Journal of Financial Economics 118 (2015) 211–244

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Regression-based estimation of dynamic asset pricing models ☆

Tobias Adrian *, Richard K. Crump, Emanuel Moench

Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045, United States

A R T I C L E I N F O

Article history:
Received 11 April 2013
Received in revised form
19 June 2014
Accepted 27 June 2014
Available online 30 July 2015

JEL classification:
G10
G12
C58

Keywords:
Dynamic asset pricing
Fama-MacBeth regressions
Time-varying betas
Generalized method of moments
Minimum distance estimation
Reduced rank regression

A B S T R A C T

We propose regression-based estimators for beta representations of dynamic asset pricing models with an affine pricing kernel specification. We allow for state variables that are cross-sectional pricing factors, forecasting variables for the price of risk, and factors that are both. The estimators explicitly allow for time-varying prices of risk, time-varying betas, and serially dependent pricing factors. Our approach nests the Fama-MacBeth two-pass estimator as a special case. We provide asymptotic multistage standard errors necessary to conduct inference for asset pricing tests. We illustrate our new estimators in an application to the joint pricing of stocks and bonds. The application features strongly time-varying, highly significant prices of risk that are found to be quantitatively more important than time-varying betas in reducing pricing errors.

1. Introduction

Overwhelming evidence exists that risk premiums vary over time (Campbell and Shiller, 1988; Cochrane, 2011). Yet, widely used empirical asset pricing methods such as Fama and MacBeth (1973) two-pass regressions rely on the assumption that prices of risk are constant.

This paper proposes regression-based estimators for dynamic asset pricing models (DAPMs) with time-varying prices of risk. The estimators and associated standard errors are computationally as simple as Fama-MacBeth regressions, but they explicitly provide estimates of time-varying prices of risk, as well as estimates of the associated state variable dynamics. Our model combines key assumptions of the dynamic asset pricing models from fixed income applications with the computational ease of Fama-MacBeth regressions that are popular in empirical equity market research. The setup can also be viewed as a

☆ We are grateful to the referee, Stijn Van Nieuwerburgh, for a number of helpful suggestions that substantially improved the paper. We would like to thank Andrew Ang, Matias Cattaneo, Fernando Duarte, Darrell Duffie, Robert Engle, Arturo Estrella, Andreas Fuster, Eric Ghysels, Benjamin Mills, Monika Piazzesi, Karen Shen, Michael Sockin, Jonathan Wright, seminar participants at the Federal Reserve Bank of New York, the National Bureau of Economic Research (NBER) Summer Institute, and the Verein für Socialpolitik for helpful comments and discussions. A special thanks goes to Wayne Ferson for his detailed comments as our discussant at the NBER Asset Pricing meeting. Daniel Green, Ariel Zucker, and Benjamin Mills provided excellent research assistance. The views expressed in this paper are those of the authors and do not necessarily represent those of the Federal Reserve Bank of New York or the Federal Reserve System.

* Corresponding author.
E-mail addresses: tobias.adrian@ny.frb.org (T. Adrian), richard.crump@ny.frb.org (R.K. Crump), emanuel.moench@ny.frb.org (E. Moench).

http://dx.doi.org/10.1016/j.jfineco.2015.07.004
0304-405X/

reduced form representation of dynamic macro-finance models with time-varying prices of risk.

We distinguish three different types of aggregate state variables: risk factors, price of risk factors, and factors that are both. By risk factors, we refer to variables that are significant factors for the cross section of asset returns, i.e. they have nonzero betas. By price of risk factors, we refer to variables that significantly forecast the time series variation of excess returns but do not necessarily have nonzero betas. Prices of risk are assumed to be affine functions of price of risk factors. We show that by introducing this risk price specification into generic asset pricing models, one can derive simple regression-based estimators for all model parameters that are consistent and asymptotically normal under mild conditions.

Our baseline estimator is a three-step regression that can be described as follows. In the first step, shocks to the state variables are obtained from a time series vector autoregression (VAR). In the second step, asset returns are regressed in the time series on lagged price of risk factors and the contemporaneous innovations to the cross sectional pricing factors, generating predictive slopes and risk betas for each test asset. In the third step, price of risk parameters are obtained by regressing the constant and the predictive slopes from the time series regression on the betas cross-sectionally. We give asymptotic variance formulas that allow for conditional heteroskedasticity and correct for the additional estimation uncertainty arising from using generated regressors.

We show that this three-step estimator coincides with the Fama-MacBeth estimator when two conditions are met. First, state variables have to be uncorrelated across time. Second, prices of risk have to be constant. Our approach can thus be viewed as a dynamic version of the Fama-MacBeth estimator, nesting the popular unconditional estimator as a special case.

We also introduce an additional (quasi-) maximum likelihood estimator (QMLE). This estimator is replacing the third regression step with a simple eigenvalue decomposition. The QMLE is asymptotically equivalent to the three step regression estimator even in the case of conditional heteroskedasticity in the return errors. We show that in our model generalized method of moments (GMM) and minimum distance (MD) estimation are exactly equivalent and that the QMLE is a special case of this more general class of estimation approaches for certain choices of weighting matrix.

While our main results are extensions of classic results in the cross-sectional pricing literature to a dynamic setting, we provide new interpretations of results in the model when prices of risk are constant. For example, the equivalence between GMM and MD estimation implies that the cross sectional T^2 statistic of Shanken (1985) could be directly interpreted as a J-test for the moment restrictions of the static model.

We also extend the three-step regression estimator to the case where betas and the parameters in the vector autoregression of the state variables are time-varying. We assume that these parameters evolve smoothly over time and estimate them using a kernel regression approach pioneered by Robinson (1989). Kernel regressions have the appealing feature of nesting least squares rolling window regressions which are often used in the empirical literature (see, for example, Fama and French, 1997; Lewellen and Nagel, 2006; among many others). In our implementation, however, we use a Gaussian kernel estimator with data-driven bandwidth choice following Ang and Kristensen (2012).

