SS_{k+1} + (k+1)g(n,T)$.

The following model selection rule extends the previous one. We select $\mathcal{M}_1(k)$ if $\xi(k) < 0$, and we select $\mathcal{M}_2(k)$ if $\xi(k) > 0$, since (a) $Pr[\xi(k) < 0|\mathcal{M}_1(k)] \to 1$ and (b) $Pr[\xi(k) > 0|\mathcal{M}_2(k)] \to 1$, when $n, T \to \infty$.

The proof of the validity of that second selection rule is more complicated than the proof of the first one. We need additional arguments to derive an asymptotic upper bound when we look at the $(k+1)$th eigenvalue of a symmetric matrix, and this further complexity explains why we have developed the first selection rule as a special case. We rely on the Courant-Fischer min-max theorem and Courant-Fischer formula which represent eigenvalues as solutions of constrained quadratic optimization problems. We cannot directly exploit standard inequalities or bounds associated to a norm when we investigate the asymptotic behavior of the spectrum beyond its largest element. We know that the largest eigenvalue $\mu_1(A)$ of a symmetric positive semi-definite matrix A is equal to its operator norm. There is no such useful norm interpretation for the smaller eigenvalues $\mu_k(A)$, $k \geq 2$. We cannot either exploit distributional results from random matrix theory since we also allow for $T/n = o(1)$. The slow convergence rate $\sqrt{T}$ for the individual estimates $\hat{\beta}_i$ also complicates the proof. In the presence of homogeneous regression coefficients $\beta_i = \beta$ for all i, the estimate $\hat{\beta}$ in Bai (2009) and Moon and Weidner (2015) has a fast convergence rate $\sqrt{nT}$. In that case, controlling for the estimation error in $\hat{\varepsilon}_{i,t} = \varepsilon_{i,t} + x'_{i,t}(\beta - \hat{\beta})$ is straightforward due to the small asymptotic contribution of $(\beta - \hat{\beta})$. Hence our results also apply to diagnose the absence of omitted interactive effects before applying a difference-in-differences procedure to avoid bias. The approach of Onatski (2010) requires the convergence of the upper edge of the spectrum (i.e., the first k largest eigenvalues of the covariance matrix, with $k/T = o(1)$) to a constant, while the approach of Ahn and Horenstein (2013) requires an asymptotic lower bound on the eigenvalues. Extending these approaches for residuals of an unbalanced panel when $T/n = o(1)$ looks challenging.

We can use the results of the selection rule in order to estimate the number of unobservable factors. It suffices to choose the minimum k such that $\xi(k) < 0$. We get the consistency of that estimate even in the presence of a degenerate distribution of the eigenvalues, and without needing to give conditions on the growth rate of the maximum possible number $kmax$ of factors as in Onatski (2010) and Ahn and Horenstein (2013). We believe that it is a strong advantage since there are many possible choices for $kmax$ and the estimated number of factors is sometimes sensitive to the choice of $kmax$ (see the simulation results in those papers).

4.2 Estimation of conditional risk premia

In the linear regression (12), the coefficients associated to $x_{1,i,t}$ and $x_{2,i,t}$ are $\beta_i = \left(\beta_{1,i}', \beta_{2,i}'\right)'$ such that

$$\beta_{1,i} = \left(\left(N_p \left[(\Lambda - F)' \otimes I_p\right] vec\left[B_i'\right]\right)', \left(\left[(\Lambda - F)' \otimes I_q\right] vec\left[C_i'\right]\right)'\right)', \qquad N_p = \frac{1}{2} D_p^{+}(W_p + I_{p^2}),$$
$$\beta_{2,i} = \left(vec\left[B_i'\right]', vec\left[C_i'\right]'\right)', \tag{20}$$

where parameter matrices $B_i = B(\gamma_i)$, $C_i = C(\gamma_i)$, Λ and F are defined in (10). The vector operator $vec\left[\cdot\right]$ stacks the elements of a $m \times n$ matrix as a $mn \times 1$ vector. The matrix D_p^{+} is the $p(p+1)/2 \times p^2$ Moore-Penrose inverse of the duplication matrix D_p, such that $vech\left[A\right] = D_p^{+} vec\left[A\right]$ for any matrix $A \in \mathbb{R}^{p \times p}$ (see Chapter 3 in Magnus and Neudecker (2007)). The commutation matrix $W_{p,q}$ is such that $vec[A'] = W_{p,q} vec[A]$, for any matrix $A \in \mathbb{R}^{p \times q}$, and $W_p := W_{p,p}$. When $Z_t = 1$ and $Z_{i,t} = 0$, we have $p = 1$ and $q = 0$, and the model in (12) reduces to a factor model with time-invariant coefficients and regressor x_t common across assets (scaled factors).

In Equations (20), the $d_1 \times 1$ vector $\beta_{1,i}$ is a linear transformation of the $d_2 \times 1$ vector $\beta_{2,i}$. It clarifies that the asset pricing restriction (8) implies a constraint on the distribution of the random vector β_i via its support. The coefficients of the linear transformation depend on matrix $\Lambda - F$. For the purpose of estimating the loading coefficients of the risk premia in matrix Λ, we rewrite the parameter restrictions as:

$$\beta_{1,i} = \beta_{3,i}\nu, \qquad \nu = vec\left[\Lambda' - F'\right], \qquad \beta_{3,i} = \left(\left[N_p\left(B_i' \otimes I_p\right)\right]', \left[W_{p,q}\left(C_i' \otimes I_p\right)\right]'\right)'. \tag{21}$$

Furthermore, we can relate the $d_1 \times Kp$ matrix $\beta_{3,i}$ to the vector $\beta_{2,i}$:

$$vec\left[\beta_{3,i}'\right] = J_a \beta_{2,i}, \tag{22}$$

where the $d_1 pK \times d_2$ block-diagonal matrix of constants J_a is given by $J_a = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix}$ with diagonal blocks $J_1 = W_{p(p+1)/2,pK}\left(I_K \otimes \left[\left(I_p \otimes N_p\right)\left(W_p \otimes I_p\right)\left(I_p \otimes vec\left[I_p\right]\right)\right]\right)$ and $J_2 = W_{pq,pK}\left(I_K \otimes \left[\left(I_p \otimes W_{p,q}\right)\left(W_{p,q} \otimes I_p\right)\left(I_q \otimes vec\left[I_p\right]\right)\right]\right)$. The no-arbitrage restrictions (21) and the link (22) induced by the conditional setting were unknown to the literature and are instrumental for the derivation of the asymptotic theory and for empirics. They hold in the population, and we should use them for inference with fixed n and large T as well. For the fixed n case, it yields a different estimation procedure compared to those reviewed in Section 2. In fact, the GMM estimator reviewed in Section 2 relies on a conditionally linear specification of the risk prices $\delta_{1,t}$. No parametric specification for the betas is required, however we need to estimate the conditional variance of the factor $V_t[f_{t+1}]$ to get an estimate of the conditional risk premia (see Section 2). Instead, the approach reviewed in this section relies on linear specifications for the risk premia, the betas, and the conditional expectation of the factor (see (10)), but no specification for the conditional variance of the factor is needed. In the time-invariant setting, $\beta_{1,i} = a_i$, $\beta_{2,i} = \beta_{3,i} = b_i$, and the matrix J is equal to I_K. Hence, Equations (21) and (22) in the time-varying case are the counterparts of restriction $a_i = b_i'\nu$ in the time-invariant case.

Let us now describe the two-pass approach to estimate the factor risk premia. The first pass consists in computing time-series OLS estimators $\hat{\beta}_i$, and was described in the previous subsection (see Equation (17)). The second pass consists in computing a cross-sectional estimator of ν by regressing the $\hat{\beta}_{1,i}$ on the $\hat{\beta}_{3,i}$ keeping non-trimmed assets only. We use a multivariate WLS approach. The weights are estimates of $w_i = (diag\,[v_i])^{-1}$, where the matrices v_i are the asymptotic variances of the standardized errors $\sqrt{T}\left(\hat{\beta}_{1,i} - \hat{\beta}_{3,i}\nu\right)$ in the cross-sectional regression for large T. We have $v_i = \tau_i C_\nu' Q_{x,i}^{-1} S_{ii} Q_{x,i}^{-1} C_\nu$, where $Q_{x,i} = E\left[x_{i,t}x_{i,t}'|\gamma_i\right]$, $S_{ii} = \underset{T\to\infty}{\text{plim}} \frac{1}{T}\sum_t \sigma_{ii,t} x_{i,t} x_{i,t}' = E\left[\varepsilon_{i,t}^2 x_{i,t} x_{i,t}'|\gamma_i\right]$, $\tau_i = \underset{T\to\infty}{\text{plim}}\, \tau_{i,T} = E\left[I_{i,t}|\gamma_i\right]^{-1}$, $C_\nu = \left(E_1' - \left(I_{d_1} \otimes \nu'\right) J_a E_2'\right)'$, with $E_1 = (I_{d_1} : 0_{d_1 \times d_2})'$, $E_2 = (0_{d_2 \times d_1} : I_{d_2})'$. We use the estimates $\hat{v}_i = \tau_{i,T} C_{\hat{\nu}_1}' \hat{Q}_{x,i}^{-1} \hat{S}_{ii} \hat{Q}_{x,i}^{-1} C_{\hat{\nu}_1}$, where $\hat{S}_{ii} = \frac{1}{T_i}\sum_t I_{i,t}\hat{\varepsilon}_{i,t}^2 x_{i,t} x_{i,t}'$, $\hat{\varepsilon}_{i,t} = R_{i,t} - \hat{\beta}_i' x_{i,t}$ and $C_{\hat{\nu}_1} = \left(E_1' - \left(I_{d_1} \otimes \hat{\nu}_1'\right) J_a E_2'\right)'$. To estimate C_ν, we use the multivariate OLS estimator $\hat{\nu}_1 = \left(\sum_i \mathbf{1}_i^\chi \hat{\beta}_{3,i}' \hat{\beta}_{3,i}\right)^{-1} \sum_i \mathbf{1}_i^\chi \hat{\beta}_{3,i}' \hat{\beta}_{1,i}$, i.e., a first-step estimator with unit weights. The WLS estimator is:

$$\hat{\nu} = \hat{Q}_{\beta_3}^{-1} \frac{1}{n} \sum_i \hat{\beta}_{3,i}' \hat{w}_i \hat{\beta}_{1,i}, \tag{23}$$

where $\hat{Q}_{\beta_3} = \frac{1}{n}\sum_i \hat{\beta}_{3,i}' \hat{w}_i \hat{\beta}_{3,i}$ and $\hat{w}_i = \mathbf{1}_i^\chi \left(diag\,[\hat{v}_i]\right)^{-1}$. Weighting accounts for the statistical precision of the first-pass estimates and includes trimming. The final estimator of the risk premia is $\hat{\lambda}_t = \hat{\Lambda} Z_{t-1}$, where we deduce $\hat{\Lambda}$ from the relationship $vec\left[\hat{\Lambda}'\right] = \hat{\nu} + vec\left[\hat{F}'\right]$ with the estimator $\hat{F}$ obtained by a SUR regression of factors f_t on lagged instruments Z_{t-1}: $\hat{F} = \sum_t f_t Z_{t-1}' \left(\sum_t Z_{t-1} Z_{t-1}'\right)^{-1}$.

In the time-invariant case, the estimator of the risk premia vector simplifies to

$$\hat{\lambda} = \hat{\nu} + \frac{1}{T}\sum_t f_t, \qquad \hat{\nu} = \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i \hat{b}_i \hat{a}_i, \tag{24}$$

where $\hat{Q}_b = \frac{1}{n}\sum_i \hat{w}_i \hat{b}_i \hat{b}_i'$ and $(\hat{a}_i, \hat{b}_i')' = \hat{Q}_{x,i}^{-1}\frac{1}{T_i}\sum_t I_{i,t}x_t R_{i,t}$. Hence, we estimate the model coefficients a_i and b_i by time series OLS regression, and the risk premium by cross-sectional WLS regression of the $\hat{a}_i$s on the $\hat{b}_i$s augmented by the factor mean. Moreover, under conditional homoskedasticity $\sigma_{ii,t} = \sigma_{ii}$ and a balanced panel $\tau_{i,T} = 1$, we have $v_i = c_\nu' Q_x^{-1} c_\nu \sigma_{ii}$, where $c_\nu = (1, -\nu')'$ and $Q_x = E[x_t x_t']$. Then, v_i is directly proportional to σ_{ii}, and we can simply pick the weights as $\hat{w}_i = \hat{\sigma}_{ii}^{-1}$, where $\hat{\sigma}_{ii} = \frac{1}{T}\sum_t \hat{\varepsilon}_{i,t}^2$ (Shanken (1992)). In the time-invariant case, we can avoid the trimming on the condition number if we substitute $\hat{Q}_x = \frac{1}{T}\sum_t x_t x_t$ for $\hat{Q}_{x,i}$ in the first-pass estimator definition. However, it increases the asymptotic variance of the bias corrected estimator of ν, and does not extend to the time-varying case. Starting from the asset pricing restriction $E[R_{i,t}] = b_i'\lambda$ in the time-invariant case, another estimator of λ is $\bar{\lambda} = \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i \hat{b}_i \bar{R}_i$, where $\bar{R}_i = \frac{1}{T_i}\sum_t I_{i,t}R_{i,t}$. This estimator is numerically equivalent to $\hat{\lambda}$ in the balanced case, where $I_{i,t} = 1$ for all i and t. In the unbalanced case, it is equal to $\bar{\lambda} = \hat{\nu} + \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i \hat{b}_i \hat{b}_i' \bar{f}_i$, where $\bar{f}_i = \frac{1}{T_i}\sum_t I_{i,t} f_t$. Estimator $\bar{\lambda}$ is often studied by the literature (see, e.g., Shanken (1992), Kandel and Stambaugh (1995), Jagannathan and Wang (1998)), and is also consistent. Estimating $E[f_t]$ with a simple average of the observed factor instead of a weighted average based on estimated betas simplifies the form of the asymptotic distribution in the unbalanced case. It explains our preference for $\hat{\lambda}$ over $\bar{\lambda}$.

We get consistency and asymptotic normality of $\hat{\nu}$ and $\hat{\Lambda}$ under the asymptotics in (14) with $1/\underline{\gamma} < 3$. The estimator $\hat{\nu}$ has a fast convergence rate $\sqrt{nT}$ and features an asymptotic bias term: $\sqrt{nT}\left(\hat{\nu} - \nu - \hat{B}_\nu/T\right) \Rightarrow N(0, \Sigma_\nu)$. Both $\hat{\beta}_{1,i}$ and $\hat{\beta}_{3,i}$ in the definition of $\hat{\nu}$ contain an estimation error; for $\hat{\beta}_{3,i}$, it is the well-known Error-In-Variable (EIV) problem. The EIV problem does not impede consistency since we let T grow to infinity. However, it induces a bias term $\hat{B}_\nu/T$ which centers the asymptotic distribution of $\hat{\nu}$. Ang, Liu, and Schwarz (2020) look at a maximum likelihood analysis with a single asymptotic treatment (large T, n fixed) and balanced panel under a particular approximate Gaussian factor structure (block diagonal covariance matrix of residuals) and time-invariant coefficients. Their setting further assumes that the factors have zero mean. Such an assumption gives $\hat{\lambda} = \hat{\nu}$ in a time-invariant setting.

Under a zero mean (or a known mean, i.e., not to be estimated), the asymptotic variance of $\hat{\lambda}$ corresponds to the asymptotic variance Σ_ν of $\hat{\nu}$ and the rate of convergence is $\sqrt{nT}$. On the contrary, if we do not know the mean of the factor and need to estimate it, we have $\hat{\lambda} = \hat{\nu} + \frac{1}{T}\sum f_t$. The asymptotic variance of $\hat{\lambda}$ corresponds to the asymptotic variance Σ_f of the sample average of the factors, and the rate of convergence is $\sqrt{T}$. Jagannathan and Wang (2002) is an early reference on the impact of knowing or not the mean of the factors for asymptotic analysis. With an unknown mean, only the variability of the factor drives the asymptotic distribution of $\hat{\lambda}$, since the estimation error $O_p\left(1/\sqrt{T}\right)$ of the sample average $\frac{1}{T}\sum_t f_t$ dominates the estimation error $O_p\left(1/\sqrt{nT} + 1/T\right)$ of $\hat{\nu}$. This result is an oracle property for $\hat{\lambda}$, namely that its asymptotic distribution is the same irrespective of the knowledge of ν. This property is in sharp difference with the single asymptotics with a fixed n and $T \to \infty$. In the balanced case and with homoskedastic errors for the time-invariant case, Theorem 1 of Shanken (1992) shows that the rate of convergence of $\hat{\lambda}$ is $\sqrt{T}$ and that its asymptotic variance is $\Sigma_{\lambda,n} = \Sigma_f + \frac{1}{n}\Sigma_{\nu,n}$, for fixed n and $T \to \infty$. The two components in $\Sigma_{\lambda,n}$ come from estimation of $E[f_t]$ and ν, respectively (see also Theorem 1 in Jagannathan and Wang (1998), or Theorem 3.2 in Jagannathan, Skoulakis, and Wang (2009)). Letting $n \to \infty$ gives Σ_f under weak cross-sectional dependence. Thus, exploiting the full cross-section of assets improves efficiency asymptotically, and the positive definite matrix $\Sigma_{\lambda,n} - \Sigma_f$ corresponds to the efficiency gain. Using a large number of assets instead of a small number of portfolios does help to eliminate the contribution coming from estimation of ν. The use of portfolios is often motivated by their intrinsic construction yielding stable betas and thus mitigating the contribution of the additional component $\frac{1}{n}\Sigma_{\nu,n}$. We see that this contribution vanishes when we have a large pool of test assets, and we often use portfolios for the wrong reason.

We can exploit the analytical bias correction $\hat{B}_\nu/T$ and use the estimator $\hat{\nu}_B = \hat{\nu} - \frac{1}{T}\hat{B}_\nu$ instead of $\hat{\nu}$. In the time-invariant setting, $\hat{\lambda}_B = \hat{\nu}_B + \frac{1}{T}\sum_t f_t$ delivers a bias-free estimator of λ at order $1/T$, which shares the same root-T asymptotic distribution as $\hat{\lambda}$. We can relate that suggestion to bias-corrected estimation accounting for the well-known incidental parameter problem (Neyman and Scott (1948)) in the panel literature (see Lancaster (2000) for a review). To highlight the main idea, let us focus on the model with time-invariant coefficients. We can write the factor model under restriction $a_i = b_i'\nu$ as $R_{i,t} = b_i'(f_t + \nu) + \varepsilon_{i,t}$. In the likelihood setting of Hahn and Newey (2004) (see also Hahn and Kuersteiner (2002)), the b_is correspond to the individual fixed effects and ν to the common parameter of interest. Here, the individual effects enter multiplicatively instead of additively as in a standard panel. Available results on the fixed-effect approach tell us: (i) the ML estimator of ν is inconsistent if n goes to infinity while T is held fixed, (ii) the ML estimator of ν is asymptotically biased even if T grows at the same rate as n, (iii) an analytical bias correction may yield an estimator of ν that is root-(nT) asymptotically normal and centered at the truth if T grows faster than $n^{1/3}$. The two-pass estimators $\hat{\nu}$ and $\hat{\nu}_B$ exhibit the properties (i)-(iii) as expected by analogy with unbiased estimation in large panels. This clear link with the incidental parameter literature highlights another advantage of working with ν in the second-pass regression. Chamberlain (1992) considers a general random coefficient model nesting the factor model with time-invariant coefficients. He establishes asymptotic normality of an estimator of ν for fixed T and balanced panel data. His estimator does not admit a closed-form and requires a numerical optimization. It leads to computational difficulties in the conditional setting. It also makes the study of his estimator under double asymptotics and cross-sectional dependence challenging. Recent advances on the incidental parameter problem in random coefficient models for fixed T are Arellano and Bonhomme (2012) and Bonhomme (2012).

Finally, let us discuss confidence intervals. Their construction for components of $\hat{\Lambda}$ to achieve valid asymptotic coverage is straightforward through the use of standard HAC estimators such as in Newey and West (1994) or Andrews and Monahan (1992). The construction of confidence intervals for the components of $\hat{\nu}$ is more difficult. Indeed, the asymptotic variance involves a limiting double cross-sectional sum scaled by n and not n^2. A naive approach consists in replacing unknown quantities by any consistent estimator, but it does not work here. To handle this, we can rely on recent proposals in the statistical literature on consistent estimation of large dimensional sparse covariance matrices by hard thresholding (Bickel and Levina (2008), El Karoui (2008)). Fan, Liao, and Mincheva (2011) focus on the estimation of the variance-covariance matrix of the errors in large balanced panel with nonrandom time-invariant coefficients and i.i.d. disturbances. Another possibility is to rely on a multiple testing approach to find the sparsity structure. Bailey et al. (2019) develop an estimator which consistently recovers the support of the population covariance matrix under Gaussian and non-Gaussian observations, and show that the true positive rate tends to one with probability 1, and the false positive rate and the false discovery rate tend to zero with probability 1, even if n tends to infinity faster than T.