The affine price of risk specification we use closely resembles affine term structure models.[1] Our approach thus lends itself to asset pricing applications across different asset classes. We present an empirical application for the cross section of size-sorted equity portfolios and maturity-sorted Treasury portfolios. We show that a parsimonious model with two pricing factors, two price of risk factors, and one factor that serves both roles fits this cross section of test assets very well on average, while, at the same time, giving rise to strongly significant time variation in risk premiums. We further find that allowing for time variation in prices of risk is more important than modeling time variation in factor risk exposures in terms of minimizing squared pricing errors of the model. In our application, traditional estimation approaches such as the one by Fama and MacBeth (1973) and Ferson and Harvey (1991) imply substantially larger pricing errors than the estimators we propose.

The remainder of the paper is organized as follows. Section 2 provides a discussion of the contribution of this paper to the existing literature. We present the dynamic asset pricing model in Section 3. We discuss estimation and inference when betas are assumed to be constant in Section 4. In Section 4.1, we formally present the link of the dynamic asset pricing estimator to the static Fama-MacBeth estimator, and we explain the contributions of our results to the existing literature in detail. In Section 5, we derive the corresponding estimator under the assumption that betas vary over time. We illustrate our estimators in an empirical application in Section 6. Section 7 concludes.

2. Related literature

Our approach can be seen as a generalization of the static Fama and MacBeth (1973) cross sectional asset pricing approach to dynamic asset pricing models. The empirical applications of the static Fama-MacBeth approach are too numerous to list, but some of the seminal works are Chen, Roll, and Ross (1986) and Fama and French (1992).

Some previous authors have extended the Fama-MacBeth approach to conditional asset pricing models. Ferson and Harvey (1991) use period-by-period Fama-MacBeth regressions to obtain estimates of time-varying market prices of risk, which they then regress on lagged conditioning variables. They find evidence for predictable variation in prices of risk and associate most of the

[1] For regression-based approaches to term structure models featuring an exponentially affine pricing kernel, see Adrian, Crump, and Moench (2013) and Abrahams, Adrian, Crump, and Moench (2014).

predictable variation in stock returns to time variation in risk compensation instead of time variation in betas. Our estimation approach generalizes the one used in Ferson and Harvey (1991) by allowing for estimation in the presence of serially correlated pricing factors and explicitly incorporating time variation of prices of risk. In addition, we provide asymptotic standard errors for all parameters of the model taking into account the uncertainty generated at each step of the estimation. Jagannathan and Wang (1996), Lettau and Ludvigson (2001), and others have used the Fama-MacBeth technology to estimate scaled factor models. The beta representations of such models are nested in our more general framework. Moreover, in contrast to our proposed estimators, the scaled factor approaches typically do not explicitly provide estimates for the price of risk parameters and the number of parameters grows quickly with the number of factors.

Our paper is further related to Balduzzi and Robotti (2010), who estimate time-varying risk premiums for maximum-correlation portfolios, i.e., portfolios resulting from the projection of a candidate pricing kernel on the set of test assets. Moreover, Gagliardini, Ossola, and Scaillet (2014) and Chordia, Goyal, and Shanken (2013) present alternative estimation approaches for models with time-varying risk premiums using Fama-MacBeth-type estimators when both the number of assets and the number of time series observations tend to infinity. Ang, Liu, and Schwarz (2010) study the implications for efficiency of using individual stocks versus portfolios in estimating cross-sectional pricing models. Finally, another strand of the literature investigates the implications of model misspecification in cross sectional asset pricing models. For example, Kan, Robotti, and Shanken (2013) derive the asymptotic distribution of the cross-sectional R^2 and develop model comparison tests which accommodate model misspecification. Here, instead, we assume that the model is correctly specified.

Our empirical application is closest to Ferson and Harvey (1991) and Campbell (1996), who use similar test assets and similar pricing factors in models with time-varying and constant prices of risk, respectively. A number of recent papers estimate dynamic pricing kernels for the cross section of stocks and bonds (see, for example, Mamaysky, 2002; Bekaert, Engstrom, and Grenadier, 2010; Lettau and Wachter, 2011; Ang and Ulrich, 2012; and Koijen, Lustig, and van Nieuwerburgh, 2013). What distinguishes our approach from that literature is the regression-based estimation methodology, which is simple to implement, is computationally robust, and allows for standard specification tests. We show that our empirical application features good pricing properties across stocks and bonds and that it implies notable time variation of expected returns associated with highly significant dynamic price of risk parameters. Moreover, the dynamic asset pricing model that we estimate yields substantially smaller mean squared pricing errors than several alternative models with constant prices of risk.

Some prior literature on conditional factor pricing models has assumed that betas are (linear) functions of observable variables, see, for example, Shanken (1990), Ferson and Harvey (1999), and, recently, Gagliardini, Ossola, and Scaillet (2014) and Chordia, Goyal, and Shanken (2013). A drawback to this approach is that it requires the correct specification for the functional form of the betas. As pointed out by Ghysels (1998) and Harvey (2001), models with misspecified betas often feature larger pricing errors than models with constant betas. In contrast, the kernel estimator that we use imposes less structure than assuming a specific functional form for the parameters and, therefore, is likely more robust to misspecification. Moreover, we show that our Gaussian kernel estimator yields smaller pricing errors than simple rolling window regressions for both specifications with constant and time-varying prices of risk.

We provide a further comparison of our results to the existing literature throughout the remainder of the paper.

3. Pricing kernel and return generating process

Before describing the model, it is convenient to introduce the following notations that are used throughout the paper. The symbol $\otimes$ represents the Kronecker product; vec$(\cdot)$, the vectorization operator. I_m and ι_n denote the $m \times m$ identity matrix and a $n \times 1$ column vector of ones, respectively. Moreover, let $[\Gamma_1 \,\vdots\, \Gamma_2]$ be the matrix formed by appending the columns of the matrix Γ_2 to the columns of the matrix Γ_1. Finally, throughout the paper, equalities involving conditional expectations are understood to hold almost surely.

We assume that systematic risk in the economy is captured by a $K \times 1$ vector of state variables X_t that follow a stationary vector autoregression:

$$X_{t+1} = \mu + \Phi X_t + v_{t+1}, \quad t = 0, \ldots, T-1, \tag{1}$$

with initial condition X_0. The dynamics of these state variables can be assumed to be generated by an equilibrium model of the macroeconomy.

The state variables can be risk factors, price of risk factors, or both. By risk factors, we refer to variables that are significant factors for the cross section. By price of risk factors, we refer to variables that significantly forecast the time variation of excess returns.[2] While some state variables act as both price of risk and risk factors, many commonly used state variables act exclusively as one or the other. This setup thus nests that of Campbell (1996), who argues that innovations in variables that have been shown to forecast stock returns should be used in cross-sectional asset pricing studies.

As a consequence, we partition the state variables into three categories: $X_{1,t} \in \mathbb{R}^{K_1}$: risk factor only, $X_{2,t} \in \mathbb{R}^{K_2}$: risk and price of risk factor, and $X_{3,t} \in \mathbb{R}^{K_3}$: price of risk factor only. In Section 6 we use all three types of factors in an application investigating the cross

[2] Variables which predict excess returns but are not contemporaneously correlated with excess returns are sometimes referred to as "unspanned" factors. For applications to affine term structure models with unspanned factors see, for example, Joslin, Priebsch, and Singleton (2012) or Adrian, Crump, and Moench (2013).

section of equity and bond returns. For simplicity of notation, we define

$$C_t = \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix}, \quad F_t = \begin{bmatrix} X_{2,t} \\ X_{3,t} \end{bmatrix}, \quad u_t = \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}, \tag{2}$$

where C_t is for cross section and F_t is for forecasting. Let $K_C = K_1 + K_2$, $K_F = K_2 + K_3$, and $K = K_1 + K_2 + K_3$. We assume that

$$\mathbb{E}[v_{t+1}|\mathcal{F}_t] = 0, \quad \mathbb{V}[v_{t+1}|\mathcal{F}_t] = \Sigma_{v,t}, \tag{3}$$

where $\mathcal{F}_t$ denotes the information set at time t. We denote holding period returns in excess of the risk free rate of asset i by $R_{i,t+1}$. We assume the existence of a pricing kernel M_{t+1} such that

$$\mathbb{E}[M_{t+1} R_{i,t+1}|\mathcal{F}_t] = 0. \tag{4}$$

Moreover, we assume that the pricing kernel has the linear form

$$\frac{M_{t+1} - \mathbb{E}[M_{t+1}|\mathcal{F}_t]}{\mathbb{E}[M_{t+1}|\mathcal{F}_t]} = -\lambda_t' \Sigma_{u,t}^{-1/2} u_{t+1}, \tag{5}$$

where λ_t is the $K_C \times 1$ vector of period t prices of risk and where the $K_C \times K_C$ matrix $\Sigma_{u,t}$ is the conditional variance of u_{t+1}. It is important to point out that the above form for the pricing kernel incorporates that the covariance $\mathbb{C}[R_{i,t+1}, v_{3,t+1}|\mathcal{F}_t] = 0$ for all t. The same restriction is imposed in term structure models that feature unspanned factors.

As in Duffee (2002), we assume that prices of risk are affine functions of the price of risk factors F_t, so that

$$\lambda_t = \Sigma_{u,t}^{-1/2} (\lambda_0 + \Lambda_1 F_t), \tag{6}$$

where λ_0 is a $K_C \times 1$ vector, Λ_1 is a $K_C \times K_F$ matrix, and $\Lambda = [\lambda_0 \,\vdots\, \Lambda_1]$ has full row rank. We then find the following beta representation of expected returns:

$$\begin{aligned} \mathbb{E}[R_{i,t+1}|\mathcal{F}_t] &= -\frac{\mathbb{C}[M_{t+1}, R_{i,t+1}|\mathcal{F}_t]}{\mathbb{E}[M_{t+1}|\mathcal{F}_t]} \\ &= \lambda_t' \Sigma_{u,t}^{-1/2} \mathbb{C}[u_{t+1}, R_{i,t+1}|\mathcal{F}_t] \\ &= (\lambda_0 + \Lambda_1 F_t)' \Sigma_{u,t}^{-1} \mathbb{C}[C_{t+1}, R_{i,t+1}|\mathcal{F}_t]. \end{aligned} \tag{7}$$

Thus,

$$\mathbb{E}[R_{i,t+1}|\mathcal{F}_t] = \beta_{i,t}' (\lambda_0 + \Lambda_1 F_t), \tag{8}$$

where $\beta_{i,t}$ is a (time-varying) K_C-dimensional exposure vector,

$$\beta_{i,t} = \Sigma_{u,t}^{-1} \mathbb{C}[C_{t+1}, R_{i,t+1}|\mathcal{F}_t] \tag{9}$$

We can then decompose excess returns into an expected and an unexpected component:

$$R_{i,t+1} = \beta_{i,t}' (\lambda_0 + \Lambda_1 F_t) + (R_{i,t+1} - \mathbb{E}[R_{i,t+1}|\mathcal{F}_t]). \tag{10}$$

The unexpected excess return $R_{i,t+1} - \mathbb{E}[R_{i,t+1}|\mathcal{F}_t]$ can be further decomposed into a component that is conditionally correlated with the innovations of the risk factors, $u_{t+1} = C_{t+1} - \mathbb{E}[C_{t+1}|\mathcal{F}_t]$, and a return pricing error $e_{i,t+1}$ that is conditionally orthogonal to the risk factor innovations:

$$\begin{aligned} R_{i,t+1} - \mathbb{E}[R_{i,t+1}|\mathcal{F}_t] &= \gamma_{i,t}' (C_{t+1} - \mathbb{E}[C_{t+1}|\mathcal{F}_t]) \\ &\quad + e_{i,t+1} = \gamma_{i,t}' u_{t+1} + e_{i,t+1}. \end{aligned} \tag{11}$$

By definition of $\beta_{i,t}$,

$$\gamma_{i,t} = \Sigma_{u,t}^{-1} \mathbb{C}[C_{t+1}, R_{i,t+1}|\mathcal{F}_t] = \beta_{i,t}, \tag{12}$$

so that

$$R_{i,t+1} = \beta_{i,t}' (\lambda_0 + \Lambda_1 F_t) + \beta_{i,t}' u_{t+1} + e_{i,t+1}. \tag{13}$$

The excess returns, $R_{i,t+1}$, thus depend on the expected excess return, $\beta_{i,t}'(\lambda_0 + \Lambda_1 F_t)$, the component that is conditionally correlated with the innovations to the risk factors, $\beta_{i,t}' u_{t+1}$, and a return pricing error, $e_{i,t+1}$, that is conditionally orthogonal to the risk factor innovations. Therefore, the innovations to the pricing factors C_t capture systematic risk exposure, and the levels of the price of risk factors F_t are forecasting variables.