Besides, if we face a short time series panel (for example a 5-year window), and without the availability of high-frequency data (see the discussion at the end of Section 5), asymptotics with fixed T and large n are better suited. In this context, assuming time-invariant factor loadings and risk premia is coherent with the data time span. Here, keeping T fixed impedes consistent estimation of the risk premia, and inference has to focus on ex-post risk premia (Shanken (1992)). Ex-post risk premia differ from the risk premia we have defined in the theory above and we estimate in the empirics below. Kim and Skoulakis (2018), Raponi et al. (2020) propose methodologies for estimating with observable factors and evaluating asset pricing models on balanced panels when T is fixed and n is large. Kim and Skoulakis (2018) employ the regression calibration approach used in EIV models to derive a $\sqrt{n}$-consistent estimator of ex-post risk premia in a two-pass cross-sectional regression setting. Raponi et al. (2020) propose a consistent and asymptotically normally distributed estimator of ex-post risk premia following the bias-adjusted estimator of Shanken (1992). They discuss extensions of their results to time variation in risk premia and factor loadings, potential misspecification, and unbalanced panels.

4.3 Testing asset pricing restrictions

From (21), the null hypothesis underlying the asset pricing restriction (8) is

$$\mathcal{H}_0: \text{ there exists } \nu \in \mathbb{R}^{pK} \text{ such that } \beta_1(\gamma) = \beta_3(\gamma)\nu, \text{ for almost all } \gamma \in [0,1],$$

where $\beta_1(\gamma)$ and $\beta_3(\gamma)$ are defined as $\beta_{1,i}$ and $\beta_{3,i}$ in Equations (20) and (21) replacing $B(\gamma)$ and $C(\gamma)$ for B_i and C_i. This null hypothesis is written on the continuum of assets. Under $\mathcal{H}_0$, we have $E\left[(\beta_{1,i} - \beta_{3,i}\nu)'(\beta_{1,i} - \beta_{3,i}\nu)\right] = 0$. Since we estimate ν via the WLS cross-sectional regression of the estimates $\hat{\beta}_{1,i}$ on the estimates $\hat{\beta}_{3,i}$, we can use a test based on the weighted sum of squared residuals SSR of the cross-sectional regression. The weighted SSR is $\hat{Q}_e = \frac{1}{n}\sum_i \hat{e}_i'\hat{w}_i\hat{e}_i$, with $\hat{e}_i = \hat{\beta}_{1,i} - \hat{\beta}_{3,i}\hat{\nu} = C_{\hat{\nu}}'\hat{\beta}_i$, which is an empirical counterpart of $E\left[(\beta_{1,i} - \beta_{3,i}\nu)' w_i (\beta_{1,i} - \beta_{3,i}\nu)\right]$. Let us now introduce the following statistic $\hat{\xi}_{nT} = T\sqrt{n}\left(\hat{Q}_e - \frac{1}{T}\hat{B}_\xi\right)$, where the recentering term simplifies to $\hat{B}_\xi = d_1$ thanks to the weighting scheme. Under the null hypothesis $\mathcal{H}_0$ and asymptotics (14) with $1/\underline{\gamma} < 2$, we have $\hat{\xi}_{nT} \Rightarrow N(0, \Sigma_\xi)$, and get a feasible testing procedure by exploiting a consistent estimate of the asymptotic variance Σ_ξ.

Finally, we can derive a test for the null hypothesis when the factors come from tradable assets, i.e., are portfolio excess returns:

$$\mathcal{H}_0: \quad \beta_1(\gamma) = 0 \text{ for almost all } \gamma \in [0,1],$$

against the alternative hypothesis

$$\mathcal{H}_1: \quad E\left[\beta_{1,i}'\beta_{1,i}\right] > 0.$$

We only have to substitute $\hat{Q}_a = \frac{1}{n}\sum_i \hat{\beta}_{1,i}'\hat{w}_i\hat{\beta}_{1,i}$ for $\hat{Q}_e$. Since the constrained form of $\beta_{1,i}$ in (21) comes from (8), we take directly into account the no-arbitrage restrictions imposed by the model specification. It gives an extension of Gibbons, Ross and Shanken (1989) to the conditional case with double asymptotics. Implementing the original GRS test, which uses a weighting matrix corresponding to an inverted estimated large variance-covariance matrix, becomes quickly problematic. We face a large number nd_1 of restrictions; each $\beta_{1,i}$ is of dimension $d_1 \times 1$, and the estimated covariance matrix to invert is of dimension $nd_1 \times nd_1$. We expect to compensate the potential loss of power induced by a diagonal weighting via the larger number of restrictions since we use a large number n of assets. Monte Carlo simulations show that the test exhibits good power properties against both risk-based and non risk-based alternatives (e.g. MacKinlay (1995)) already for a thousand assets with a time series dimension similar to the one in the empirical analysis. Fan et al. (2015) discuss power enhancement in high dimensional cross-sectional tests.

Finally, let us mention that Ma et al. (2020) has recently developed a test of the nullity of the alphas when the alphas and betas are taken as smooth functions of time in a "large n, large T" setting (see Li and Yang (2011) and Ang and Kristensen (2012) for the "small n, large T" case). For fixed T and large n, Kim and Skoulakis (2018) develop inference for a test statistic of the asset pricing restrictions based on a measure of aggregate mispricing (GRS type test based on their regression calibration approach).

5 Inference in models with unobservable factors

In this section, we review methodologies for inference in large dimensional conditional factor models when the factor values are unobserved by the econometrician. In this setting, we cannot use standard Principal Component Analysis (PCA) to extract the factor space since PCA assumes either constant factor loadings (Stock and Watson (2002a,b), Bai (2003, 2009), Bai and Ng (2002, 2006)) or at most small instabilities in the factors loadings (Bates et al. (2013)). Intuitively, invalidity of standard PCA in a conditional framework comes from a factor with time-varying loading being potentially confused with multiple static factors.

The model specification is:

$$R_{i,t} = a_{i,t} + b'_{i,t} f_t + \varepsilon_{i,t}, \tag{25}$$

where f_t is the K-dimensional vector of the unobservable factor values. Several estimation approaches are based on assuming that the intercepts $a_{i,t}$ and the factor loadings $b_{i,t}$ are either parametric or nonparametric functions of lagged time-varying observable variables, with or without imposing the no arbitrage restrictions. Among the parametric approaches, Kelly et al. (2017, 2019) model the coefficients as linear functions of characteristics plus some noise term:

$$\begin{aligned} a_{i,t} &= A' Z_{i,t-1} + \nu_{i,t}, & (26) \\ b_{i,t} &= B' Z_{i,t-1} + \eta_{i,t}, & (27) \end{aligned}$$

where $Z_{i,t}$ is a vector of observed characteristics, A and B are a vector and a matrix of unknown parameters, and $\nu_{i,t}$ and $\eta_{i,t}$ are unobservable noise terms. By plugging (26) and (27) into (25), we get $R_{i,t} = Z'_{i,t-1}A + Z'_{i,t-1}Bf_t + \varepsilon^*_{i,t}$, where the composite error term is $\varepsilon^*_{i,t} = \varepsilon_{i,t} + \nu_{i,t} + \eta'_{i,t} f_t$. The Instrumented Principal Component Analysis (IPCA) estimator of Kelly et al. (2017, 2019) is obtained by minimizing a LS criterion w.r.t. parameter matrices A, B and the factor values f_t, $t = 1, ..., T$, i.e., $\min_{A,B,f_t,t=1,...,T} \sum_{i=1}^{n} \sum_{t=1}^{T} (R_{i,t} - Z'_{i,t-1}A - Z'_{i,t-1}Bf_t)^2$ subject to the static normalization restrictions that the matrix BB' is diagonal, and $\frac{1}{T} \sum_{t=1}^{T} f_t = 0$, $\frac{1}{T} \sum_{t=1}^{T} f_t f'_t = I_K$. They propose an iterative numerical procedure to perform the optimization.

In the nonparametric setting, an early contribution is provided by an extension of the model considered by Connor and Linton (2007) and Connor et al. (2012), in which the factor loadings are functions of observed covariates:

$$R_{i,t} = \sum_{k=1}^{K} b_k(Z_{k,i,t-1}) f_{k,t} + \varepsilon_{i,t}, \tag{28}$$

where $b_k(\cdot)$ is an unknown smooth function of observable variable $Z_{k,i,t-1}$, for $k = 1, ..., K$ (see Connor et al. (2012) p. 728 for a short discussion of this extension; their base model assumes that the covariates are time-invariant). The identification scheme requires that the characteristics differ across factors. Indeed, Connor et al. (2012) estimate model (28) by deploying the property that it corresponds to an additive non-parametric regression at each date t. They propose an iterative procedure that alternates at each step the cross-sectional estimation of (i) the loadings functions via the backfitting projection algorithm, and (ii) the factor values by least-squares regression, subject to normalization restrictions. They obtain the final estimates of the loadings functions $b_k(\cdot)$ by averaging across time the cross-sectional estimates. Fan et al. (2016b) extend the characteristic-based modeling in Connor and Linton (2007) and Connor et al. (2012) by allowing the betas $b_k(Z_i) + \gamma_{i,k}$ to include unknown asset-specific additive constants (see Liao and Yang (2018) for the continuous-time case under infill asymptotics for high frequency data). They propose a so-called Projected PCA method to estimate this specification with time-invariant loadings. It is an open question whether we can extend this estimation approach to accommodate time variation in the characteristics. Pelger and Xiong (2019) instead let the factor loadings be functions of an observable state variable. They consider the model:

$$R_{i,t} = \sum_{k=1}^{K} b_{i,k}(Z_{t-1}) f_{k,t} + \varepsilon_{i,t}, \tag{29}$$

where the $b_{i,k}(\cdot)$ are smooth functions and Z_t is a vector of observed variables, common across assets. Pelger and Xiong (2019) estimate model (29) by minimizing a local version of the least-squares criterion underlying PCA, where localization is implemented by kernel smoothing. In practice, the number of conditioning variables, which we can accommodate, is small. Su and Wang (2017) develop a similar approach with piece-wise smooth functions $b_{i,k}(\cdot)$ of t/T on $(0, 1]$. They find strong evidence of structural changes in the factor loadings in U.S. macroeconomic data. A literature related to such a specification stems from large dimensional factor models with structural instabilities, including e.g. Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015). In those papers, the loadings may have a small number of large changes (structural breaks). Cheng et al. (2016) further allow for changes in the number of factors and the space spanned by the loadings.

Among the nonparametric approaches, some recent work takes advantage of machine learning methods to achieve greater flexibility in the modeling of time-varying betas and accommodate the large dimensionality of the set of potential characteristics and state variables. Gu et al. (2019, forthcoming) consider the setting where the loadings are a nonparametric function of a large dimensional vector of characteristics: $b_{i,t} = b(Z_{i,t-1})$, and use an autoencoder to estimate this relationship. Autoencoder is a class of universal approximators in the realm of Artificial Neural Networks (see Gu et al. (2019, forthcoming) and references therein). Using L_b hidden layers and an activator function g, each component of the loadings vector is approximated as:

$$\begin{aligned}
b_{i,t,k}(\theta_b) &= G_k(Z_{i,t-1}, L_b, g, \theta_b) := A_k + B_k' Z_{i,t-1}^{(L)}, \\
Z_{i,t-1,j}^{(\ell)} &= g\left(A_j^{(\ell-1)} + B_j^{(\ell-1)\prime} Z_{i,t-1}^{(\ell-1)}\right),\ \ell = 1, ..., L, \\
Z_{i,t-1}^{(0)} &= Z_{i,t-1},
\end{aligned}$$

where the parameter vector θ_b includes the A_k, B_k, $A_j^{(\ell-1)}$, $B_j^{(\ell-1)}$ for all k, j, ℓ. Gu et al. (2019, forthcoming) approximate the factor values also with an autoencoder as $f_{t,k}(\theta_f) = G_k(\hat{x}_t, L_f, g, \theta_f)$ using as input the standardized cross-sectional averages $\hat{x}_t = \left(\frac{1}{n}\sum_i Z_{i,t-1}Z_{i,t-1}'\right)^{-1}\frac{1}{n}\sum_i Z_{i,t-1}R_{i,t}$, i.e., characteristics-based portfolio returns. They minimize the penalized least-squares criterion $\min_\theta \sum_i \sum_t \left(R_{i,t} - b_{i,t}(\theta_b)' f_t(\theta_f)\right)^2 + \lambda\|\theta\|_1$, where $\|\theta\|_1$ is the L^1 norm of the parameter vector $\theta = (\theta_b', \theta_f')'$. The inferential theory for this estimator is unknown.

In the rest of this section, we review a recent proposal for inference in time-varying statistical factor models developed by Gagliardini and Ma (2019). As the focus of these authors is on the problem of conducting inference on the conditional factor space, including its dimension, the adopted nonparametric framework is general regarding the beta dynamics and encompasses the linear and nonlinear beta specifications of e.g. Kelly et al. (2017, 2019), and Gu et al. (2019, forthcoming). The framework allows for time variation in the number of conditional factors as an effect of the changing macroeconomic environment. The main idea is to see the estimation of factor values as a cross-sectional Instrumental Variable (IV) problem and deploy a well-chosen (conditional) normalization of the factor vector to accommodate an essentially unspecified beta dynamics. Here we review the main results in the simpler framework with constant number of conditional factors, and refer the reader to Gagliardini and Ma (2019) for the more general setting, the regularity conditions, and the derivation of the results.

After imposing the no arbitrage restrictions $a_{i,t} = b_{i,t}'\nu_t$ (see Equation (8)), model (25) becomes: $R_{i,t} = b_{i,t}'g_t + \varepsilon_{i,t}$, where $g_t = f_t + \nu_t$. Gagliardini and Ma (2019) assume that the m-dimensional lagged instruments $Z_{i,t-1}$ are cross-sectionally uncorrelated with errors and correlated with betas under a full rank condition:

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_i Z_{i,t-1}\varepsilon_{i,t} &= 0,\\ \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_i Z_{i,t-1}b'_{i,t} &=: \Gamma_t \text{ is a full column-rank matrix,}\end{aligned} \tag{30}$$

for all t, which implies the order condition $m \geq K$. Being the limit of a cross-sectional average of predetermined variables, the matrix Γ_t is measurable w.r.t. the information set $\mathcal{G}_{t-1}$ of aggregate shocks at time $t-1$, i.e., the non-diversifiable shocks (see Gagliardini and Ma (2019) for more details). It is assumed that $\mathcal{G}_t$ is generated by the vector process Z_t, and that the econometrician observes Z_t. Under (30), it holds:

$$\xi_t := \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_i Z_{i,t-1}R_{i,t} = \Gamma_t g_t. \tag{31}$$

Process ξ_t is identifiable from population moments. Its conditional variance given $\mathcal{G}_{t-1}$ is $V[\xi_t|\mathcal{G}_{t-1}] = \Gamma_t V[g_t|\mathcal{G}_{t-1}]\Gamma'_t$. Thus, the number of non-zero eigenvalues of $V[\xi_t|\mathcal{G}_{t-1}]$ equals the number of factors K, and the associated eigenvectors span the column space of matrix Γ_t. It allows to identify g_t from (31) up to a non-singular transformation matrix which is $\mathcal{G}_{t-1}$-measurable. In fact, the conditional factor space in model (25) is identifiable up to transformations $f_t \to c_{t-1} + A_{t-1}f_t$, where c_{t-1} and A_{t-1} are $\mathcal{G}_{t-1}$-measurable. Gagliardini and Ma (2019) show how to choose a convenient normalization of the factor space in order to get a closed form expression for g_t. Specifically, Gagliardini and Ma (2019) normalize the latent factors such that $E[f_t|\mathcal{G}_{t-1}] = 0$, and $\Gamma_t = J_t$, where J_t is the matrix whose columns are the K normalized eigenvectors of $V[\xi_t|\mathcal{G}_{t-1}]$ associated with the non-zero eigenvalues. Under this normalization, it follows $g_t = (J'_t\Omega_t J_t)^{-1}J'_t\Omega_t\xi_t$, where Ω_t is any $m \times m$ positive definite matrix measurable w.r.t. $\mathcal{G}_{t-1}$, and $f_t = g_t - E[g_t|\mathcal{G}_{t-1}]$.

In a setting with $n, T \to \infty$, Gagliardini and Ma (2019) define consistent estimators for the conditional factor space and for its dimension by replacing population (cross-sectional, or conditional) expectations with sample analogues. They get $\hat{g}_t = (\hat{J}'_t\Omega_t\hat{J}_t)^{-1}\hat{J}'_t\Omega_t\hat{\xi}_t$, and $\hat{f}_t = \hat{g}_t - \hat{E}[\hat{g}_t|\mathcal{G}_{t-1}]$, where $\hat{\xi}_t = \frac{1}{n}\sum_i Z_{i,t-1}R_{i,t}$. Here, $\hat{J}_t$ is the matrix of the standardized eigenvectors to the K largest eigenvalues of $\hat{V}[\hat{\xi}_t|\mathcal{G}_{t-1}]$, and $\hat{E}[\cdot|\mathcal{G}_{t-1}]$ and $\hat{V}[\cdot|\mathcal{G}_{t-1}]$ are nonparametric estimators for the conditional expectation given Z_{t-1}. Given the potentially large dimension of vector Z_t, Gagliardini and Ma (2019) consider estimators based on post-Lasso and Artificial Neural Networks (ANN). The estimator of the conditional factor space is in closed form up to the nonparametric regression given Z_{t-1}. The estimator of the number of factors is $\hat{K} = \text{mode}\left\{\hat{k}_t, t \geq 1\right\}$, and $\hat{k}_t = \underset{1 \leq k \leq k_{max}}{\arg\max} \dfrac{\mu_{k+1}(\hat{V}[\hat{\xi}_t|\mathcal{G}_{t-1}])}{\mu_k(\hat{V}[\hat{\xi}_t|\mathcal{G}_{t-1}])}$, where $\mu_k(\cdot)$ denotes the k largest eigenvalue of a symmetric matrix. Estimator $\hat{k}_t$ exploits the idea of the eigenvalue ratio test but in a different context than Ahn and Horenstein (2013), since $\hat{V}[\hat{\xi}_t|\mathcal{G}_{t-1}]$ is not a large dimensional sample variance-covariance matrix.

In the framework of Kelly et al. (2017, 2019), Equation (27) yields $\Gamma_t = Q_{Z,t-1}B$, where $Q_{Z,t-1} = \underset{n\to\infty}{\text{plim}} \dfrac{1}{n} \sum_i Z_{i,t-1} Z'_{i,t-1}$, which implies a constraint on the time variation of Γ_t. In Gu et al. (2019, forthcoming), we have $x_t := \underset{n\to\infty}{\text{plim}} \hat{x}_t = Q^{-1}_{Z,t-1}\xi_t = Q^{-1}_{Z,t-1}\Gamma_t g_t$. Hence, for large n, the autoencoder mapping for the latent factor essentially amounts to fixing a normalization of the latent factor such that some $k \times k$ block of $Q^{-1}_{Z,t-1}\Gamma_t$ is time-invariant, so that we can write g_t as a time-invariant function of x_t. This function is linear. The methodology of Gagliardini and Ma (2019) does not impose constraints on the dynamics of Γ_t and deploys the structural linear link between ξ_t and g_t conditional on $\mathcal{G}_{t-1}$.

Among the possible extensions of the model setting, we can further impose a group structure on the latent factor space in order to accommodate the presence of both common pervasive factors and group-specific pervasive factors. The former affect all series in the panel while the latter have an impact on subgroups of assets. The subgroups can correspond to e.g. economic sectors, asset classes, markets or countries. Andreou et al. (2019) develop inference procedures in a “large n, large T" setting for estimating the common and group-specific numbers of factors and the corresponding spanned factor spaces.

Finally, let us mention that there is also work on inference for large dimensional models with unobservable factors with high frequency data (Fan et al. (2016a), Ait-Sahalia and Xiu (2017), Pelger (2019a,b)), but extensions to the conditional case with instruments still need to be developed there. Fan and Kim (2018) discuss how to robustify such methods and Kim and Fan (2019) how to impose a dynamic parametric structure based on a factor GARCH-Itô process for prediction. Li et al. (2019) develop tests for deciding whether a large cross-section of asset prices obey an exact factor structure at the times of factor jumps with infill asymptotics. In the context of short time series panel, but when high-frequency data are not available, Zaffaroni (2019) provides inferential theory with unobservable factors for fixed T and large n in a conditional setting.