Previous approaches have been taken to model the time variation in risk premiums in equity returns (for example, in Gibbons and Ferson, 1985; Campbell, 1987; Ferson and Harvey, 1991; Lettau and Ludvigson, 2001; among others). However, most, if not all, of these approaches can be viewed as special cases of our more general framework which has been derived from first principles. Affine prices of risk are also commonly used in the fixed income literature, see e.g., Duffee (2002), Dai and Singleton (2002), or Ang and Piazzesi (2003).

The system of equations (13) for $i = 1, \ldots, N$ embeds the no-arbitrage restrictions that were derived from the form of the pricing kernel introduced in Eq. (5). Relative to a seemingly unrelated regressions (SUR) model in which $R_{i,t+1} = a_{i,t} + c_{i,t} F_t + \beta_{i,t}' u_{t+1} + e_{i,t+1}$, the assumption of no-arbitrage implies $a_{i,t} = \beta_{i,t}' \lambda_0$ and $c_{i,t} = \beta_{i,t}' \Lambda_1$. These are reduced rank restrictions resulting in a smaller number of parameters to estimate. To the extent that the model is well specified, then the parameter restrictions imposed by no-arbitrage help in increasing the predictive accuracy for the entire cross-section of excess returns. Hence, in our dynamic asset pricing model, a clear connection exists between the cross-sectional pricing performance and the predictive ability of a given set of model factors.

Standard, static cross-sectional asset pricing models make two additional assumptions: $\Lambda_1 = 0$ in Eq. (13), and $\Phi = 0$ in Eq. (1) (see the reviews by Campbell, Lo, and MacKinlay, 1997 and Cochrane, 2005). We consider these special cases in the following sections. However, the main contribution of this paper is to study the dynamic case in which $\Phi \neq 0$ and $\Lambda_1 \neq 0$.

While the focus of this paper is the estimation of the beta representation of dynamic asset pricing models, an extensive literature estimates the stochastic discount factor (SDF) representation using GMM (Hansen, 1982). In that literature, the expression $\mathbb{E}[M_{t+1} R_{i,t+1}|\mathcal{F}_t] = 0$ is estimated directly (see Harvey, 1989, 1991). Singleton (2006) provides an overview of dynamic asset pricing estimators, Nagel and Singleton (2011) provide a GMM estimator with an optimal weighting matrix, and Roussanov (2014) proposes a nonparametric approach to estimating the SDF model.

4. Estimation with constant betas

In this section, we assume that $\beta_{i,t} = \beta_i$ for all i and t, and we analyze an extension of the model with time-varying $\beta_{i,t}$ in Section 5. We can then stack this model as

$$R = B\lambda_0 \iota_T' + B\Lambda_1 F_- + BU + E \tag{14}$$

and

$$X = \mu + \Phi X_- + V, \tag{15}$$

where $R = [R_1 \cdots R_T]$ is $N \times T$ with $R_t = (R_{1,t}, \ldots, R_{N,t})'$, $F_- = [F_0 \cdots F_{T-1}]$ is $K_F \times T$, $U = [u_1 \cdots u_T]$ is $K_C \times T$, $E = [e_1 \cdots e_T]$ is $N \times T$ with $e_t = (e_{1,t}, \ldots, e_{N,t})'$, $X = [X_1 \cdots X_T]$ is $K \times T$, and $V = [v_1 \cdots v_T]$ is $K \times T$. Hereafter, we assume that $N \geq K_C$. The parameters of the return equation are the stacked risk exposures B, which is a $N \times K_C$ matrix with rows composed of $\{\beta_i : 1 \leq i \leq N\}$ and the prices of risk, Λ.

We can nest the model in the SUR model,

$$R = A_0 \iota_T' + A_1 F_- + BU + E = A\tilde{Z} + E, \tag{16}$$

where A is a $N \times (K_C + K_F + 1)$ matrix, $\tilde{Z} = \left[\iota_T \mid F_-' \mid U'\right]'$ is of dimension $(K_C + K_F + 1) \times T$, and

$$A_0 = B\lambda_0, \quad A_1 = B\Lambda_1, \quad A = [A_0 \mid A_1 \mid B]. \tag{17}$$

In practice, we do not observe U so that we replace it with the residuals from OLS estimation of the VAR. The asymptotic variance formulas we provide in Theorem 1 incorporate the additional estimation uncertainty generated by replacing U with $\hat{U}$. In Appendix A, we provide explicit instructions on how to construct estimators and their associated standard errors. In Appendix B, we discuss how to impose linear restrictions on the parameters B and Λ and conduct inference on these restricted estimators. Here we focus on developing intuition for the form of the estimators and discussing their properties.

Let $\hat{Z} = \left[\iota_T \mid F_-' \mid \hat{U}'\right]'$ and $\hat{A}_{\text{ols}} = R\hat{Z}'\left(\hat{Z}\hat{Z}'\right)^{-1}$, and partition the estimator $\hat{A}_{\text{ols}}$ as $\hat{A}_{0,\text{ols}}$, $\hat{A}_{1,\text{ols}}$, and $\hat{B}_{\text{ols}}$, respectively, with associated heteroskedasticity-robust variance matrix estimator $\hat{\mathcal{V}}_{\text{rob}}$ [so that $\hat{\mathcal{V}}_{\text{rob}} \to_p \mathcal{V}_{\text{rob}}$ and $\sqrt{T}\left(\text{vec}\left(\hat{A}_{\text{ols}} - A\right)\right) \to_d \mathcal{N}(0, \mathcal{V}_{\text{rob}})$].