6 Empirical findings

In this section, we provide some empirical findings based on a large number of financial factor models. We provide contrast analysis based on monthly returns of individual stocks and standard sets of portfolios. The latter is the typical approach adopted in previous empirical applications of time-invariant factor models in finance. The empirical analysis presented below is made possible thanks to the unified econometric toolkit outlined in Section 4 for large dimensional conditional factor models with observable factors. Available literature studying large cross-sectional equity data sets has mainly relied on latent factor approaches as reviewed in Section 5.

6.1 Data description and factor models

Our dataset includes monthly excess returns of stocks data from CRSP database. We proxy the risk free rate with the monthly 30-day T-bill beginning-of-month yield. 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 $n = 10,827$ stocks, and covers the period from July 1963 to December 2017 with $T = 654$ months. Table 1 provides the distribution of asset returns of stocks w.r.t. T_i the number of observations available for each asset. About half of the stocks in the panel have more than 120 monthly return observations. We observe the complete time series of observations for only 2% of the stocks. Table 2 provides the distribution of stocks w.r.t. the classification of industry in Ferson and Harvey (1999). The two most frequent industry categories are Professional Services (2282) and Healthcare (1194), while the two less frequent ones are Aerospace (64) and Paper (129).

For comparison purposes with a standard methodology for small n, we consider i) the 25 Fama-French (FF) portfolios and ii) the 44 industry (Indu.) portfolios excluding four financial sectors (banking, insurance, real estate, and trading) as base assets.

We consider several linear factor models that involve financial variables (see GOS2 for models with macroeconomic variables). Table 3 lists the financial models, the factors, the number of parameters to estimate, and the trimmed cross-sectional dimensions n^{χ} considering time-invariant and time-varying specifications. The three factors of Fama and French (1993) are the monthly excess return $r_{m,t}$ on CRSP NYSE/AMEX/Nasdaq value-weighted market portfolio over the risk free rate, and the monthly returns on zero-investment factor-mimicking portfolios for size and book-to-market, denoted by $r_{smb,t}$ and $r_{hml,t}$. We denote the monthly returns on portfolio for momentum by $r_{mom,t}$ (Carhart (1997)). The two operative profitability factors of Fama and French (2015) are the difference between monthly returns on diversified portfolios with robust and weak profitability and investments, and with low and high investment stocks, denoted by $r_{rmw,t}$ and $r_{cma,t}$. We have downloaded the time series of these factors from the website of Kenneth French. We also consider a model with long-only factors, that should be more immune to market imperfections (e.g., transaction costs). We build the long-only factors from the six FF research portfolios available on the website of Ken French. The excess return of the "Small" factor (denoted by $r_{s,t}$) is the average excess return of the three small portfolios, and the excess return of the "Value" factor (denoted by $r_{h,t}$) is the average excess return of the two value portfolios. Furthermore, we include quality minus junk (qmj_t) and bet against beta (bab_t) factors as described in Asness et al. (2019) and Frazzini and Pedersen (2014). The factor return qmj_t is the average return on the two high quality portfolios minus the average return on the two low quality (junk) portfolios. The bet against beta factor is a portfolio that is long low-beta securities and short high-beta securities. We have downloaded these data from the website of AQR. As additional specifications, from the website of Kenneth French, we consider the two reversal factors which are monthly returns on portfolios for short-term and long-term reversals, denoted by $r_{strev,t}$ and $r_{ltrev,t}$.

To account for time-varying alphas, betas and risk premia, we use a conditional specification based on one common variable and a firm-level variable. We take the instruments $Z_{t-1} = (1, divY_{t-1})'$, where $divY_{t-1}$ is the lagged dividend yield and the asset specific instrument $bm_{i,t-1}$ corresponds to the lagged book-to-market equity of firm i. We compute the book-to-market equity of firm i as defined in logarithmic terms by Fama and French (2008). We compute the firm characteristics from Compustat as in the appendix of Fama and French (2008). We consider all the assets for which the book-to-market equity is always positive over the sample period, as in Fama and French (2008). The number of assets reduces to $n = 8,570$ for the estimation of the time-varying specifications. We refer to Avramov and Chordia (2006) for convincing theoretical and empirical arguments in favor of the chosen conditional specification. In Table 3, the vector $x_{i,t}$ has maximum dimension $d = 23$ (CAR and REV model), and parsimony explains why we have not included e.g. the size of firm i as an additional stock specific instrument. We have downloaded time series of portfolio characteristics from the website of Kenneth French.

6.2 Time-invariant specifications

Let us first focus on the time-invariant specifications (i.e. $Z_t = 1$ and $Z_{i,t} = 0$) in order to benchmark the results of the next section for the time-varying specifications. We use $\chi_{1,T} = 15$ as advocated by Greene (2008), together with $\chi_{2,T} = 546/60$. The number of assets whose condition number is below 15 is $7,754$ for each model specification.

First, we compute the diagnostic criterion and the number k of omitted factors. Table 4 reports the contribution in percentage of the first eigenvalue μ_1 with respect to the variance of normalized residuals $\dfrac{1}{n^{\chi}T}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'$, that is equal to one by construction under our variance scaling for each time series of residuals. We also report the selected number of omitted factors k, the contribution of the first k eigenvalues, i.e., $\sum_{j=1}^{k}\mu_j$, and the incremental contribution of the $(k+1)$-th eigenvalue μ_{k+1}. For each model, we have to specify the numerical value of the penalisation function $g\left(n^{\chi}, T\right)$. We use the penalisation

$$g(n,T) = c\frac{\left(\sqrt{n}+\sqrt{T}\right)^2}{nT}\ln\left(\frac{nT}{\left(\sqrt{n}+\sqrt{T}\right)^2}\right), \tag{32}$$

with a data-driven constant c based on the proposal of Alessi et al. (2010); see also Hallin and Liška (2007) in the general dynamic factor model. The multiplicative factor $\frac{\left(\sqrt{n}+\sqrt{T}\right)^2}{nT}$ in (32) is the order of the largest eigenvalue predicted by random matrix theory if the residuals were independent standard Gaussian variates (Geman (1980), Johnstone (2001)). The multiplicative factor $\ln\left(\frac{nT}{\left(\sqrt{n}+\sqrt{T}\right)^2}\right)$ ensures that $\min\{n,T\}g(n,T) \to \infty$, so that we get a valid selection procedure.

The number k of omitted factors is larger than one for the most popular financial models, e.g., the CAPM (Sharpe (1964)) and the three-factor Fama-French model (FF). On the contrary, for the the four-factor Carhart (1997) model (CAR), the five-factor Fama-French model (5FF), quality minus junk (QMJ), and models involving the reversal factors, we find no omitted latent factor. We observe that adding observable factors helps to reduce the contribution of the first eigenvalue μ_1 to the variance of residuals. However, when we face latent factors, the omitted systematic contribution $\sum_{j=1}^{k}\mu_j$ only accounts for a small proportion of the residual variance. For instance, we find $k = 1$ omitted factors in the CAPM. That latent factor only contributes to $\mu_1 = 2.39\%$ of the residual variance. Figures 1, 3, 5 summarize this information graphically by displaying the penalized scree plots and the plots of cumulated eigenvalues for the CAPM, the three

Fama-French factors model and the four-factor CAR model. For instance, $\mu_2 = 1.54\%$ lies below the horizontal line $g(n^{\chi}, T) = 1.55\%$ in Panel A for the time-invariant CAPM, so that $k = 1$. In Panel B for the time-invariant CAPM, the vertical bar $\mu_1 + \mu_2 = 3.93\%$ is divided into the contribution of $\mu_1 = 2.39\%$ (light grey area) and that of $\mu_2 = 1.54\%$ (dark grey area). Figure 2 Panel A displays the scree plots of squared eigenvalues for the CAPM and the square $g^2(n^{\chi}, T)$ of the penalisation function relative to the squared Frobenius norm $\sum_{l=1}^{T} \mu_l^2 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i' \right)$. By construction, the conclusion of the number of omitted factor is the same as for the scree plot shown in Figure 1. For example, we get that the sum of the square of the two first eigenvalues accounts for 21.45% of the square of the Frobenius norm for the time-invariant CAPM. Thus, the two latent factors are much more representative of the off-diagonal components. We conclude similarly for the time-invariant FF model (see Figure 4), even if the correlation explanation provided by the single omitted factor is lower.

Tables 5-8 gather the estimated annual risk premia and the estimates of the components of ν, with the corresponding confidence intervals at 95% level, for the ten time-invariant models listed in Table 3. For individual stocks, we use bias-corrected estimates for λ and ν. In order to build the confidence intervals, we use the HAC estimators $\hat{\Sigma}_f$ defined as in Newey and West (1994) and $\tilde{\Sigma}_\nu$ defined in GOS. When we consider the 25 FF and 44 Indu. portfolios as base assets, we use asymptotics for fixed n and $T \to \infty$. In particular, we compute the estimates of the variance-covariance matrices $\Sigma_{\lambda,n}$ and $\Sigma_{\nu,n}$ defined in GOS. The estimated risk premia for the market factor are of the same magnitude and all positive across the three universes of assets and all financial models. In Table 7, for the four-factor CAR model and the individual stocks, the size factor is positively remunerated (3.5430%) and it is significantly different from zero. The value factor commands a significant negative reward (-4.9265%). The momentum factor is largely remunerated (8.0947%) and significantly different from zero. For the 25 FF portfolios, we observe that the size factor is not significantly positively remunerated while the value factor is significantly positively remunerated (2.5028% and 4.1996%). The momentum factor bears a significant positive reward (34.6689%). For λ_m, λ_{smb}, λ_{hml}, we obtain similar inferential results when we consider the Fama-French model in Table 8. Our point estimates of λ_m, λ_{smb}, and λ_{hml} for large n agree with Ang et al. (2020). Our point estimates and confidence intervals for λ_m, λ_{smb}, and λ_{hml} agree with the results reported by Shanken and Zhou (2007) for the 25 FF portfolios. The large, but imprecise, estimate for the momentum premium when $n = 25$

comes from the estimate for ν_{mom} (26.7559%) that is much larger and less accurate than the estimates for ν_m, ν_{smb}, and ν_{hml} (0.9447%, -0.0225%, -0.3662%). Moreover, while the estimates of ν_m, ν_{smb}, and ν_{hml} are statistically not significant for the 25 FF portfolios, the estimates of ν_m, ν_{smb}, and ν_{hml} are statistically different from zero for individual stocks. In particular, the estimate of ν_{hml} is large and negative. It explains the negative estimate on the value premium for individual stocks displayed in Table 7, despite the positive time average of the value factor. Phalippou (2007) obtained a similar growth premium for portfolios built on stocks with a high institutional ownership. The results with the 44 Indu. portfolios sharply differ from those with the 25 FF portfolios. The former are more like the results for individual stocks; in particular, they yield negative estimates of coefficient ν_{hml} and value premium λ_{hml} (albeit the latter not statistically significant). In Table 6, the 5FF model also exhibits large differences between estimated risk premia on individual stocks, FF and Indu. portfolios. For example, we get a significant $\lambda_{rmw} = 5.3198\%$ for the FF portfolios and an insignificant $\lambda_{rmw} = 0.8911\%$ for individual stocks. On the contrary, we get an insignificant $\lambda_{cma} = 0.8787\%$ for the FF portfolios (with a large confidence interval) and a significant $\lambda_{cma} = -3.2867\%$ for individual stocks. The estimated risk premia on the Indu. portfolios exhibit large confidence intervals. For example, we get insignificant $\lambda_{rmw} = 2.3817\%$ and $\lambda_{cma} = -0.3614\%$.

The size, value, and momentum factors are tradable in theory. In practice, their implementation faces transaction costs due to rebalancing and short selling. A nonzero ν might capture these market imperfections (Cremers et al. (2012)). In Table 8, we also get zero estimates with the FF portfolios except for value, and nonzero estimates with the Indu. portfolios and the individual stocks for market and value, when we use a time-invariant model with long-only factors derived from the FF methodology. Market imperfections are probably not the key drivers here (see Frazzini et al. (2012)) for empirical support based on live trading data from a large institutional money manager).

A potential explanation of the discrepancies revealed in Tables 5-8 between individual stocks and the FF portfolios is the much larger heterogeneity of the factor loadings for the former. As already discussed in Lewellen et al. (2010), the FF portfolio betas are all concentrated in the middle of the cross-sectional distribution obtained from the individual stocks. Creating portfolios with an ad hoc methodology distorts information by shrinking the dispersion of betas. The estimation results for the momentum factor on the FF portfolios exemplify the problems related to a small number of portfolios exhibiting a tight factor structure.

Another potential explanation of the discrepancy revealed in Tables 5-8 is the effect of model misspecification on the risk premia because of omitted factors as observed in Table 4 for the three-factor FF model.

6.3 Time-varying specifications

We use $\chi_{1,T} = 15$ and $\chi_{2,T} = 546/60$. The number of assets whose condition number is below 15 is often between 2,000 and 3,000, for instance 2,578 for the four-factor CAR model.

For the time-varying specifications of Table 3, we still find one omitted factor for the CAPM and the 4-factor MOM and REV model in Table 4. The other time-varying models pass the diagnostic criterion. As already discussed in the Introduction, this diagnostic step is crucial to decide whether we can feel comfortable with the chosen set of observable factors before proceeding further in an empirical analysis of a large cross sectional equity data set under the APT setting. The time-varying specification is more parsimonious for the factor space in the conditional sense, but less parsimonious for the parameter space. From an econometric point of view, it is not clear which parsimony we should favor to decide between the time-invariant specification (more factors, fewer parameters) and the time-varying specification (fewer factors, more parameters). For investment purposes, the first one is better suited for static (unconditional) decisions while the second one is better suited for dynamic (conditional) decisions. The choice between the two models should meet the investor needs or answer the empirical research question at hand.

Figure 7 plots the estimated time-varying paths of the four risk premia estimated assuming the four-factor CAR model and using the individual stocks (see GOS for a formal test of time variation based on the estimated coefficients $\hat{F}$ and $\hat{\nu}$). For comparison purpose, we also plot the time-invariant estimates and the average lambdas over time. A well-known bias coming from market-timing and volatility-timing (Jagannathan and Wang (1996), Lewellen and Nagel (2006), Boguth et al. (2011)) explains the discrepancy between the time-invariant estimate and the average over time. After trimming, we compute the risk premia on $n^{\chi} = 2,549$ individual assets in the four-factor CAR model. The observed discrepancy w.r.t. the average over time is only marginally explained by the larger size of the stock universe used for the time-invariant estimates. The risk premia for the factors feature a counter-cyclical pattern most of the time. Indeed, these risk premia increase during economic contractions and decrease during economic booms. Gomes et al. (2003) and Zhang (2005) constructed equilibrium models exhibiting a counter-cyclical behavior in size and book-to-market effects. Furthermore, time-varying estimates of the value premium are negative and might take positive values because of the large confidence intervals around recessions. Growth firms are riskier in boom times because of their in-the-money growth options; value firms are riskier in recession times because of default risk. However, empirical evidence for such an interpretation is mixed. Some papers find that distress is related to size and book-to-market effects (Griffin and Lemmon (2002), Vassalou and Xing (2004)) while other papers find the opposite (Dichev (1998), Campbell et al. (2008)). Chava and Purnanandam (2010) find support for a positive relation and argued that conclusions regarding the risk return trade-off can change significantly depending on how the expected return is measured. Gomes and Schmid (2010) and Garlappi and Yan (2011) argue that financial leverage provides a rationale for a positive relation. The time-varying estimates of the size premium are most of the time slightly positive.

Figure 8 plots the estimated time-varying path of the four risk premia from the 25 FF portfolios. We also plot the time-invariant estimates and the average lambdas over time. The discrepancy between the time-invariant estimates and the averages over time is also observed for $n = 25$. The time-varying point estimates for $\lambda_{mom,t}$ are typically smaller than the time-invariant estimate in Table 7, but both estimates are rather inaccurate. Finally, by comparing Figures 7 and 8, we observe that the patterns of risk premia look similar in terms of cyclicality except for the book-to-market factor. Indeed, the risk premium for the value effect estimated from the 25 portfolios is pro-cyclical, contradicting the counter-cyclical behavior predicted by finance theory. The paths of risk premia in the Fama-French model estimated from the 25 FF portfolios look similar to the corresponding estimates for the four-factor CAR model in Figure 8. The time-varying paths of risk premia for the 44 Indu. portfolios look similar to the corresponding estimates on individual stocks. This similarity, also observed in Section 6.2 with time-invariant models, is likely linked with the relative stability of the time-varying portfolio weights for the 44 Indu. portfolios compared to the weights of the 25 FF portfolios.

6.4 Asset pricing restriction tests

As already discussed in Lewellen et al. (2010), the 25 FF portfolios have four-factor CAR market and momentum betas close to 1 and zero, respectively. As depicted in Figure 1 by Lewellen, Nagel, and Shanken (2010), this empirical concentration implies that it is easy to get artificially large estimates $\hat{\rho}^2$ of the cross-sectional R^2 for three-factor FF and four-factor CAR models. On the contrary, the observed heterogeneity in the betas coming from the individual stocks impedes this. It suggests that it is much less easy to find factors that explain the cross-sectional variation of expected excess returns on individual stocks than on portfolios. Reporting large $\hat{\rho}^2$, or small SSR $\hat{Q}_e$, when n is large, is much more impressive than when n is small.

Tables 9 and 10 gather the results for the tests of the asset pricing restrictions in factor models with time-invariant coefficients. When n is large, we prefer working with test statistics based on the SSR $\hat{Q}_e$ instead of $\hat{\rho}^2$ since the population R^2 is not well-defined with tradable factors under the null hypothesis (its denominator is zero). For the individual stocks, we compute the feasible test statistics based on $\hat{Q}_e$ and $\hat{Q}_a$ and hard thresholding to get consistent estimates $\tilde{\Sigma}_\xi$ of covariance matrices, as well as their associated one-sided p-value. Our Monte Carlo simulations show that we need to set a stronger trimming level $\chi_{2,T}$ to compute the test statistic than to estimate the risk premium. We use $\chi_{2,T} = 546/240$. For the 25 and 44 Indu. portfolios, we compute weighted test statistics (Gibbons et al. (1989)) as well as their associated p-values. For individual stocks, the test statistics reject both null hypotheses $\mathcal{H}_0 : a(\gamma) = b(\gamma)'\nu$ and $\mathcal{H}_0 : a(\gamma) = 0$ for all specifications at 1% level. Similar conclusions are obtained when using the 25 FF portfolios as base assets. For the 44 Indu. portfolios, we do not reject the null hypothesis $\mathcal{H}_0 : a(\gamma) = b(\gamma)'\nu$, but we reject $\mathcal{H}_0 : a(\gamma) = 0$.

Tables 11 and 12 gather the results for tests of the asset pricing restrictions in time-varying specifications. We do not report results for the FF long-only model since multicollinearity problems prevent us to estimate and test that model. Contrary to the time-invariant case, we do not report the values of the weighted test statistics (Gibbons et al. (1989)) computed for portfolios because of the numerical instability in the inversion of the covariance matrix. Instead, we report the values of the test statistics $T\hat{Q}_e$ and $T\hat{Q}_a$. For individual stocks, the test statistics reject both null hypotheses $\mathcal{H}_0 : a(\gamma) = b(\gamma)'\nu$ and $\mathcal{H}_0 : a(\gamma) = 0$ for all specifications at 1% level.

In addition, we compare the cross-sectional distributions of $\hat{\beta}_{1,i}'\hat{\beta}_{1,i}$, the idiosyncratic risk (square root of residual variance), and the estimated time-series coefficient of determination $\hat{\rho}_i^2$ (ratio of explained variance and total variance) for the time-varying specifications assuming the four-factor CAR model for the excess returns. We can view those estimates as measures of limits-to-arbitrage and missing factor impact (Pontiff (2006), Lam and Wei (2011), Ang et al. (2009)). For each asset (either stock, or portfolio)

i, we compute four measures: (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 $SysRisk_i = \sqrt{\frac{ESS_i}{T_i}}$.