Given this parameterization, two natural approaches can be taken to estimating the parameters B, λ_0, and Λ_1. The first is an indirect approach based on backing out λ_0 and Λ_1 via

$$\lambda_0 = (B'WB)^{-1} B'WA_0, \quad \Lambda_1 = (B'WB)^{-1} B'WA_1, \tag{18}$$

for some positive-definite weight matrix W.[3] When $W = I_N$ this produces the regression-based counterpart to Eq. (18)

$$\hat{\lambda}_{0,\text{ols}} = \left(\hat{B}_{\text{ols}}'\hat{B}_{\text{ols}}\right)^{-1} \hat{B}_{\text{ols}}'\hat{A}_{0,\text{ols}}, \quad \hat{\Lambda}_{1,\text{ols}} = \left(\hat{B}_{\text{ols}}'\hat{B}_{\text{ols}}\right)^{-1} \hat{B}_{\text{ols}}'\hat{A}_{1,\text{ols}}. \tag{19}$$

We could consider alternative estimators that use data-dependent weight matrices, but we prefer this formulation in conjunction with heteroskedasticity-robust standard errors to avoid taking a stance on the exact form of the variance matrix of the return innovations.

The expressions in Eq. (19) can be interpreted as a three-step estimator in the following way. In the first step, shocks to the state variables are obtained from a time series vector autoregression. In the second step, asset returns are regressed in the time series on lagged price of risk factors and the contemporaneous innovations to the cross-sectional pricing factors, generating predictive slopes and risk betas for each test asset. In the third step, price of risk parameters are obtained by regressing the constant and the predictive slopes from the time series regression on the betas cross-sectionally. This three-step estimator was initially proposed by Adrian and Moench (2008) in an application to affine term structure models with a linear pricing kernel. In Section 4.1, we show that this estimator nests the two-pass regressions of Fama and MacBeth (1973), when $\Lambda_1 = 0$ and $\Phi = 0$. In Section 5, we further discuss the differences between our approach and the one proposed in Ferson and Harvey (1991). Heuristically, these authors first estimate λ_t from cross-sectional Fama-MacBeth regressions on time-varying betas and then Λ by regressing λ_t on a constant and lagged state variables.

The second regression-based approach is the following MD procedure,

$$\left(\hat{B}_{\text{md}}, \hat{\Lambda}_{\text{md}}\right) = \min_{B,\Lambda} \mathcal{Q}\left(B, \Lambda; \hat{A}_{\text{ols}}, W^{\text{md}}\right), \tag{20}$$

where

$$\mathcal{Q}\left(B, \Lambda; \hat{A}_{\text{ols}}, W^{\text{md}}\right) = T \cdot \text{vec}\left(\hat{A}_{\text{ols}} - B\left[\Lambda \mid I_{K_C}\right]\right)' W^{\text{md}} \text{vec}\left(\hat{A}_{\text{ols}} - B\left[\Lambda \mid I_{K_C}\right]\right). \tag{21}$$

This estimator finds the closest approximation of the unconstrained estimator, $\hat{A}_{\text{ols}}$, to values of B, λ_0, and Λ_1, which satisfy the restrictions in Eq. (17). This MD approach turns out to be exactly equivalent to the GMM estimator in this model and, under certain choices of W^{md}, nests the MLE if the error terms $\{e_t : 1 \leq t \leq T\}$ are jointly Gaussian.[4] Specifically, when the weighting matrix is $W^{\text{md}} = \left(\hat{Z}\hat{Z}' \otimes I_N\right)$, then the solutions to Eq. (21) are the MLEs under the assumption that $e_t \sim_{iid} \mathcal{N}(0, \sigma_e^2 \cdot I_N)$. We label these estimators as "quasi-maximum likelihood estimators" $\left(\hat{B}_{\text{qmle}}, \hat{\Lambda}_{\text{qmle}}\right)$. Closed-form expressions for these estimators are given in Appendix A. Specifically, these estimators replace the third regression step in the ordinary least squares (OLS) estimation with a simple eigenvalue decomposition.

In Theorem 1, we show that these two estimators are asymptotically equivalent under our assumptions, as they both converge to the same limiting normal distribution.

Theorem 1. Under our assumptions,

$$\sqrt{T}\,\text{vec}\left(\hat{\Lambda}_{\text{ols}} - \Lambda\right) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_\Lambda), \quad \sqrt{T}\,\text{vec}\left(\hat{\Lambda}_{\text{qmle}} - \Lambda\right) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_\Lambda),$$

[3] Here we assume that B is of full-column rank and is consequently strongly identified. For cases in which B could be weakly identified see Kleibergen (2009), Burnside (2010), Kleibergen and Zhan (2013), and Burnside (2011). In cases of weak identification, the robust test statistics of Kleibergen (2009) could be generalized to our setting. For weak identification robust inference in an SDF representation setting, see Gospodinov, Kan, and Robotti (2012).

[4] In addition, this equivalence combined with the results of Andrews and Lu (2001) could be used to produce an intuitive model-selection criterion to compare across specifications.

as $T\to\infty$, *where*

$$\mathcal{V}_\Lambda=\left(\Upsilon_{FF}^{-1}\otimes\Sigma_u\right)+\mathcal{H}_\Lambda(B,\Lambda)\mathcal{V}_{\text{rob}}\mathcal{H}_\Lambda(B,\Lambda)',$$

$$\Upsilon_{FF}=\operatorname*{p\,lim}_{T\to\infty}\tilde{F}_-\tilde{F}_-'/(T-1),\ \tilde{F}_-=[\iota_T\,\vdots\,F_-']',\ \text{and}$$

$$\mathcal{H}_\Lambda(B,\Lambda)=\left[\left(I_{(K_F+1)}\otimes(B'B)^{-1}B'\right)\vdots-\left(\Lambda'\otimes(B'B)^{-1}B'\right)\right].$$