Figures 13 and 14 compare the cross-sectional distributions of the four measures (i)-(iv) computed on the time-invariant and time-varying four-factor CAR models using the individual stocks, 25 FF and 44 Indu. portfolios as base assets. The boxplots provided by the statistical software do not take into account the presence of estimation noise, i.e, of the EIV issue coming from using estimates instead of true quantities. Barras et al. (2019) explain how to correct for the EIV bias and to modify standard deviations and confidence intervals of estimates of p.d.f., c.d.f., quantiles, and moments computed from estimated quantities such as estimated regression coefficients in a "large n, large T" setting. For comparison purposes, the cross-sectional distributions for individual stocks in both figures refer to the $n^{\chi} = 2,549$ stocks used in the estimation of the time-varying specification after trimming. The time-series (adjusted) $\hat{\rho}_i^2$ of the 25 FF portfolios are all larger than 0.80. The estimates $\hat{\rho}_i^2$ of the individual stocks are typically much smaller, with a median below 0.30. As expected, the excess returns of individual stocks also have larger idiosyncratic volatilities. The time-series adjusted $\hat{\rho}_i^2$ of individual stocks tend to be a bit larger in the time-varying model than in the time-invariant one, as a result of the explanatory power that we gain by allowing for beta dynamics. Figures 13 and 14 show that the use of the FF portfolios also shrinks the dispersion of $\hat{\rho}_i^2$, $IdiVol_i$, and $SysRisk_i$, by a large amount. The distributions for the individual stocks and the 44 Indu. portfolios are comparable and share a wide support. Figure 15 plots the cross-sectional distributions of $\hat{\beta}_{1,i}'\hat{\beta}_{1,i}$ for the three universes of assets. We observe a huge heterogeneity in $\hat{\beta}_{1,i}'\hat{\beta}_{1,i}$ for the individual stocks in Figure 15, similar to the one observed on $IdiVol_i$ in Figure 14. We may face the presence of limits-to-arbitrage and missing factors in that case. On the contrary, the estimates $\hat{\beta}_{1,i}'\hat{\beta}_{1,i}$ are concentrated close to zero for the 25 FF and 44 Indu. portfolios. The 25 FF portfolios exhibit small $\hat{\beta}_{1,i}'\hat{\beta}_{1,i}$, small idiosyncratic risks, and large estimates $\hat{\rho}_i$ compared to individual stocks as expected from the previous empirical results. Unreported preliminary results based on linear quantile regressions reveal that stocks with small size tend to yield large $\hat{\beta}_{1,i}'\hat{\beta}_{1,i}$, large idiosyncratic risks, and small estimates $\hat{\rho}_i$. We also find that firms with short observation periods tend to be associated with large values of both idiosyncratic and systematic risks (with a larger proportion of systematic risk to total risk), as well as small market capitalization.

6.5 Time-varying cost of equity

We can use the results in Section 4.2 for estimation and inference on the cost of equity in conditional factor models (see Fama and French (1997) for a fixed n approach). We can estimate the time-varying cost of equity $CE_{i,t} = r_{f,t} + b_{i,t}'\lambda_t$ of firm i with $\widehat{CE}_{i,t} = r_{f,t} + \hat{b}_{i,t}'\hat{\lambda}_t$, where $r_{f,t}$ is the risk-free rate. We have

$$\begin{aligned} \sqrt{T}\left(\widehat{CE}_{i,t} - CE_{i,t}\right) &= \psi_{i,t}' E_2' \sqrt{T}\left(\hat{\beta}_i - \beta_i\right) \\ &\quad + \left(Z_{t-1}' \otimes b_{i,t}'\right) W_{p,K} \sqrt{T} vec\left[\hat{\Lambda}' - \Lambda'\right] + o_p(1), \end{aligned} \tag{33}$$

where $\psi_{i,t} = \left(\lambda_t' \otimes Z_{t-1}', \lambda_t' \otimes Z_{i,t-1}'\right)'$. Standard results on OLS imply that estimator $\hat{\beta}_i$ is asymptotically normal, $\sqrt{T}\left(\hat{\beta}_i - \beta_i\right) \Rightarrow N\left(0, \tau_i Q_{x,i}^{-1} S_{ii} Q_{x,i}^{-1}\right)$, and independent of estimator $\hat{\Lambda}$. Then, from the asymptotic normality results for the estimator $\hat{\Lambda}$, we deduce that $\sqrt{T}\left(\widehat{CE}_{i,t} - CE_{i,t}\right) \Rightarrow N\left(0, \Sigma_{CE_{i,t}}\right)$, conditionally on Z_{t-1}, where

$$\Sigma_{CE_{i,t}} = \tau_i \psi_{i,t}' E_2' Q_{x,i}^{-1} S_{ii} Q_{x,i}^{-1} E_2 \psi_{i,t} + \left(Z_{t-1}' \otimes b_{i,t}'\right) W_{p,K} \Sigma_\Lambda W_{K,p} \left(Z_{t-1} \otimes b_{i,t}\right).$$

Figure 16 plots the path of the estimated annualized costs of equity for Microsoft Corp, Apple, Disney Walt, and Sony. We use the time-varying four-factor CAR model estimated on individual stocks. For the last twenty years, the cost of equity rose substantially during the subprime crisis, but came back to much lower levels in the recent years. Again, using a time-invariant specification would obliterate those dynamic features, and could mislead investors in computing the return that a firm should pay to them to compensate for the risk they undertake by investing their capital.

6.6 International equity data sets

All of the empirical findings on factor structure, asset pricing restrictions tests, risk premia estimation and cost of equity discussed so far are based on evidence from a large cross-sectional equity data set for the U.S. market. It is interesting to examine these important issues in a global context. Market integration and currency risk are two main factors that distinguish international financing and investment decisions from domestic ones.

A recent paper by Chaieb, Langlois and Scaillet (2020, CLS hereafter) extends the GOS methodology to make it applicable to a very large panel of international equity returns. The sample includes more than 64,000 stocks from 23 Developed Markets (DMs) and 24 Emerging Markets (EMs). It is the first time in the literature that a large international database is analysed at the individual stock level, and time-varying risk premia are infered from. GOS methodology is particularly suitable to model USD-denominated international equity returns as it can handle the correlation implications of denominating all returns in a common currency. CLS explicitly consider the impact of currency conversion on correlations across stocks since they do not impose a priori an exact factor structure. To handle the large number of parameters needed to capture time-varying factor exposures and risk premia in multifactor models for an international setting, CLS extend the GOS methodology by automatically selecting the most important instruments for each stock while ensuring consistency with no-arbitrage conditions. This extension allows dimension reduction and renders the estimation approach applicable in an international setting through more parsimonious first-pass regressions. Using this framework, CLS document several new empirical results based on individual international stocks.

First, global or regional factor models fail to fully capture the factor structure in many DMs and EMs. It holds for models with only a market factor or models augmented with non-market factors such as size, value, momentum, profitability, and investment. The country-excess market factor - defined as the spread between the country market and the world market - is crucial to capture the factor structure in individual stock returns for both DMs and EMs. Bekaert et al. (2009), Fama and French (2012), and Fama and French (2017) find that regional factors perform better than global factors on portfolios. The results with individual stocks show that regional factors are not sufficient to fully capture the factor structure, a necessary step prior to estimating risk premia. CLS obtain a factor model specification close to a block diagonal structure for the error covariance matrix with blocks corresponding to countries. Such a sparse matrix is compatible with the notion of weak cross-sectional dependence. Therefore, the leading Fama and French (2012, 2017) four-factor and five-factor international models and Hou et al. (2015) q-factor model (when applied using global factors) should allow factor exposures to vary over time, specifically with stock characteristics, and should add a country excess market factor to capture the factor structure of international individual equity returns for both DMs and EMs.

Second, although the country-excess market factor is crucial to capture the factor structure in individual stock returns for both DMs and EMs, it carries a small risk premium for DMs but a large one in EMs.

Third, since asset pricing restrictions are often rejected, it is important to account for model misspecification. Notwithstanding a few differences across regions, CLS show that pricing errors from leading factor models constitute a large part of expected returns.

7 Concluding remarks

After an historical perspective on conditional factor models with a small number of assets, this chapter has reviewed recent advances in econometrics for conditional factor models estimated on data sets with "large n, large T" in finance. The tools studied above are simple to implement and often similar to the ones used in a "small n, large T" setting. The asymptotic treatment however differs substantially. The empirical results on individual stocks also differ substantially from an analysis relying on standard sets of portfolios, and show the importance of allowing for time variation in the loadings and risk premia. The empirics reveal the relevance of including characteristics in conditional modeling. We believe that extracting information directly from disaggregated data in finance will become increasing popular in the upcoming years as it is already the case in other fields, e.g., in labor econometrics. The current big data trend favours the development of new econometric tools, the collection of data sets at the individual level, and the improvement of computation/storage powers.

Table 1: **Distribution of individual stocks w.r.t.** T_i

T_i	Frequency
$1 \leq 12$	18
$13 \leq 24$	607
$25 \leq 60$	2514
$61 \leq 120$	2444
$121 \leq 240$	2861
$241 \leq 360$	1351
$361 \leq 480$	557
$481 \leq 600$	286
$601 \leq 654$	189

We report the frequency counts of the individual stocks w.r.t. their buckets of sample size T_i.

Table 2: **Distribution of individual stocks w.r.t. industry**

Industry	Frequency
Aerospace	64
Transportation	152
Building Materials	147
Chemicals/Plastics	276
Construction	177
Entertainment	375
Food/Beverage	314
Healthcare	1194
Industrial Machinery	322
Metals	198
Mining	364
Motor Vehicles	155
Paper	129
Petroleum	869
Printing/Publishing	194
Professional Services	2282
Retailing	577
Semiconductors	861
Telecommunications	901
Textiles/Apparel	200
Utilities	437
Wholesaling	639

We report the frequency counts of the individual stocks w.r.t. their industry category.

Table 3: Financial linear factor models.

Model	Observable Factors	time-invariant		time-varying	
		K	n^{χ}	d	n^{χ}
CAPM	$r_{m,t}$	1	7,754	8	3,215
FF	$r_{m,t}, r_{smb,t}, r_{hml,t}$	3	7,754	14	2,877
FF long-only	$r_{m,t}, r_{s,t}, r_{h,t}$	3	7,754	14	825
CAR	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}$	4	7,754	17	2,549
FF and QMJ	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{qmj,t}$	4	7,754	17	2,251
FF and BAB	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{bab,t}$	4	7,754	17	2,471
MOM and REV	$r_{m,t}, r_{mom,t}, r_{strev,t}, r_{ltrev,t}$	4	7,754	17	2,569
5FF	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{rmw,t}, r_{cma,t}$	5	7,754	20	1,928
FF and REV	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{strev,t}, r_{ltrev,t}$	5	7,754	20	2,460
CAR and REV	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}, r_{strev,t}, r_{ltrev,t}$	6	7,754	23	2,019

For each financial model, we report the list of observable factors and the trimmed cross-sectional dimension n^{χ} for estimation from monthly data. We use $\chi_{1,T} = 15$ and $\chi_{2,T} = 546/60$. For the time-invariant specifications, we report the number of observable factors K. The number of parameters to estimate is $d = K + 1$. For the time-varying specifications, we give the dimension d of vector $x_{i,t}$ using $Z_{t-1} = (1, divY_{t-1})'$ and $Z_{i,t-1} = bm_{i,t-1}$.

Table 4: Number of omitted factors.

Model	μ_1	k	$\sum_{j=1}^{k} \mu_j$	μ_{k+1}	Penalty	μ_1	k	$\sum_{j=1}^{k} \mu_j$	μ_{k+1}	Penalty
	Panel A: time-invariant models					Panel B: time-varying models				
CAPM	2.39	1	2.39	1.54	1.55	2.17	1	2.17	1.65	1.66
FF	1.55	1	1.55	1.18	1.19	1.46	0	0	1.46	1.60
FF long-only	1.57	1	1.57	1.22	1.25	1.48	0	0	1.48	1.60
CAR	1.28	0	0.00	1.28	1.52	1.27	0	0	1.27	1.60
FF and QMJ	1.50	0	0.00	1.50	1.52	1.43	0	0	1.43	1.60
FF and BAB	1.54	1	1.54	1.13	1.16	1.45	0	0	1.45	1.60
MOM and REV	2.22	1	2.22	1.28	1.31	2.03	1	2.03	1.33	1.35
5FF	1.47	0	0.00	1.47	1.52	1.39	0	0	1.39	1.60
FF and REV	1.42	0	0.00	1.42	1.52	1.37	0	0	1.37	1.60
CAR and REV	1.24	0	0.00	1.24	1.52	1.24	0	0	1.24	1.60

The table shows the contribution of the first eigenvalue μ_1 to the variance of normalised residuals, the number of omitted factors k, the contributions of the first k, and of the $(k+1)$-th eigenvalues, and the penalty term. Panel A and B report the results for time-invariant and time-varying specifications, respectively.

Table 5: Estimated annualized risk premia and ν for the time-invariant specifications (CAR and REV, FF and REV)

	Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)		Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)
				CAR and REV			
λ_m	8.3628	7.4505	8.0807	ν_m	1.9910	1.0787	1.7089
	(4.0831, 12.6424)	(3.0727, 11.8282)	(3.6164, 12.5153)		(1.5808, 2.4012)	(0.1572, 2.0003)	(0.5467, 2.8711)
λ_{smb}	3.7984	2.4349	-2.6986	ν_{smb}	1.2731	-0.0904	-5.2239
	(0.9007, 6.6961)	(-0.5710, 5.4409)	(-6.8331, 1.4360)		(0.6300, 1.9161)	(-0.8899, 0.7092)	(-8.1731, -2.2747)
λ_{hml}	-5.4866	3.9988	-1.9805	ν_{hml}	-10.0524	-0.5671	-6.5464
	(-8.4465, -2.5266)	(0.6230, 7.3746)	(-7.0536, 3.0925)		(-10.9627, -9.1422)	(-2.1903, 1.0562)	(-10.6665, -2.4263)
λ_{mom}	6.8604	37.4039	3.5191	ν_{mom}	-1.0527	29.4909	-4.3939
	(2.9055, 10.8152)	(16.9232, 57.8846)	(-8.0371, 15.0753)		(-2.8090, 0.7037)	(9.3956, 49.5861)	(-15.2524, 6.4645)
λ_{strev}	0.7385	12.2081	8.8738	ν_{strev}	-5.0588	6.4108	3.0766
	(-1.9118, 3.3888)	(-5.6558, 30.0719)	(-2.9362, 20.6839)		(-6.2475, -3.8700)	(-11.2554, 24.0770)	(-8.4322, 14.5854)
λ_{ltrev}	-3.8142	4.4558	-4.3534	ν_{ltrev}	-6.8584	1.4115	-7.3976
	(-6.4351, -1.1933)	(-8.2076, 17.1191)	(-13.5046, 4.7979)		(-7.5727, -6.1442)	(-10.9776, 13.8007)	(-16.1655, 1.3703)
				FF and REV			
λ_m	8.0307	6.8128	7.9634	ν_m	1.6589	0.441	1.5917
	(3.7510, 12.3103)	(2.5093, 11.1162)	(3.5576, 12.3693)		(1.2624, 2.0554)	(-0.0108, 0.8929)	(0.5449, 2.6385)
λ_{smb}	3.5417	2.187	-3.2215	ν_{smb}	1.0164	-0.3384	-5.7468
	(0.6440, 6.4393)	(-0.7619, 5.1358)	(-7.2361, 0.7931)		(0.3973, 1.6355)	(-0.8855, 0.2088)	(-8.5254, -2.9682)
λ_{hml}	-5.7603	4.3557	-2.7668	ν_{hml}	-10.3261	-0.2102	-7.3326
	(-8.7202, -2.8003)	(1.0337, 7.6776)	(-7.11503, 1.6168)		(-11.1277, -9.5245)	(-1.7183, 1.2979)	(-10.5650, -4.0993)
λ_{strev}	-0.3091	-2.3783	4.5768	ν_{strev}	-6.1063	-8.1756	-1.2204
	(-2.9594, 2.3412)	(-13.6653, 8.9086)	(-6.2186, 15.3723)		(-7.2106, -5.0021)	(-19.1469, 2.7958)	(-11.6855, 9.2447)
λ_{ltrev}	-4.7160	7.3225	-7.9227	ν_{ltrev}	-7.7603	4.2783	-10.9669
	(-7.3369, -2.0952)	(-0.4908, 15.1358)	(-15.9998, 0.1544)		(-8.4400, -7.0805)	(-3.0823, 11.6389)	(-18.6070, -3.3269)

The table contains the estimates and the corresponding confidence intervals of the annualized risk premia and of the components of vector ν for the time-invariant specifications CAR and REV, and FF and REV estimated using the three different sets of base assets.

Table 6: Estimated annualized risk premia and ν for the time-invariant specifications (5FF, MOM and REV)

	Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)		Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)
				5FF			
λ_m	7.2385	5.7861	7.2065	ν_m	0.8668	-0.5856	0.8648
	(2.9589, 11.5181)	(1.4686, 10.1037)	(2.8474, 11.5657)		(0.5105, 1.2230)	(-1.1565, -0.0147)	(0.0060, 1.6636)
λ_{smb}	3.9635	3.8616	-1.5358	ν_{smb}	0.9712	0.8693	-4.5281
	(1.0341, 6.8928)	(0.8784, 6.8448)	(-5.5843, 2.5126)		(0.4585, 1.4838)	(0.3051, 1.4335)	(-7.3225, -1.7338)
λ_{hml}	-6.4111	3.6934	-3.6409	ν_{hml}	-10.5696	-0.4652	-7.7994
	(-9.4377, -3.3845)	(0.6118, 6.7749)	(-8.1114, 0.8296)		(-11.1010, -10.0382)	(-1.0446, 0.1143)	(-11.0896, -4.5093)
λ_{rmw}	0.8911	5.3198	2.3817	ν_{rmw}	-2.1874	2.2413	-0.6969
	(-1.3642, 3.1464)	(1.0680, 9.5716)	(-1.6346, 6.3980)		(-2.7289, -1.6459)	(-1.3631, 5.8457)	(-4.0201, 2.6264)
λ_{cma}	-3.2867	0.8787	-0.3614	ν_{cma}	-6.7305	-2.5652	-3.8063
	(-5.3748, -1.1986)	(-3.3306, 5.0880)	(-4.8824, 4.1596)		(-7.2786, -6.1825)	(-6.2200, 1.0897)	(-7.8152, 0.2047)
				MOM and REV			
λ_m	9.3559	6.8911	7.7548	ν_m	2.9842	0.5194	1.383
	(5.0763, 13.6356)	(2.5690, 11.2132)	(3.3423, 12.1672)		(2.3647, 3.6036)	(-0.0846, 1.1234)	(0.3088, 2.4572)
λ_{mom}	9.0771	-4.245	-2.7548	ν_{mom}	1.1641	-12.158	-10.6678
	(5.1223, 13.0320)	(-24.1192, 15.6292)	(-13.0607, 7.5512)		(-0.7642, 3.0924)	(-31.6348, 7.3187)	(-20.1847, -1.1509)
λ_{strev}	0.4665	-6.657	2.7492	ν_{strev}	-5.3307	-12.4542	-3.048
	(-2.1838, 3.1168)	(-21.8234, 8.5094)	(-8.5381, 14.0366)		(-6.6386, -4.0228)	(-27.3873, 2.4788)	(-14.0198, 7.9238)
λ_{ltrev}	-2.1908	7.558	-8.3057	ν_{ltrev}	-5.2351	4.5138	-11.3499
	(-4.8117, 0.4300)	(2.5750, 12.5411)	(-16.4515, -0.1599)		(-6.4058, -4.0644)	(0.2757, 8.7520)	(-19.0625, -3.6372)

The table contains the estimates and the corresponding confidence intervals of the annualized risk premia and of the components of vector ν for the time-invariant specifications 5FF, and MOM and REV estimated using the three different sets of base assets.