The first term of $\mathcal{V}_\Lambda$ accounts for replacing the unobserved innovations U by estimated innovations. The second term accounts for all other sources of estimation uncertainty, including that of using an estimate of B to construct the estimator of Λ. Relative to the existing literature, Theorem 1 provides a number of insights. First, it extends feasible inference from the static Fama-MacBeth approach that assumes $\Phi=0$ and $\Lambda_1=0$ to the case with persistent factors and time-varying prices of risk. Second, Theorem 1 provides a generalization of Theorem 1 of Shanken (1992), which provides a correction for the uncertainty generated by estimating B to a setting with persistent factors and time-varying prices of risk [under conditional homoskedasticity, i.e., when $\mathcal{V}_{\text{rob}}=\left(\text{plim}_{T\to\infty}\left(\hat{Z}\hat{Z}'/T\right)\otimes\Sigma_e\right)$ for a positive-definite variance matrix Σ_e]. More generally, the results allow for conditionally heteroskedastic errors in the spirit of Theorem 1 of Jagannathan and Wang (1998), and so those results are extended to the dynamic setting as well. Finally, we show the asymptotic equivalence of the QML approach (a special case of GMM/MD, as mentioned above) and the OLS approach even under conditional heteroskedasticity, which is also an extension of Theorem 4 of Shanken (1992) both for constant and time-varying prices of risk.

Remark 1. (i) Although $\hat{\Lambda}_{\text{ols}}$ and $\hat{\Lambda}_{\text{qmle}}$ are asymptotically equivalent, the associated estimators of B are generally not. This is because the estimator $\hat{B}_{\text{ols}}$ is not constructed under the restrictions in Eq. (17). However, with a simple additional step, we can construct an estimator of B based on $\hat{\Lambda}_{\text{ols}}$ that is asymptotically equivalent to $\hat{B}_{\text{qmle}}$:

$$\hat{B}_{4\text{ols}}=R\left(\hat{\Lambda}_{\text{ols}}\tilde{F}_-+\hat{U}\right)'\left[\left(\hat{\Lambda}_{\text{ols}}\tilde{F}_-+\hat{U}\right)\left(\hat{\Lambda}_{\text{ols}}\tilde{F}_-+\hat{U}\right)'\right]^{-1}.$$

Intuitively, $\hat{B}_{4\text{ols}}$ is the *OLS* estimator of B taking the estimated prices of risk $\hat{\Lambda}_{\text{ols}}$ as given.

(ii) Under the assumption that $e_t|\mathcal{F}_{t-1}\sim_{iid}\mathcal{N}(0,\sigma_e^2\cdot I_N)$ and all variables are X_2-type variables, the estimators $\hat{\Lambda}_{\text{ols}}$ and $\hat{\Lambda}_{\text{qmle}}$ are asymptotically efficient. $\hat{B}_{\text{qmle}}$ and $\hat{B}_{4\text{ols}}$ are also asymptotically efficient, although $\hat{B}_{\text{ols}}$ is asymptotically efficient only when $N=K_C$.

In traditional asset pricing models with constant prices of risk, the parameter λ_0 determines whether a risk factor is priced in the cross section of test assets. However, when prices of risk are time varying, this parameter is no longer of independent interest. Instead, to gauge whether differential exposures to a given pricing factor result in significant spreads of expected excess returns, one has to test whether a specific element of $\bar{\lambda}$ is equal to zero, where

$$\bar{\lambda}=\lambda_0+\Lambda_1\mathbb{E}[F_t]. \tag{22}$$

Theorem 2. Under our assumptions,

$$\sqrt{T}\,\text{vec}\left(\hat{\bar{\lambda}}_{\text{ols}}-\bar{\lambda}\right)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{V}_{\bar{\lambda}}\right),\quad\sqrt{T}\,\text{vec}\left(\hat{\bar{\lambda}}_{\text{qmle}}-\bar{\lambda}\right)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{V}_{\bar{\lambda}}\right),$$

as $T\to\infty$, *where* $\mathcal{V}_{\bar{\lambda}}$ *is given in Appendix D.1.*

In Appendix D.1 we show that $\mathcal{V}_{\bar{\lambda}}$ is a simple expression that invokes quantities that are known in closed form and easy to compute. Using this result, we can form a t-statistic of the null hypothesis that the sample average of the market price of risk for a given pricing factor is equal to zero. This allows us to test whether a given factor is unconditionally priced in the cross section of test assets.

4.1. Relation to Fama-MacBeth regressions

Standard factor pricing models assume that prices of risk are constant and that the pricing factors are unforecastable. Hence, the prevalent factor model used in the literature implicitly assumes that data are generated by[5]

$$R_{i,t+1}=\beta_i'\lambda_0+\beta_i'v_{t+1}+e_{i,t+1} \tag{23}$$

and

$$X_{t+1}=\mu+v_{t+1},\quad t=0,\ldots,T-1; \tag{24}$$

see, for example, Cochrane (2005, p. 276). This setup is nested in our model if $\Phi=0$ and $\Lambda_1=0$. This model is most commonly estimated by the two-pass Fama-MacBeth estimator (Fama and MacBeth, 1973) whose properties have been studied by Shanken (1992), Jagannathan and Wang (1998), and Shanken and Zhou (2007) among many others. The Fama-MacBeth estimator for λ_0 is

$$\hat{\lambda}_{0,\text{ols}}^{\text{FM}}=\left(\hat{B}_{\text{ols}}'\hat{B}_{\text{ols}}\right)^{-1}\hat{B}_{\text{ols}}'\hat{A}_{0,\text{ols}}, \tag{25}$$

where A_0 is the estimated constant term from a contemporaneous regression of returns on de-meaned factors. For comparison with Theorem 1, note that under our assumptions it can be shown that

$$\sqrt{T}\left(\hat{\lambda}_{0,\text{ols}}^{\text{FM}}-\lambda_0\right)\xrightarrow{d}\mathcal{N}(0,\mathcal{V}_\Lambda^{\text{FM}}),$$
$$\mathcal{V}_\Lambda^{\text{FM}}=\Sigma_u+\mathcal{H}_\Lambda^{\text{FM}}(B,\lambda_0)\mathcal{V}_{\text{rob}}^{\text{FM}}\mathcal{H}_\Lambda^{\text{FM}}(B,\lambda_0)',$$
$$\mathcal{H}_\Lambda^{\text{FM}}(B,\Lambda)=\left[(B'B)^{-1}B'\,\vdots\,-\left(\lambda_0'\otimes(B'B)^{-1}B'\right)\right], \tag{26}$$