Table 7: Estimated annualized risk premia and ν for the time-invariant specifications (FF and BAB, FF and QMJ, CAR)

	Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)		Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)
FF and BAB							
λ_m	8.2095	6.4955	7.7839	ν_m	1.8377	0.1238	1.4121
	(3.9298, 12.4891)	(2.1693, 10.8218)	(3.4167, 12.1510)		(1.4718, 2.2036)	(-0.5095, 0.7570)	(0.5423, 2.2820)
λ_{smb}	3.6204	2.5213	-1.4281	ν_{smb}	1.0951	-0.004	-3.9534
	(0.7227, 6.5180)	(-0.4447, 5.4873)	(-5.3160, 2.4599)		(0.4817, 1.7084)	(-0.6371, 0.6291)	(-6.5456, -1.3612)
λ_{hml}	-6.0898	4.0495	-3.2287	ν_{hml}	-10.6556	-0.5163	-7.7946
	(-9.0497, -3.1298)	(1.0114, 7.0877)	(-7.8795, 1.4220)		(-11.3520, -9.9592)	(-1.2015, 0.1689)	(-11.3819, -4.2074)
λ_{bab}	0.9911	11.4076	5.091	ν_{bab}	-8.4720	1.9445	-4.3721
	(-2.6448, 4.6269)	(4.5581, 18.2571)	(-1.8042, 11.9863)		(-9.1717, -7.7723)	(-3.8604, 7.7494)	(-10.2309, 1.4867)
FF and QMJ							
λ_m	8.1452	6.0886	7.4174	ν_m	1.7735	-0.2831	1.0457
	(3.8656, 12.4249)	(1.7682, 10.4091)	(3.0666, 11.7682)		(1.4321, 2.1149)	(-0.8754, 0.3092)	(0.2617, 1.8297)
λ_{smb}	4.0889	2.825	-1.865	ν_{smb}	1.5636	0.2997	-4.3904
	(1.1913, 6.9866)	(-0.1202, 5.7702)	(-5.6541, 1.9240)		(1.0059, 2.1213)	(-0.2273, 0.8267)	(-6.8318, -1.9489)
λ_{hml}	-6.2321	4.0693	-1.871	ν_{hml}	-10.7979	-0.4966	-6.4369
	(-9.1920, -3.2721)	(1.03,7.1086)	(-6.0150, 2.2729)		(-11.4106, -10.1853)	(-1.1867, 0.1936)	(-9.3372, -3.5367)
λ_{qmj}	0.5346	3.7478	1.5075	ν_{qmj}	-4.0602	-0.8471	-3.0873
	(-1.8421, 2.9113)	(0.4677, 7.0278)	(-2.1228, 5.1378)		(-4.6324, -3.4880)	(-3.1076, 1.4134)	(-5.8315, -0.3432)
CAR							
λ_m	8.5656	7.3165	8.2175	ν_m	2.1939	0.9447	1.8458
	(4.2860, 12.8453)	(2.9662, 11.6668)	(3.8134, 12.6216)		(1.8101, 2.5776)	(0.1637, 1.7257)	(0.8062, 2.8854)
λ_{smb}	3.5430	2.5028	-2.1261	ν_{smb}	1.0176	-0.0225	-4.6514
	(0.6453, 6.4406)	(-0.5464, 5.5520)	(-5.9563, 1.7041)		(0.3382, 1.6971)	(-0.9719, 0.9269)	(-7.1562, -2.1466)
λ_{hml}	-4.9265	4.1996	-1.1333	ν_{hml}	-9.4923	-0.3662	-5.6992
	(-7.8864, -1.9665)	(1.0187, 7.3806)	(-5.4517, 3.1850)		(-10.4596, -8.5251)	(-1.5313, 0.7989)	(-8.8436, -2.5548)
λ_{mom}	8.0947	34.6689	6.4697	ν_{mom}	0.1817	26.7559	-1.4434
	(4.1398, 12.0495)	(16.6801, 52.6576)	(-4.5222, 17.4615)		(-1.9046, 2.2679)	(9.2072, 44.3045)	(-11.6991, 8.8123)

The table contains the estimates and the corresponding confidence intervals of the annualized risk premia and of the components of vector ν for the time-invariant specifications FF and BAB, FF and QMJ, and CAR estimated using the three different sets of base assets.

Table 8: Estimated annualized risk premia and ν for the time-invariant specifications (FF long-only, FF, CAPM)

	Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)		Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)
			FF long-only				
λ_m	7.4291	6.4138	7.9642	ν_m	1.0573	0.0421	1.5925
	(3.1494, 11.7087)	(2.0913, 10.7363)	(3.5971, 12.3314)		(0.6042, 1.5105)	(-0.5651, 0.6492)	(0.7227, 2.4622)
λ_s	9.6986	9.8810	5.6996	ν_s	0.0061	0.1886	-3.6629
	(4.0202, 15.3770)	(4.1980, 15.5641)	(-0.3736, 11.7728)		(-0.6064, 0.6187)	(-0.0413, 0.4185)	(-6.1469, -1.8388)
λ_h	5.9767	11.1603	6.3689	ν_h	-4.4425	0.7411	-4.0503
	(1.0205, 10.9330)	(6.1494, 16.1712)	(0.763877, 11.9739)		(-4.9540, -3.9309)	(0.0033, 1.4789)	(-6.6678, -1.4328)
			FF				
λ_m	8.0976	6.5214	7.9523	ν_m	1.7258	0.1497	1.5806
	(3.8179, 12.3772)	(2.2034, 10.8394)	(3.5786, 12.3260)		(1.3556, 2.0961)	(-0.4246, 0.7239)	(0.6783, 2.4828)
λ_{smb}	3.8179	2.3082	-2.6931	ν_{smb}	0.6037	-0.2171	-5.2184
	(0.2313, 6.0267)	(-0.6515, 5.2679)	(-6.4954, 1.1091)		(0.0112, 1.1962)	(-0.8201, 0.3859)	(-7.6803, -2.7566)
λ_{hml}	-5.3153	4.6262	-1.4475	ν_{hml}	-9.8812	0.0603	-6.0134
	(-8.2753, -2.3554)	(1.5374, 7.7150)	(-5.6417, 2.7467)		(-10.6116, -9.1509)	(-0.8228, 0.9434)	(-8.9849, -3.0418)
			CAPM				
λ_m	8.0110	7.9705	7.7133	ν_m	1.6392	1.5987	0.8016
	(3.8179, 12.3772)	(3.5459, 12.3951)	(2.7573, 11.5894)		(1.2939, 1.9849)	(0.4754, 2.7220)	(-0.2874, 1.8906)

The table contains the estimates and the corresponding confidence intervals of the annualized risk premia and of the components of vector ν for the time-invariant specifications FF long-only, FF and CAPM estimated using the three different sets of base assets.

Table 9: Test results for asset pricing restrictions for the time-invariant specifications (I)

	Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)	Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)
	Test of the null hypothesis $\mathcal{H}_0 : a(\gamma) = b(\gamma)'\nu$			Test of the null hypothesis $\mathcal{H}_0 : a(\gamma) = 0$		
			CAR and REV			
Test statistic	4.7465	29.2689	47.9956	13.2657	89.4889	103.8036
p-value	0.0000	0.0619	0.1284	0.0000	0.0000	0.0000
			FF and REV			
Test statistic	6.1670	75.0417	47.8065	23.0638	104.2557	108.6600
p-value	0.0000	0.0000	0.1575	0.0000	0.0000	0.0000
			5FF			
Test statistic	4.6307	64.4533	45.9120	22.3504	77.2595	129.9196
p-value	0.0000	0.0000	0.2075	0.0000	0.0000	0.0000
			MOM and REV			
Test statistic	4.9874	56.7129	48.8531	4.3279	107.8952	79.3954
p-value	0.0000	0.0000	0.1591	0.0000	0.0000	0.0009
			FF and BAB			
Test statistic	4.6153	69.5203	48.8758	21.1069	79.3056	109.2380
p-value	0.0000	0.0000	0.1585	0.0000	0.0000	0.0000
			FF and QMJ			
Test statistic	5.8987	50.4669	51.3657	23.8266	50.3339	103.5257
p-value	0.0003	0.0003	0.1075	0.0000	0.0019	0.0000

We compute the statistics $\tilde{\Sigma}_{\xi}^{-1/2}\hat{\xi}_{nT}$ based on $\hat{Q}_e$ and $\hat{Q}_a$ for the individual stocks to test the null hypotheses $\mathcal{H}_0 : \; a(\gamma) = b(\gamma)'\nu$ and $\mathcal{H}_0 : \; a(\gamma) = 0$, respectively. For $n = 25$ and $n = 44$, we compute the weighted statistics $T\hat{e}'\hat{V}^{-1}\hat{e}$ and $T\hat{a}'\hat{V}_a^{-1}\hat{a}$ (Gibbons, Ross and Shanken (1989)), where $\hat{e}$ and $\hat{a}$ are $n \times 1$ vectors with elements $\hat{e}_i$ and $\hat{a}_i$, and $\hat{V} = (\hat{v}_{ij})$ and $\hat{V}_a = (\hat{v}_{a,ij})$ are $n \times n$ matrices with elements $\hat{v}_{ij} = \hat{c}_{\nu}'\hat{Q}_x^{-1}\hat{S}_{ij}\hat{Q}_x^{-1}\hat{c}_{\nu}$, and $\hat{v}_{a,ij} = E_1'\hat{Q}_x^{-1}\hat{S}_{ij}\hat{Q}_x^{-1}E_1$.The table reports the p-values of the statistics.

Table 10: Test results for asset pricing restrictions for the time-invariant specifications (II)

	Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)	Stocks ($n = 10,827$)	FF ($n = 25$)	Indu. ($n = 44$)
	Test of the null hypothesis $\mathcal{H}_0 : a(\gamma) = b(\gamma)'\nu$			Test of the null hypothesis $\mathcal{H}_0 : a(\gamma) = 0$		
			CAR			
Test statistic	4.5831	33.0933	51.8536	13.8409	87.6456	101.6536
p-value	0.0000	0.0452	0.0992	0.0000	0.0000	0.0000
			FF long-only			
Test statistic	6.6873	100.6665	53.0506	18.8767	104.0156	101.1693
p-value	0.0000	0.0000	0.0983	0.0000	0.0000	0.0000
			FF			
Test statistic	7.0130	95.1795	52.8127	24.4025	99.7995	109.0464
p-value	0.0000	0.0000	0.1022	0.0000	0.0000	0.0000
			CAPM			
Test statistic	8.2587	117.8019	57.0650	11.1168	120.4961	69.8462
p-value	0.0000	0.0000	0.0739	0.0000	0.0000	0.0078

We compute the statistics $\tilde{\Sigma}_{\xi}^{-1/2}\hat{\xi}_{nT}$ based on $\hat{Q}_e$ and $\hat{Q}_a$ for the individual stocks to test the null hypotheses $\mathcal{H}_0 : \; a(\gamma) = b(\gamma)' \nu$ and $\mathcal{H}_0 : \; a(\gamma) = 0$, respectively. For $n = 25$ and $n = 44$, we compute the weighted statistics $T\hat{e}'\hat{V}^{-1}\hat{e}$ and $T\hat{a}'\hat{V}_a^{-1}\hat{a}$ (Gibbons, Ross and Shanken (1989)), where $\hat{e}$ and $\hat{a}$ are $n \times 1$ vectors with elements $\hat{e}_i$ and $\hat{a}_i$, and $\hat{V} = (\hat{v}_{ij})$ and $\hat{V}_a = (\hat{v}_{a,ij})$ are $n \times n$ matrices with elements $\hat{v}_{ij} = \hat{c}_{\nu}'\hat{Q}_x^{-1}\hat{S}_{ij}\hat{Q}_x^{-1}\hat{c}_{\nu}$, and $\hat{v}_{a,ij} = E_1'\hat{Q}_x^{-1}\hat{S}_{ij}\hat{Q}_x^{-1}E_1$.The table reports the p-values of the statistics.

Table 11: Test results for asset pricing restrictions for the time-varying specifications (I)

	Stocks ($n = 8,570$)	FF ($n = 25$)	Indu. ($n = 44$)	Stocks ($n = 8,570$)	FF ($n = 25$)	Indu. ($n = 44$)
	Test of the null hypothesis $\mathcal{H}_0 : \beta_1(\gamma) = \beta_3(\gamma)\nu$			Test of the null hypothesis $\mathcal{H}_0 : \beta_1(\gamma) = 0$		
			CAR and REV			
Test statistic	3.0770	6.2271	5.6791	7.0525	7.8362	6.5985
p-value	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000
			FF and REV			
Test statistic	3.9808	6.8372	5.5273	27.1026	8.3474	6.4567
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
			5FF			
Test statistic	1.2788	6.8708	5.2323	26.3653	7.9588	6.7106
p-value	0.1005	0.0000	0.0000	0.0000	0.0000	0.0000
			MOM and REV			
Test statistic	4.8628	5.2675	5.7136	4.7098	11.0140	6.7836
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
			FF and BAB			
Test statistic	2.2573	7.1090	5.7684	27.0739	7.4952	6.5827
p-value	0.0120	0.0000	0.0000	0.0000	0.0000	0.0000
			FF and QMJ			
Test statistic	3.4664	6.2304	5.3363	28.2372	6.4029	5.9830
p-value	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000

We compute the statistics $\tilde{\Sigma}_{\xi}^{-1/2}\hat{\xi}_{nT}$ to test the null hypotheses $\mathcal{H}_0 : \ \beta_1(\gamma) = \beta_3(\gamma)\nu$ and $\mathcal{H}_0 : \ \beta_1(\gamma) = 0$. The trimming levels are $\chi_{1,T} = 15$ and $\chi_{2,T} = 546/240$. For $n = 25$ and $n = 44$, we compute the test statistics $T\hat{Q}_e$ and $T\hat{Q}_a$. The table reports the p-values of the statistics.

Table 12: Test results for asset pricing restrictions for the time-varying specifications (II)

	Stocks ($n = 8,570$)	FF ($n = 25$)	Indu. ($n = 44$)	Stocks ($n = 8,570$)	FF ($n = 25$)	Indu. ($n = 44$)
	Test of the null hypothesis $\mathcal{H}_0 : \beta_1(\gamma) = \beta_3(\gamma)\nu$			Test of the null hypothesis $\mathcal{H}_0 : \beta_1(\gamma) = 0$		
			CAR			
Test statistic	2.8610	7.1562	5.6106	15.8094	7.9416	6.1547
p-value	0.0021	0.0000	0.0000	0.0000	0.0000	0.0000
			FF long-only			
Test statistic	4.5113	—	5.6740	20.3813	—	22.2895
p-value	0.0000	—	0.0000	0.0000	—	0.0000
			FF			
Test statistic	4.5848	7.6766	5.4059	4.6112	8.0618	5.9396
p-value	0.0000	0.0012	0.0026	0.0000	0.0000	0.0000
			CAPM			
Test statistic	7.6748	13.9932	7.5388	2.3109	15.7887	7.9218
p-value	0.0000	0.0015	0.0014	0.0104	0.0000	0.0000

We compute the statistics $\tilde{\Sigma}_{\xi}^{-1/2}\hat{\xi}_{nT}$ to test the null hypotheses $\mathcal{H}_0 : \ \beta_1(\gamma) = \beta_3(\gamma)\nu$ and $\mathcal{H}_0 : \ \beta_1(\gamma) = 0$. The trimming levels are $\chi_{1,T} = 15$ and $\chi_{2,T} = 546/240$. For $n = 25$ and $n = 44$, we compute the test statistics $T\hat{Q}_e$ and $T\hat{Q}_a$ for $n = 25$ and $n = 44$. The table reports the p-values of the statistics.

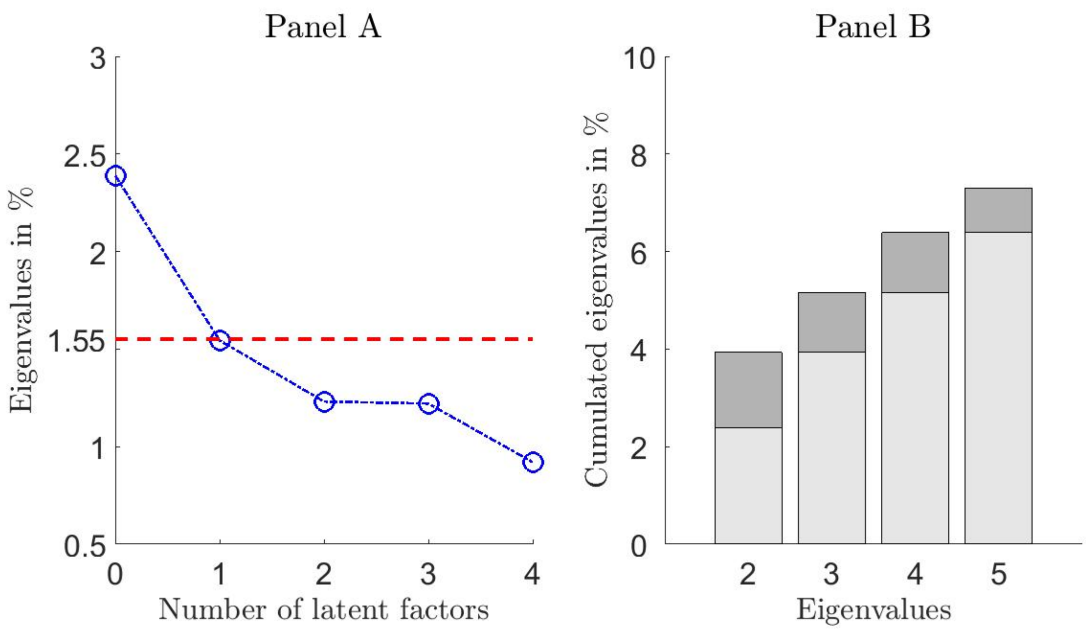


Figure 1: **Number of omitted factors and cumulated eigenvalues for the time-invariant CAPM model.** Panel A plots the scree-plot of the values of the first five eigenvalues in percentage, i.e., $\mu_j\left(\frac{1}{n^{\chi}T}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right)$ with $j = 1,...,5$. The horizonal line corresponds to the penalty function $g\left(n^{\chi},T\right)$. Panel B plots the cumulated eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1}\mu_l\left(\frac{1}{n^{\chi}T}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right)$, the dark grey area is the contribution of the j-th eigenvalue in percentage.

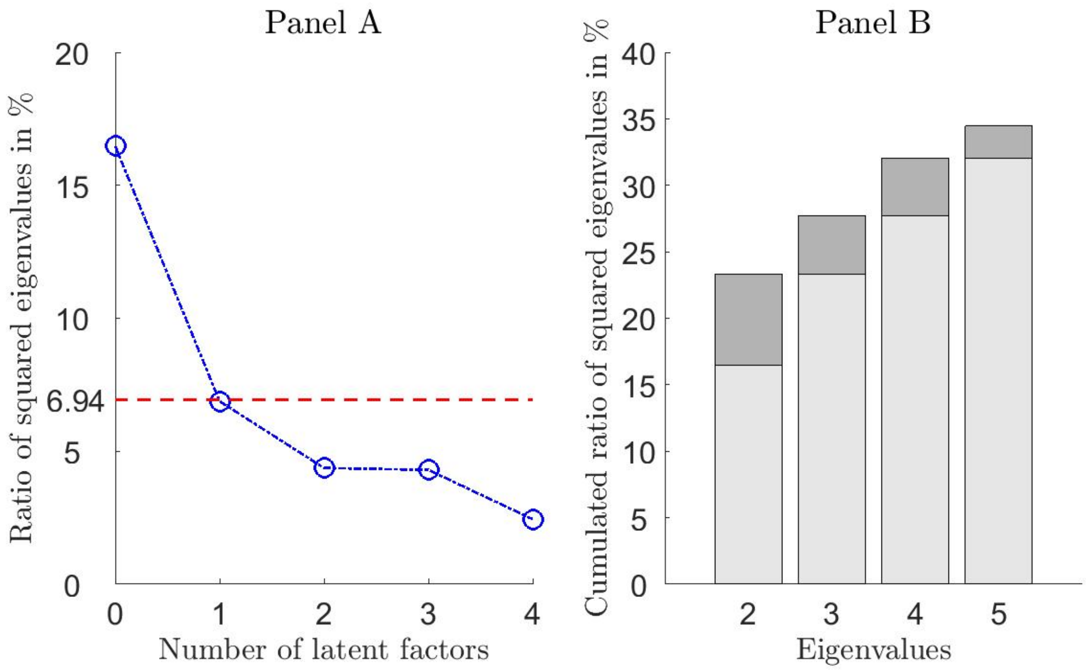


Figure 2: **Number of omitted factors and cumulated squared eigenvalues for the time-invariant CAPM model.** Panel A plots the scree-plot of the values of the first five squared eigenvalues in percentage, i.e., $\mu_j^2\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right) / \sum_{l=1}^{T}\mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$ with $j = 1, ..., 5$. The horizonal line corresponds to the penalty function $g\left(n^\chi, T\right)^2 / \sum_{l=1}^{T}\mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$. Panel B plots the cumulated squared eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1}\mu_l^2\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right) / \sum_{l=1}^{T}\mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$, the dark grey area is the contribution of the j-th squared eigenvalue in percentage.

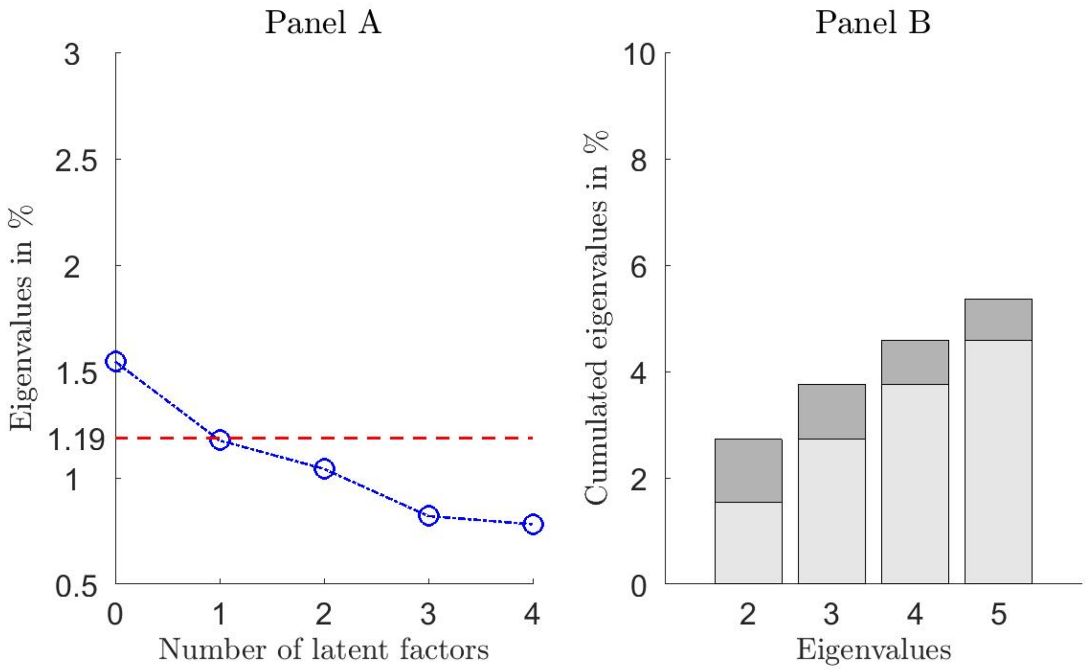


Figure 3: **Number of omitted factors and cumulated eigenvalues for the time-invariant three-factor Fama-French model.** Panel A plots the scree-plot of the values of the first five eigenvalues in percentage, i.e., $\mu_j \left(\frac{1}{n^{\chi} T} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right)$ with $j = 1, ..., 5$. The horizonal line corresponds to the penalty function $g(n^{\chi}, T)$. Panel B plots the cumulated eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1} \mu_l \left(\frac{1}{n^{\chi} T} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right)$, the dark grey area is the contribution of the j-th eigenvalue in percentage.