where $\mathcal{V}_{\text{rob}}^{\text{FM}}$ is the probability limit of the heteroskedasticity-robust variance matrix from a contemporaneous regression of returns on factors and a constant. Because we allow for conditional heteroskedasticity, the variance matrix $\mathcal{V}_\Lambda^{\text{FM}}$ is in the spirit of that obtained by Jagannathan and Wang (1998) when the risk-free rate is observed. Similarly, the variance expression derived in Shanken (1992) can be obtained by using $\mathcal{V}_\Lambda^{\text{FM}}$ with $\mathcal{V}_{\text{rob}}^{\text{FM}}$

[5] Here we are assuming that the risk-free rate is observed, so the model does not include the zero-beta rate. Similar results can be obtained with the inclusion of a zero-beta rate.

formed under the assumption of conditionally homoskedastic errors.

The analogous estimator, $\hat{\lambda}_{0,\text{qmle}}^{\text{FM}}$, has received relatively less attention in the literature than its counterpart, derived under the assumption that $e_t \sim_{iid} \mathcal{N}(0, \Sigma_e)$.[6] As in the more general case above, $\hat{\lambda}_{0,\text{ols}}^{\text{FM}}$ and $\hat{\lambda}_{0,\text{qmle}}^{\text{FM}}$ are still asymptotically equivalent so that $\sqrt{T}\left(\hat{\lambda}_{0,\text{qmle}}^{\text{FM}} - \lambda_0\right) \rightarrow_d \mathcal{N}(0, \mathcal{V}_\Lambda^{\text{FM}})$ even in the presence of conditional heteroskedasticity. To our knowledge, this has not previously been pointed out in the literature. Following similar steps as in Appendix C, even with the inclusion of a zero-beta rate, the direct equivalence between the MD and GMM estimators (for any choice of weight matrix) and the MLE (for specific choices of weight matrix) can be established for the model of Eqs. (23) and (24). Special cases of this result have been pointed out in the literature. Ahn and Gadarowski (1999) discuss, and Kan and Chen (2005) show, the equivalence between the MD estimator and the MLE. More recently, Shanken and Zhou (2007) show the equivalence between the GMM estimator and the MLE (see also Zhou, 1994; Kleibergen, 1998).

It follows from the equivalence between MD and GMM estimation for the model of Eqs. (23) and (24) that the J-statistic is equivalent to the MD criterion function [i.e., Eq. (21)]. Thus, the cross-sectional T^2 statistic of Shanken (1985) [see Lewellen, Nagel, and Shanken (2010) for a detailed discussion of the test statistic], which corresponds to the MD criterion function when there is an unknown zero-beta rate (evaluated at the two-pass estimators) may be interpreted directly as a J-test of the moment restrictions for the model. This is an intuitively appealing interpretation because the J-statistic is then a direct joint test of the cross-sectional asset pricing restrictions imposed by the assumption of no-arbitrage. This is consistent with Lewellen, Nagel, and Shanken (2010), which emphasizes the importance of analyzing the estimators of all the parameters of the model instead of solely focusing on the price of risk. More generally, one key part of our contribution is to extend the static setting discussed here to the dynamic setting introduced in Section 3 without compromising the simplicity of implementation that has made the Fama-MacBeth estimator so popular in the applied finance literature.

Some authors apply the Fama-MacBeth estimator in model specifications with constant prices of risk, where the pricing factors are given by the VAR(1) innovations of a vector of state variables (see, for example, Chen, Roll, and Ross, 1986; Campbell, 1996; Petkova, 2006). These specifications thus rely on the return generating process:

$$R_{i,t+1} = \beta_i'\lambda_0 + \beta_i' v_{t+1} + e_{i,t+1} \tag{27}$$

and

$$X_{t+1} = \mu + \Phi X_t + v_{t+1}, \quad t = 0, \ldots, T-1. \tag{28}$$

[6] See, for example, Gibbons (1982), Kandel (1984), Roll (1985), Shanken (1985, 1986), Kan and Chen (2005), Shanken and Zhou (2007), and Kleibergen (2009), among others.

As an exercise, consider the case in which the true data generating process is governed by Eqs. (1) and (13) so that the prices of risk vary over time but are mistakenly assumed to be governed by Eqs. (27) and (28) and estimated via two-pass Fama-MacBeth regressions. Interestingly, it can be shown that in this case $\sqrt{T}\left(\hat{\lambda}_0^{\text{FM}} - \overline{\lambda}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{V}_{\overline{\lambda}}\right)$ (see Theorem 2). Thus, the conventional estimator is consistent for the parameter $\overline{\lambda}$. However, Wald-type test statistics would commonly be constructed using a plug-in version of the variance formula of Shanken (1992), which under technical conditions, converges in probability to $\Sigma_v + \left(1 + \overline{\lambda}'\Sigma_v^{-1}\overline{\lambda}\right) \cdot (B'B)^{-1} B'\Sigma_e B(B'B)^{-1}$. Comparing this expression and that of $\mathcal{V}_{\overline{\lambda}}$ from Appendix D.1 shows that the bias of the standard variance estimator depends on the values of Λ, Φ, and Σ_v.