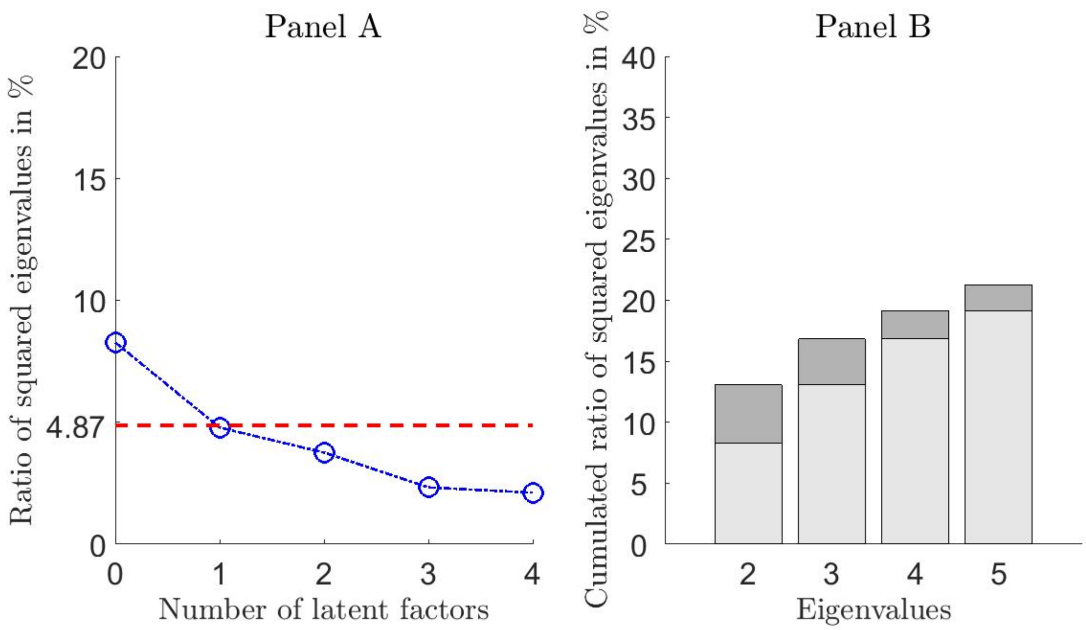


Figure 4: **Number of omitted factors and cumulated squared eigenvalues for the time-invariant three-factor Fama-French model.** Panel A plots the scree-plot of the values of the first five squared eigenvalues in percentage, i.e., $\mu_j^2\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)/\sum_{l=1}^T \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$ with $j = 1, ..., 5$. The horizonal line corresponds to the penalty function $g\left(n^\chi, T\right)^2/\sum_{l=1}^T \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$. Panel B plots the cumulated squared eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1} \mu_l^2\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)/\sum_{l=1}^T \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$, the dark grey area is the contribution of the j-th squared eigenvalue in percentage.

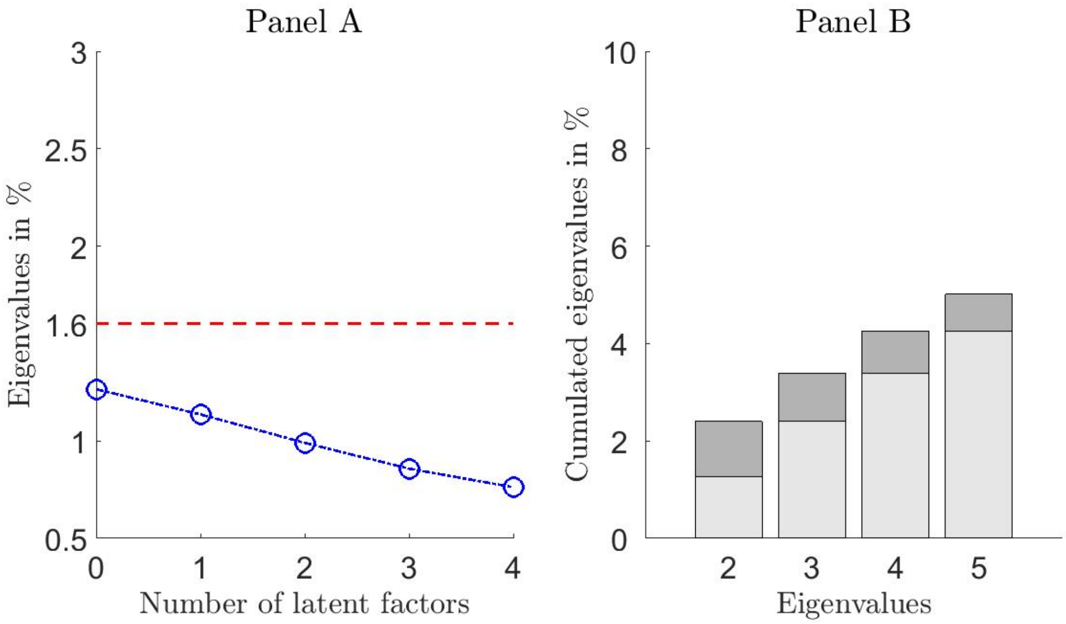


Figure 5: **Number of omitted factors and cumulated eigenvalues for the time-invariant four-factor CAR model.** Panel A plots the scree-plot of the values of the first five eigenvalues in percentage, i.e., $\mu_j \left(\frac{1}{n^{\chi} T} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right)$ with $j = 1, ..., 5$. The horizonal line corresponds to the penalty function $g\left(n^{\chi}, T\right)$. Panel B plots the cumulated eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1} \mu_l \left(\frac{1}{n^{\chi} T} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right)$, the dark grey area is the contribution of the j-th eigenvalue in percentage.

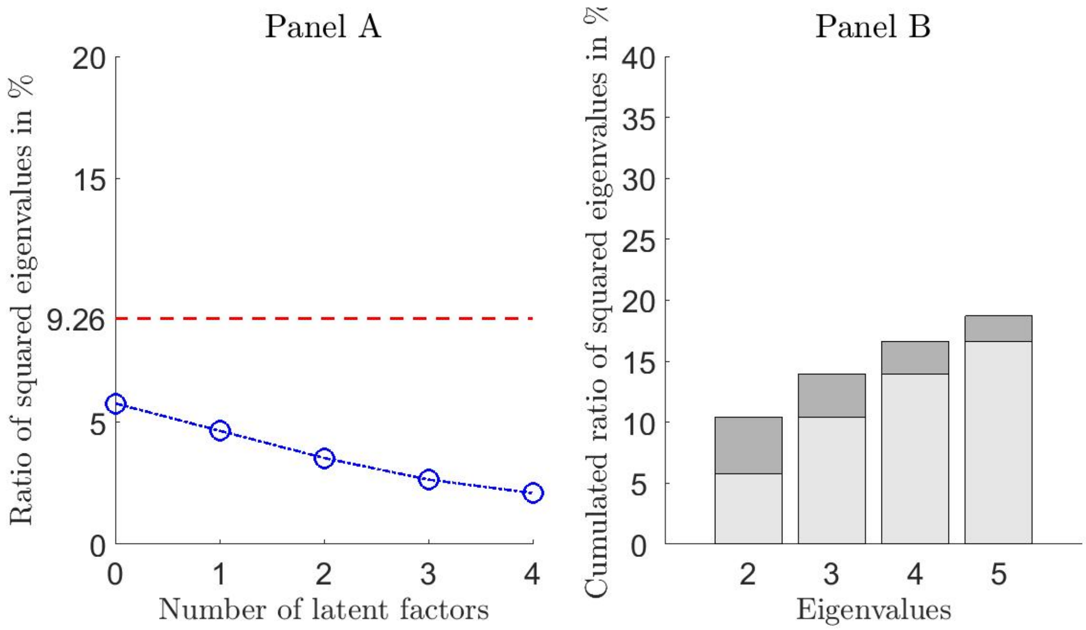
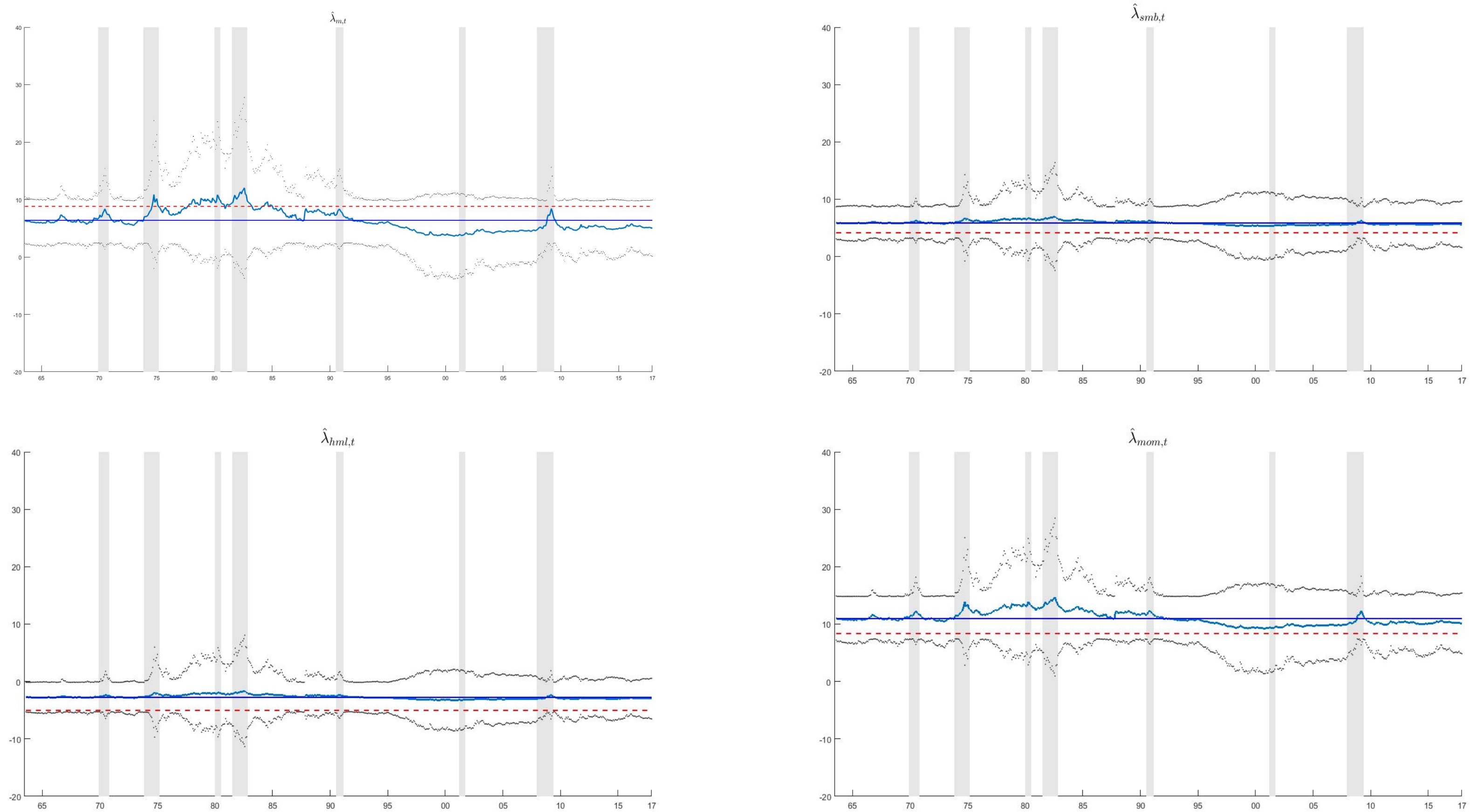


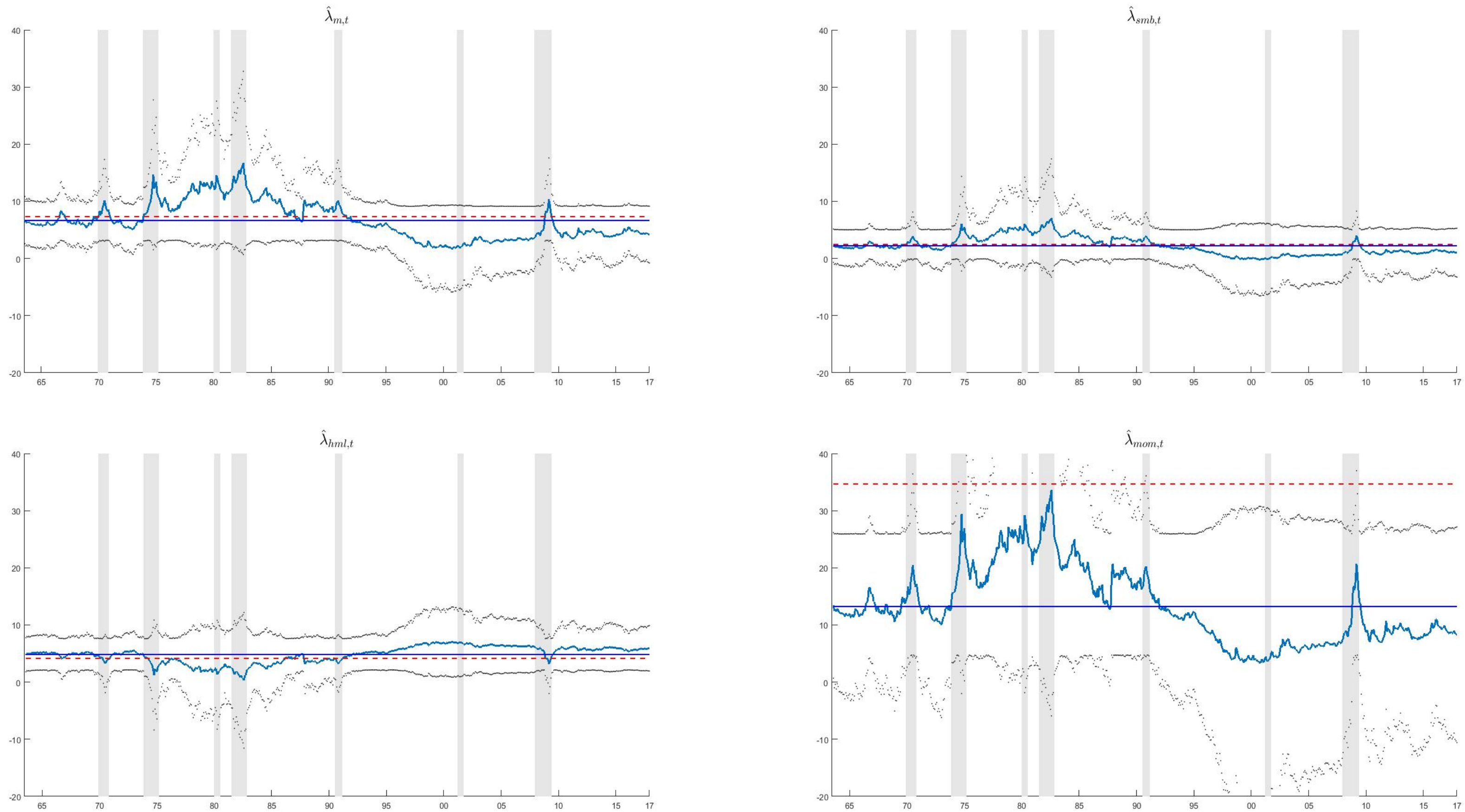
Figure 6: **Number of omitted factors and cumulated squared eigenvalues for the time-invariant four-factor CAR model.** Panel A plots the scree-plot of the values of the first five squared eigenvalues in percentage, i.e., $\mu_j^2\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right) / \sum_{l=1}^{T} \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$ with $j = 1, ..., 5$. The horizonal line corresponds to the penalty function $g\left(n^\chi, T\right)^2 / \sum_{l=1}^{T} \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$. Panel B plots the cumulated squared eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1} \mu_l^2\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right) / \sum_{l=1}^{T} \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$, the dark grey area is the contribution of the j-th squared eigenvalue in percentage.

Figure 7: Path of estimated annualized risk premia with $n = 8,570$ in the four-factor CAR model



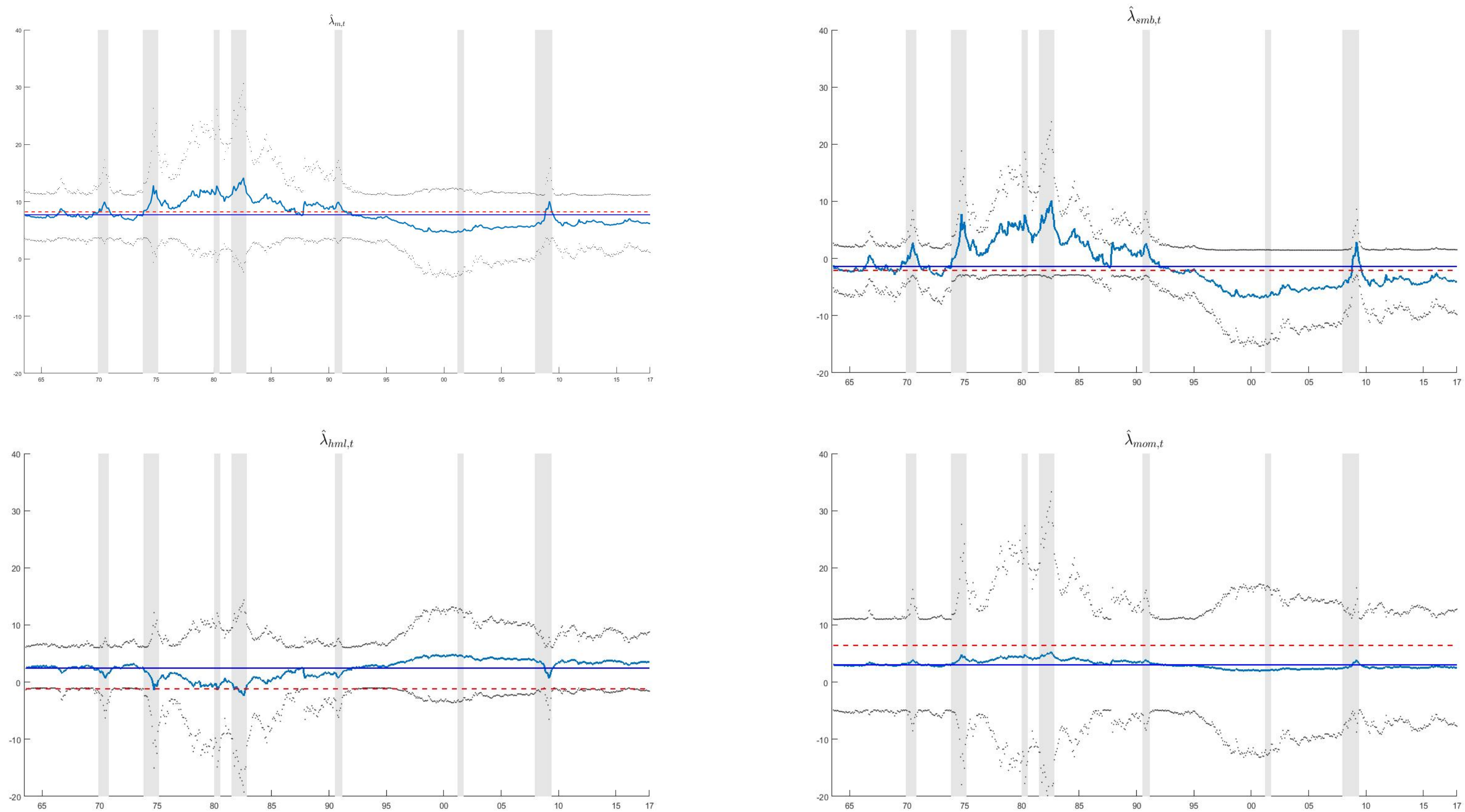
The figure plots the path of estimated annualized risk premia $\hat{\lambda}_{m,t}$, $\hat{\lambda}_{smb,t}$, $\hat{\lambda}_{hml,t}$, and $\hat{\lambda}_{mom,t}$ and their pointwise confidence intervals at 95% probability level in the four-factor CAR model. We also report the time-invariant (dashed horizontal line) and the average conditional estimate (solid horizontal line). We consider all stocks as base assets ($n = 8,570$ and $n^{\chi} = 2,549$). The vertical shaded areas denote recessions determined by the National Bureau of Economic Research (NBER). The recessions start at the peak of a business cycle and end at the trough.

Figure 8: Path of estimated annualized risk premia with $n = 25$ in the four-factor CAR model



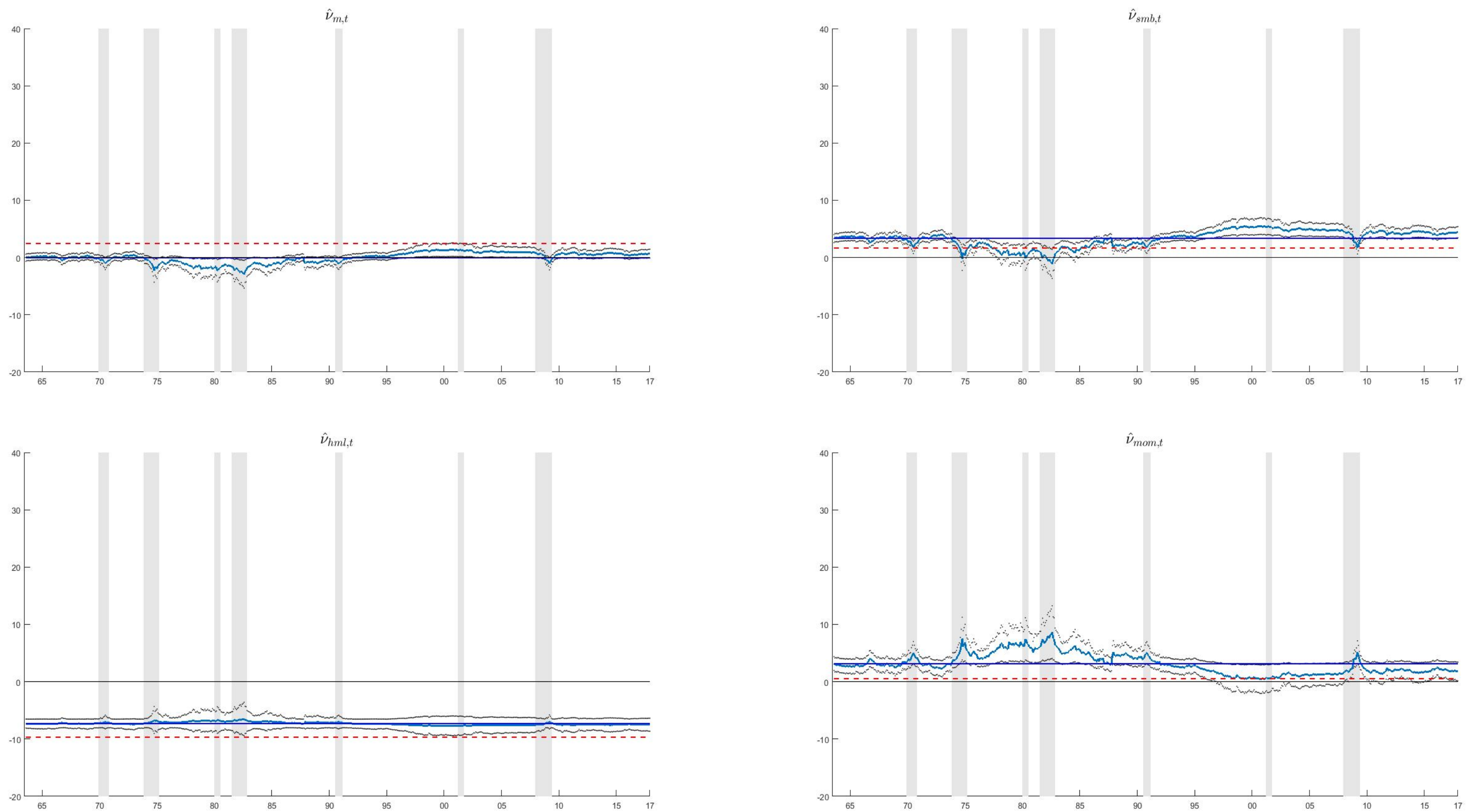
The figure plots the path of estimated annualized risk premia $\hat{\lambda}_{m,t}$, $\hat{\lambda}_{smb,t}$, $\hat{\lambda}_{hml,t}$, and $\hat{\lambda}_{mom,t}$ and their pointwise confidence intervals at 95% probability level in the four-factor CAR model. We use the returns of the 25 FF portfolios. We also report the time-invariant (dashed horizontal line) and the average conditional estimate (solid horizontal line). The vertical shaded areas denote recessions determined by the National Bureau of Economic Research (NBER).

Figure 9: Path of estimated annualized risk premia with $n = 44$ in the four-factor CAR model



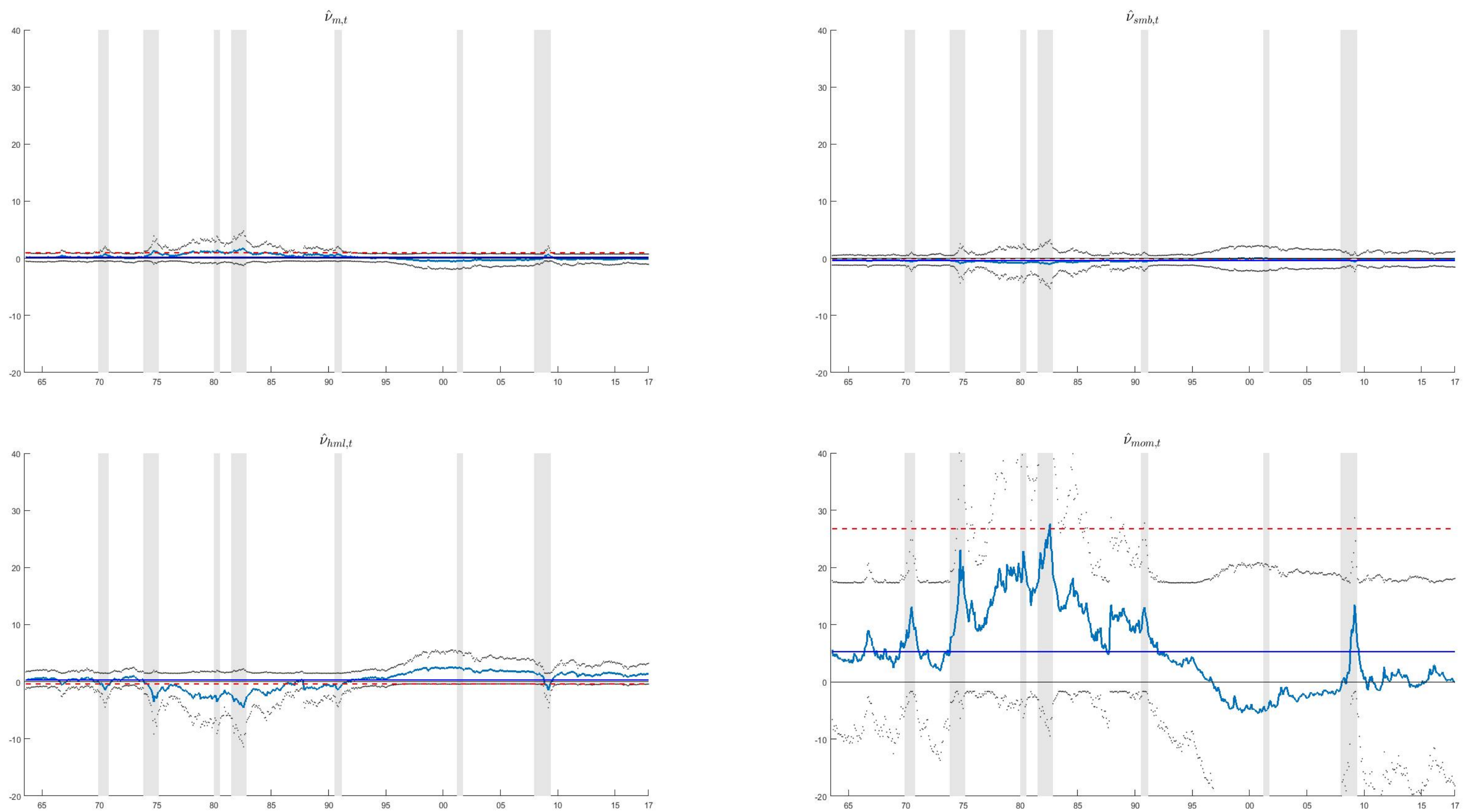
The figure plots the path of estimated annualized risk premia $\hat{\lambda}_{m,t}$, $\hat{\lambda}_{smb,t}$, $\hat{\lambda}_{hml,t}$, and $\hat{\lambda}_{mom,t}$ and their pointwise confidence intervals at 95% probability level in the four-factor CAR model. We use the returns of the 44 industry portfolios. We also report the time-invariant (dashed horizontal line) and the average conditional estimate (solid horizontal line). The vertical shaded areas denote recessions determined by the National Bureau of Economic Research (NBER).

Figure 10: Path of estimated annualized ν_t with $n = 8,570$ in the four-factor CAR model



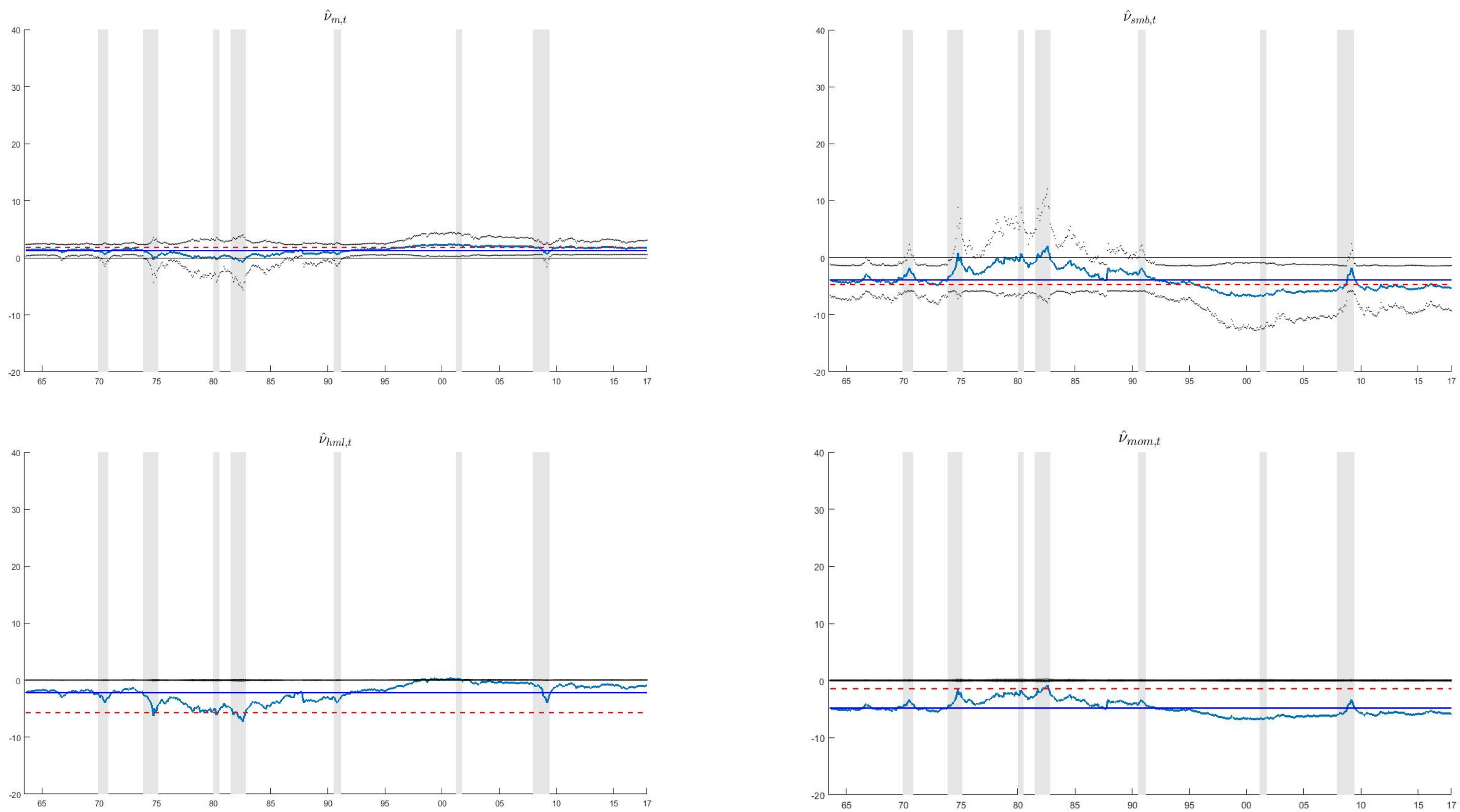
The figure plots the path of estimated annualized $\hat{\nu}_{m,t}$, $\hat{\nu}_{smb,t}$, $\hat{\nu}_{hml,t}$, and $\hat{\nu}_{mom,t}$ and their pointwise confidence intervals at 95% probability level in the four-factor CAR model. We also report the time-invariant (dashed horizontal line) and the average conditional estimate (solid horizontal line). We consider all stocks as base assets ($n = 8,570$ and $n^{\chi} = 2,549$). The vertical shaded areas denote recessions determined by the National Bureau of Economic Research (NBER). The recessions start at the peak of a business cycle and end at the trough.

Figure 11: Path of estimated annualized ν_t with $n = 25$ in the four-factor CAR model



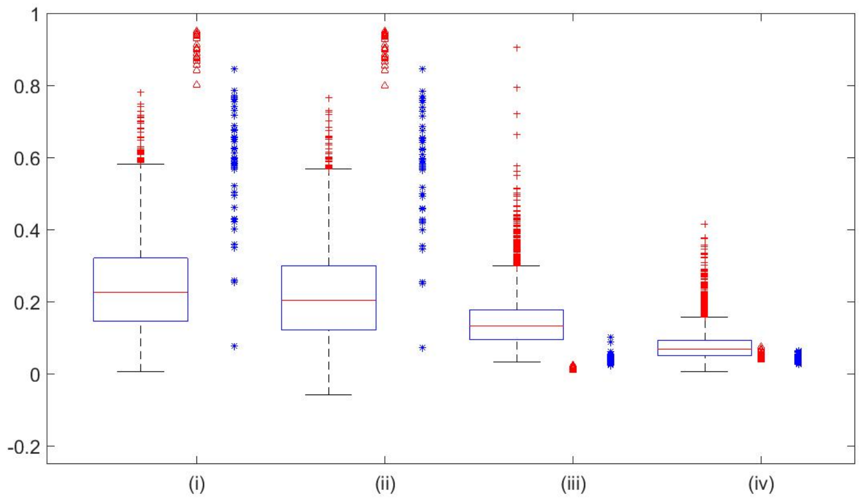
The figure plots the path of estimated annualized $\hat{\nu}_{m,t}$, $\hat{\nu}_{smb,t}$, $\hat{\nu}_{hml,t}$, and $\hat{\nu}_{mom,t}$ and their pointwise confidence intervals at 95% probability level in the four-factor CAR model. We use the returns of the 25 FF portfolios. We also report the time-invariant (dashed horizontal line) and the average conditional estimate (solid horizontal line). The vertical shaded areas denote recessions determined by the National Bureau of Economic Research (NBER).

Figure 12: Path of estimated annualized ν_t with $n = 44$ in the four-factor CAR model



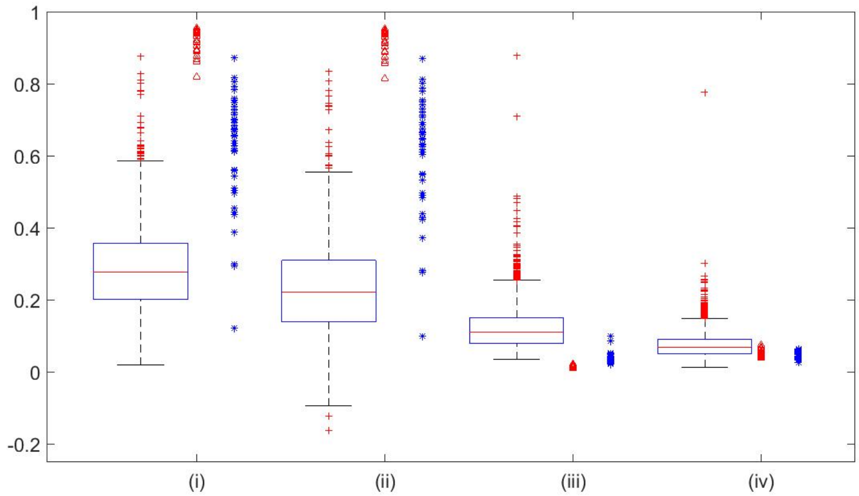
The figure plots the path of estimated annualized $\hat{\nu}_{m,t}$, $\hat{\nu}_{smb,t}$, $\hat{\nu}_{hml,t}$, and $\hat{\nu}_{mom,t}$ and their pointwise confidence intervals at 95% probability level in the four-factor CAR model. We use the returns of the 44 industry portfolios. We also report the time-invariant (dashed horizontal line) and the average conditional estimate (solid horizontal line). The vertical shaded areas denote recessions determined by the National Bureau of Economic Research (NBER).

Figure 13: Cross-sectional distributions of $\hat{\rho}_i^2$, $\hat{\rho}_{ad,i}^2$, $IdiVol_i$, and $SysRisk_i$ for the time-invariant four-factor CAR model



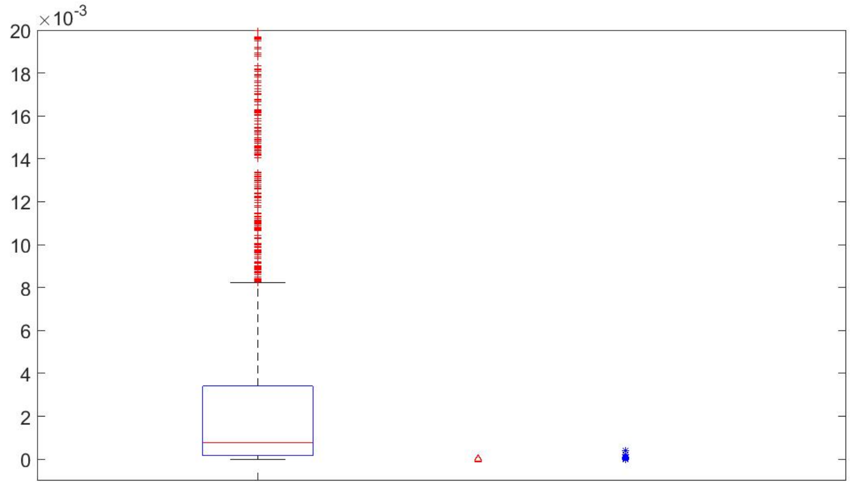
The figure displays the cross-sectional distributions of (i) the estimated coefficients of determination $\hat{\rho}_i^2$, (ii) the estimated adjusted coefficients of determination $\hat{\rho}_{ad,i}^2$, (iii) the idiosyncratic risks $IdiVol_i$, and (iv) the systematic risks $SysRisk_i$ for the individual stocks (box-plots), the 25 FF portfolios (red triangles) and the 44 Indu. portfolios (blue stars). Estimates are for the time-invariant four-factor CAR model. For comparison purposes, the cross-sectional distribution for individual stocks refers to the $n^{\chi} = 2,549$ stocks that are used in the estimation of the time-varying model after trimming.

Figure 14: Cross-sectional distributions of $\hat{\rho}_i^2$, $\hat{\rho}_{ad,i}^2$, $IdiVol_i$, and $SysRisk_i$ for the time-varying four-factor CAR model



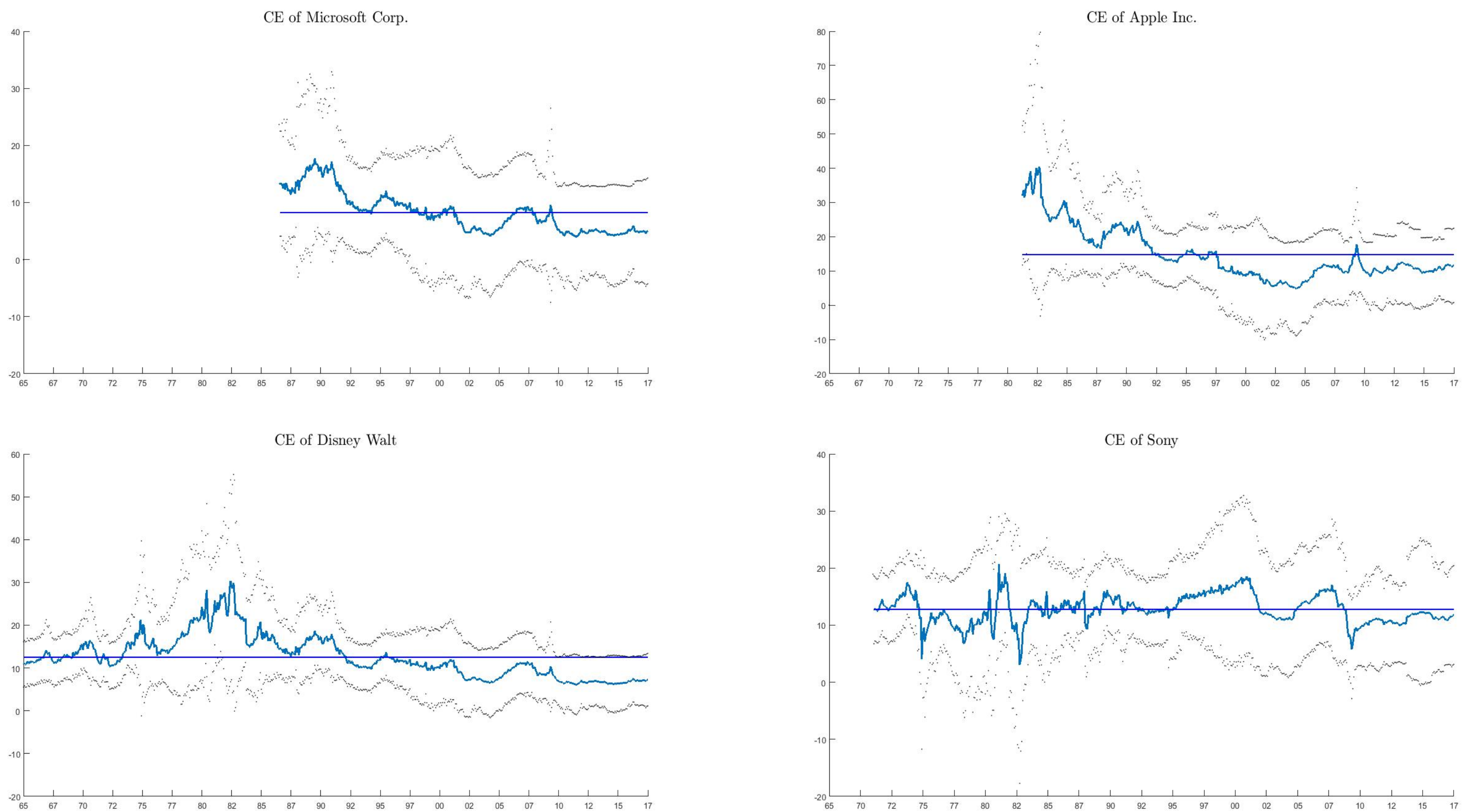
The figure displays the cross-sectional distributions of (i) the estimated coefficients of determination $\hat{\rho}_i^2$, (ii) the estimated adjusted coefficients of determination $\hat{\rho}_{ad,i}^2$, (iii) the idiosyncratic risks $IdiVol_i$, and (iv) the systematic risks $SysRisk_i$ for the $n^{\chi} = 2,549$ individual stocks (box-plots), the 25 FF portfolios (red triangles) and the 44 Indu. portfolios (blue stars). Estimates are for the time-varying four-factor CAR model.

Figure 15: Cross-sectional distributions of $\hat{\beta}_{1,i}'\hat{\beta}_{1,i}$ for the time-varying four-factor CAR model



The figure plots the cross-sectional distributions of $\hat{\beta}_{1,i}'\hat{\beta}_{1,i}$ for the $n^{\chi} = 2,549$ individual stocks (box-plot), the 25 FF portfolios (red triangles) and the 44 Indu. portfolios (blue stars). Estimated $\hat{\beta}_{1,i}$ are for the time-varying four-factor CAR model.

Figure 16: Path of estimated annualized costs of equity



The figure plots the path of estimated annualized costs of equity for Microsoft Inc., Apple Inc., Disney Walt, and Sony and their pointwise confidence intervals at 95% probability level. We use the time-varying four-factor CAR model estimated on individual stocks ($n = 8,570,\ n^{\chi} = 2,549$). We also report the average conditional estimate (solid horizontal line).

References

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

C. Ai and X. Chen. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6):1795–1843, 2003.

Y. Ait-Sahalia and D. Xiu. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 2:384–399, 2017.

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.

L. Alessi, M. Barigozzi, and M. Capasso. Improved penalization for determining the number of factors in approximate static factor models. *Statistics and Probability Letters*, 80:1806–1813, 2010.

S. Anatolyev and A. Mikusheva. Factor models with many assets: strong factors, weak factors, and the two-pass procedure. *Working Paper*, 2018.

E. Andreou, P. Gagliardini, E. Ghysels, and M. Rubin. Inference in group factor models with an application to mixed frequency data. *Econometrica*, 87(4):1267–1305, 2019.

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

D. W. K. Andrews and J. C. Monahan. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4):953–966, 1992.

A. Ang and D. Kristensen. Testing conditional factor models. *Journal of Financial Economics*, 106:132–156, 2012.

A. Ang, R. Hodrick, Y. Xing, and X. Zhang. High idiosyncratic volatility and low returns: International and further U.S. evidence. *Journal of Financial Economics*, 91:1–23, 2009.

A. Ang, J. Liu, and K. Schwarz. Using stocks or portfolios in tests of factor models. *Journal of Financial and Quantitative Analysis*, 55(3):709–750, 2020.

B. Antoine, H. Bonnal, and E. Renault. On the efficient use of the informational content of estimating equations: Implied probabilities and euclidean empirical likelihood. *Journal of Econometrics*, 138:461–487, 2007.

B. Antoine, K. Proulx, and E. Renault. Pseudo-true SDFs in conditional asset pricing model. *Journal of Financial Econometrics*, Forthcoming, 2020.

M. Arellano and S. Bonhomme. Identifying distributional characteristics in random coefficients panel data models. *Review of Economic Studies*, 79(3):987–1020, 2012.

C. S. Asness, A. Frazzini, and L.H. Pedersen. Quality minus junk. *Review of Accounting Studies*, 24(1): 34–112, 2019.

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

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

J. Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009.

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

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

Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The Annals of Probability*, 21(3):1275–1294, 1993.

N. Bailey, H. Pesaran, and V. Smith. A multiple testing approach to the regularisation of large sample correlation matrices. *Journal of Econometrics*, 208(2):507–534, 2019.

F. Bandi and V. Corradi. Nonparametric nonstationarity tests. *Econometric Theory*, 30:127–149, 2014.

L. Barras, O. Scaillet, and R. Wermers. False discoveries in mutual fund performance: measuring luck in estimated alphas. *Journal of Finance*, 65:179–216, 2010.

L. Barras, P. Gagliardini, and O. Scaillet. Skill and value creation in the mutual fund industry. *Working Paper*, 2019.

B. Bates, M. Plagborg-Moller, J. Stock, and M. Watson. Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics*, 177:289–304, 2013.

G. Bekaert, R. Hodrick, and X. Zhang. International stock return comovements. *Journal of Finance*, 64(6): 2591–2626, 2009.

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

P. J. Bickel and E. Levina. Covariance regularization by thresholding. *The Annals of Statistics*, 36(6): 2577–2604, 2008.

F. Black, M. Jensen, and M. Scholes. *The Capital Asset Pricing Model: Some empirical findings.* In Jensen, M.C. (Ed.), Studies in the Theory of Capital Markets. Praeger, New York, 1972.

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. Bonhomme. Functional differencing. *Econometrica*, 80(4):1337–1385, 2012.

J. Breitung and S. Eickmeier. Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163(1):71–84, 2011.

S. Bryzgalova. Spurious factors in linear pricing models. *Working paper*, 2016.

J. Campbell, J. Hilscher, and J. Szilagyi. In search of distress risk. *Journal of Finance*, 63(6):2899–2939, 2008.

J. Y. Campbell. Understanding risk and return. *Journal of Political Economy*, 104(2):298–345, 1996.

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

I. Chaieb, H. Langlois, and O. Scaillet. Factors and risk premia in individual international stock returns. *Journal of Financial Economics*, 2020, Forthcoming.

G. Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34:305–334, 1987.

G. Chamberlain. Efficiency bounds for semiparametric regression. *Econometrica*, 60(3):567–596, 1992.

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

S. Chava and A. Purnanandam. Is default risk negatively related to stock returns? *Review of Financial Studies*, 23(6):2523–2559, 2010.

L. Chen, J. Dolado, and J. Gonzalo. Detecting big structural breaks in large factor models. *Journal of Econometrics*, 180(1):30–48, 2014.

X. Cheng, Z. Liao, and F. Schorfheide. Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies*, 83(4):1511–1543, 2016.

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. *Asset pricing*. Princeton University Press, Princeton, 2005.

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

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. An intertemporal equilibrium beta pricing model. *Review of Financial*

Studies, 2(3):373–392, 1989.

G. Connor and O. Linton. Semiparametric estimation of a characteristic-based factor model of common stock returns. *Journal of Empirical Finance*, 14:694–717, 2007.

G. Connor, M. Hagmann, and O. Linton. Efficient semiparametric estimation of the Fama-French model and extensions. *Econometrica*, 80(2):713–754, 2012.

V. Corradi and N. Swanson. The effects of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo and a simple test. *Journal of Econometrics*, 132:195–229, 2006.

M. Cremers, A. Petajisto, and E. Zitzewitz. Should benchmark indices have alpha? Revisiting performance evaluation. *Critical Finance Review*, 2:1–48, 2012.

I. Dichev. Is the risk of bankruptcy a systematic risk? *Journal of Finance*, 53(3):1131–1147, 1998.

S. Donald, G. Imbens, and W. Newey. Choosing instrumental variables in conditional moment restriction models. *Journal of Econometrics*, 152:28–36, 2009.

B. Dumas and B. Solnik. The world price of foreign exchange risk. *Journal of Finance*, 50:445–479, 1995.

N. El Karoui. Operator norm consistent estimation of large dimensional sparse covariance matrices. *Annals of Statistics*, 36(6):2717–2756, 2008.

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. Industry costs of equity. *Journal of Financial Economics*, 43:153–193, 1997.

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

E. F. Fama and K. R. French. Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 105:457–472, 2012.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1): 1–22, 2015.

E. F. Fama and K. R. French. International tests of a five-factor asset pricing model. *Journal of Financial*

Economics, 123:441–463, 2017.

E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–36, 1973.

J. Fan and D. Kim. Robust high-dimensional volatility matrix estimation for high-frequency factor model. *Journal of the American Statistical Association*, 113:1268–1283, 2018.

J. Fan, Y. Liao, and M. Mincheva. High dimensional covariance matrix estimation in approximate factor structure. *Annals of Statistics*, 39(6):3320–3356, 2011.

J. Fan, Y. Liao, and J. Yao. Power enhancement in high dimensional cross-sectional tests. *Econometrica*, 83:1497–1541, 2015.

J. Fan, A. Furger, and D. Xiu. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data. *Journal of Business and Economic Statistics*, 34:489–503, 2016a.

J. Fan, Y. Liao, and W. Wang. Projected Principal Component Analysis in factor models. *Annals of Statistics*, 44:219–254, 2016b.

G. Feng, S. Giglio, and D. Xiu. Taming the factor zoo: A test of new factors. *Journal of Finance*, 75(3): 1327–1370, 2020.

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.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics*, 82(4):540–54, 2000.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: Consistency and rates. *Journal of Econometrics*, 119(2):231–55, 2004.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Society*, 100(4):830–40, 2005.

A. Frazzini and L.H. Pedersen. Betting against beta. *Journal of Financial Economics*, 111:1–25, 2014.

A. Frazzini, R. Israel, and T. Moskowitz. Trading costs of asset pricing anomalies. *Chicago Booth Research*, Paper No. 14-05, 2012.

J. Freyberger, A. Neuhierl, and M. Weber. Dissecting characteristics nonparametrically. *Review of Financial Studies*, 33(5):2326–2377, 2020.

P. Gagliardini and H. Ma. Extracting statistical factors when betas are time-varying. *Working Paper*, 2019.

P. Gagliardini and D. Ronchetti. Semi-parametric estimation of American option prices. *Journal of Econometrics*, 173:57–82, 2013.

P. Gagliardini and D. Ronchetti. Comparing asset pricing models by the conditional Hansen-Jagannathan distance. *Journal of Financial Econometrics*, 18(2):333–394, 2020.

P. Gagliardini, C. Gourieroux, and E. Renault. Efficient derivative pricing by the extended method of moments. *Econometrica*, 79(4):1181–1232, 2011.

P. Gagliardini, E. Ossola, and O. Scaillet. Time-varying risk premium in large cross-sectional equity datasets. *Econometrica*, 84(3):985–1046, 2016.

P. Gagliardini, E. Ossola, and O. Scaillet. A diagnostic criterion for approximate factor structure. *Journal of Econometrics*, 212(2):503–521, 2019.

L. Garlappi and H. Yan. Financial distress and the cross-section of equity returns. *Journal of Finance*, 66 (3):789–822, 2011.

S. Geman. A limit theorem for the norm of random matrices. *The Annals of Probability*, 8(2):252–261, 1980.

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.

R. Gibbons, S. A. Ross, and J. Shanken. A test of the efficiency of a given portfolio. *Econometrica*, 57(5): 1121–1152, 1989.

L. Gobillon and T. Magnac. Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3):535–551, 2016.

J. Gomes and L. Schmid. Levered returns. *Journal of Finance*, 65(2):467–494, 2010.

J. Gomes, L. Kogan, and L. Zhang. Equilibrium cross section of returns. *Journal of Political Economy*, 111 (4):693–732, 2003.

N. Gospodinov and T. Otsu. Local GMM estimation of time series models with conditional moment restrictions. *Journal of Econometrics*, 170:476–490, 2012.

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.

J. Griffin and M. Lemmon. Book-to-market equity, distress risk, and stock returns. *Journal of Finance*, 57 (5):2317–2336, 2002.

S. Gu, B. Kelly, and D. Xiu. Autoencoder asset pricing models. *Journal of Econometrics*, 2019, forthcoming.

J. Hahn and G. Kuersteiner. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica*, 70(4):1639–1657, 2002.

J. Hahn and W. Newey. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319, 2004.

M. Hallin and R. Liška. Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617, 2007.

X. Han and A. Inoue. Tests for parameter instability in dynamic factor models. *Econometric Theory*, 31(5): 1117–1152, 2015.

L. Hansen. A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics*, 30(1):203–238, 1985.

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.

L. P. Hansen and K. Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, 50(5):1269–1286, 1982.

C. R. Harvey and Y. Liu. Lucky factors. *Working Paper*, 2016.

C. R. Harvey, Y. Liu, and H. Zhu. ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68, 2016.

K. Hou, C. Xue, and L. Zhang. Digesting anomalies: an investment approach. *Review of Financial Studies*, 28:650–705, 2015.

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. Jagannathan and Z. Wang. An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *Journal of Finance*, 53(4):1285–1309, 1998.

R. Jagannathan and Z. Wang. Empirical evaluation of asset-pricing models: a comparison of the sdf and beta methods. *Journal of Finance*, 57(5):2337–2367, 2002.

R. Jagannathan, G. Skoulakis, and Z. Wang. The analysis of the cross section of security returns. *Handbook of Financial Econometrics*, 2:73–134, 2009.

I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295 – 327, 2001.

R. Kan and C. Zhang. Two-pass tests of asset pricing models with useless factors. *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.

R. Kan, C. Robotti, and J. Shanken. Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance*, 68(6):2617–2649, 2013.

S. Kandel and S. Stambaugh. Portfolio inefficiency and the cross-section of expected returns. *Journal of Finance*, 50(1):157–184, 1995.

B. Kelly, S. Pruitt, and Y. Su. Instrumented principal component analysis. *Working Paper*, 2017.

B. T. Kelly, S. Pruitt, and Y. Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524, 2019.

D. Kim and J. Fan. Factor GARCH-Itô models for high-frequency data with application to large volatility matrix prediction. *Journal of Econometrics*, 208:395–417, 2019.

S. Kim and G. Skoulakis. Ex-post risk premia estimation and asset pricing tests using large cross-sections: The regression calibration approach. *Journal of Econometrics*, 204(2):159–188, 2018.

M. King, E. Sentana, and S. Wadhwani. Volatility and links between national stock markets. *Econometrica*, 62(4):901–933, 1994.

Y. Kitamura, G. Tripathi, and H. Ahn. Empirical likelihood-based inference in conditional moment restriction models. *Econometrica*, 72:1667–1714, 2004.

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

S. Kozak, S. Nagel, and S. Santosh. Interpreting factor models. *Journal of Finance*, 73(3):1183–1223, 2018.

E. Lam and J. Wei. Limits-to-arbitrage, investment frictions, and the asset growth anomaly. *Journal of Financial Economics*, 102(1):127–149, 2011.

T. Lancaster. The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2):391–413, 2000.

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

A. Lewbel. A local generalized method of moments estimator,. *Economics Letters*, 94(1):124–128, 2007.

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

J. Lewellen, S. Nagel, and J. Shanken. A skeptical appraisal of asset-pricing tests. *Journal of Financial Economics*, 96(2):175–194, 2010.

J. Li, V. Todorov, and G. Tauchen. Jump factor models in large cross-sections. *Quantitative Economics*, 10: 419–456, 2019.

Y. Li and L. Yang. Testing conditional factor models: A nonparametric approach. *Journal of Empirical Finance*, 18:972–992, 2011.

Y. Liao and X. Yang. Uniform inference for characteristics effects of large continuous-time linear models. *Working Paper*, 2018.

S. Ludvigson. Advances in consumption-based asset pricing: Empirical tests. *Handbook of the Economics of Finance vol. 2B*, edited by G. Constantinides, M. Harris and R. Stulz, Amsterdam North-Holland: 799–906, 2013.

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

S. Ludvigson and S. Ng. Macro factors in bond risk premia. *Review of Financial Studies*, 22(12):5027–5067, 2009.

S. Ma, W. Lan, L. Su, and Tsai C.-L. Testing alphas in conditional time-varying factor models with high-dimensional assets. *Journal of Business and Economic Statistics*, 38(1):214–227, 2020.

C. MacKinlay. Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics*, 38(1):3–28, 1995.

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

H. Moon and M. Weidner. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579, 2015.

S. Nagel and K. Singleton. Estimation and evaluation of conditional asset pricing models. *Journal of Finance*, 66:873–909, 2011.

W. Newey. Efficient instrumental variables estimation of nonlinear models. *Econometrica*, 58:809–837, 1990.

W. K. Newey and K. D. West. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61(4):631–653, 1994.

J. Neyman and E. L. Scott. Consistent estimation from partially consistent observations. *Econometrica*, 16 (1):1–32, 1948.

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.

M. Pelger. Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics*, 4:23–42, 2019.

M. Pelger. Understanding systematic risk: A high-frequency approach. *Journal of Finance*, 75(4):2179–2220, 2020.

M. Pelger and R. Xiong. State-varying factor models of large dimensions. *Working Paper*, 2019.

M. H. Pesaran. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012, 2006.

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

L. Phalippou. Can risk-based theories explain the value premium? *Review of Finance*, 11(2):143–166, 2007.

J. Pontiff. Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics*, 42 (1-2):35–52, 2006.

V. Raponi, C. Robotti, and P. Zaffaroni. Testing beta-pricing models using large cross-sections. *Review of Financial Studies*, 33(6):2796–2842, 2020.

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

D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

J. Shanken. The arbitrage pricing theory: Is it testable? *Journal of Finance*, 37(5):1129–1140, 1982.

J. Shanken. Multivariate tests of the zero-beta CAPM. *Journal of Financial Economics*, 14(3):327–348, 1985.

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

J. Shanken. On the estimation of beta-pricing models. *Review of Financial Studies*, 5(1):1–33, 1992.

J. Shanken and G. Zhou. Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics*, 84(1):40–86, 2007.

W. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002a.

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

L. Su and X. Wang. On time-varying fector models: Estimation and testing. *Journal of Econometrics*, 198 (1):84–101, 2017.

L. Trapani. A randomised sequential procedure to determine the number of factors. *Journal of the American Statistical Association*, 113:1341–1349, 2018.

M. Vassalou and Y. Xing. Default risk in equity returns. *Journal of Finance*, 59(2):831–868, 2004.

Y.Q. Yin, Z.D. Bai, and P.R. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory*, 78:509–521, 1988.

P. Zaffaroni. Factor models for asset pricing. *Working paper*, 2019.

L. Zhang. The value premium. *Journal of Finance*, 60(1):67–103, 2005.