5. Estimation with time-varying betas

A large literature exists on estimating beta representations of asset pricing models assuming that the betas vary over time. For example, see Fama and MacBeth (1973), Ferson and Harvey (1991), and many more. In this section, we discuss estimation of our model in the case in which factor risk exposures as well as the parameters governing the dynamics of the factors are time-varying. The model is, therefore,

$$R_{i,t+1} = \beta_{i,t}'\lambda_0 + \beta_{i,t}'\Lambda_1 F_t + \beta_{i,t}' u_{t+1} + e_{i,t+1} \tag{29}$$

and

$$X_{t+1} = \mu_t + \Phi_t X_t + v_{t+1}. \tag{30}$$

To motivate our estimator consider the case in which the innovations $\{u_t\}$ and the betas are known. In addition, let $B_t = \left(\beta_{1,t}, \ldots, \beta_{N,t}\right)'$. Then, passing through the vectorization operator yields

$$R_{t+1} - B_t u_{t+1} = \left(\tilde{F}_t' \otimes B_t\right)\text{vec}(\Lambda) + e_{t+1}. \tag{31}$$

From there it is easy to see that the associated estimator of the price of risk is

$$\text{vec}\left(\tilde{\Lambda}_{\text{ols}}^{\text{tv}}\right) = \left(\sum_{t=0}^{T-1}\left(\tilde{F}_t\tilde{F}_t' \otimes B_t'B_t\right)\right)^{-1} \sum_{t=0}^{T-1}\left(\tilde{F}_t \otimes B_t'\right)(R_{t+1} - B_t u_{t+1}). \tag{32}$$

In practice, the estimator of Eq. (32) is infeasible without estimates of B_t, μ_t, and Φ_t. Furthermore, without additional assumptions, identification of these parameters would be impossible as the number of parameters grows too quickly as $T \rightarrow \infty$.

One approach to identify time variation in $\beta_{i,t}$ that has been used in the literature is to posit that the parameters $\beta_{i,t}$ are (linear) functions of observable variables (see, for example, Shanken, 1990; Ferson and Harvey, 1999; Gagliardini, Ossola, and Scaillet, 2014; and Chordia, Goyal, and Shanken, 2013). However, a drawback to this approach is that it requires the correct specification for the functional form of the $\beta_{i,t}$. In fact, as pointed out by

Ghysels (1998) and Harvey (2001), among others, the beta estimates obtained in this way are typically sensitive to the specification of the information set. As a consequence, the magnitude of the resulting estimated pricing errors can vary substantially with the choice of conditioning variables. Other limitations to this approach are that the number of regressors can grow large and that commonly-used conditioning variables are available only at low frequencies.

An alternative identifying assumption is that

$$\beta_{i,t}=\beta_i(t/T)+o(1),\quad \mu_t=\mu(t/T)+o(1),$$

$$\Phi_t=\Phi(t/T)+o(1), \tag{33}$$

where all $\beta_i(\cdot)$, $\mu(\cdot)$, and $\Phi(\cdot)$ are sufficiently smooth functions to estimate the parameters nonparametrically. Appendix D.2 provides some additional details about this assumption and its implications.[7] This assumption has the appeal that it implies that the betas do not vary too much over short time periods, which is consistent with both economic theory and prior empirical studies (see, for example, Braun, Nelson, and Sunier, 1995; Ghysels, 1998; and Gomes, Kogan, and Zhang, 2003). Importantly, it imposes less structure than assuming a precise functional form for the parameters and so is likely more robust to misspecification. Intuitively, the functional form assumptions in Eq. (33) imply that as T grows, the amount of local information about the function value increases.

A number of different options exist for nonparametrically estimating the $\hat{\beta}_{i,t}$. We follow Ang and Kristensen (2012) and use kernel smoothing estimators. We can then derive, at any point in time, an asymptotic distribution for all parameters of our model, including the conditional betas and the price of risk parameters obtained from the beta estimates. In addition to being more robust to misspecification, kernel smoothing estimators have the appealing feature that they nest, as a special case, rolling window estimates of $\beta_{i,t}$, which are popular in the empirical literature (for example, Chen, Roll, and Ross, 1986; Ferson and Harvey, 1991; Petkova and Zhang, 2005; among many others). Rolling beta estimates are equivalent to using a uniform one-sided kernel instead of a Gaussian two-sided kernel, as we do here. The standard approach of using backward-looking, five-year rolling regressions has two noteworthy drawbacks. First, for the estimator to be consistent, the bandwidth sequence (i.e., the window) needs to shrink to zero. However, the choice of five-year windows is not data-dependent and so might not be appropriate for many applications (see Section 6 for further discussion). Second, the order of the smoothing bias of the estimator for the betas and the price of risk parameters is larger for one-sided kernels. In fact, although the estimator of Λ based on rolling regressions (with appropriate data-dependent bandwidth choice) in Eq. (36) below is consistent, a non-negligible bias term precludes standard inference procedures without further adjustment.

Eq. (29) is nested in a time-varying equivalent of the SUR system discussed in Section 4. We solve the system by equation-by-equation weighted least squares regressions:

$$\left(\hat{A}_{0,i,t-1},\hat{A}'_{1,i,t-1},\hat{\beta}'_{i,t-1}\right)=\left(\sum_{s=1}^{T}\mathcal{K}_h((s-t)/T)z_s^{tv}z_s^{tv\prime}\right)^{-1}\times\left(\sum_{s=1}^{T}\mathcal{K}_h((s-t)/T)z_s^{tv}R_{i,s}\right) \tag{34}$$

and

$$\left(\hat{\mu}_{t-1},\hat{\Phi}_{t-1}\right)'=\left(\sum_{s=1}^{T}\mathcal{K}_b((s-t)/T)\tilde{X}_{s-1}\tilde{X}'_{s-1}\right)^{-1}\times\left(\sum_{s=1}^{T}\mathcal{K}_b((s-t)/T)\tilde{X}_{s-1}X'_s\right), \tag{35}$$

where $z_s^{tv}=(1,X'_{s-1},C'_s)'$ and $\mathcal{K}_h(x)=\mathcal{K}(x/h)$ for some kernel function $\mathcal{K}(\cdot)$ and bandwidths $h=h_T$ and $b=b_T$ are positive sequences that converge to zero. The set of regressors, z_t^{tv}, is different than in the constant beta case in which estimated innovations, $\hat{u}_t$, were used instead of C_t to estimate the betas. When betas are time-varying, it is technically convenient to make this change as we can then directly rely on results from Kristensen (2009).

Intuitively, the kernel function in Eqs. (34) and (35) places more weight on observations nearby and less weight on those farther away, where the rate of decay is governed by the bandwidths h and b, respectively. Moreover, because we smooth only in the time dimension, our approach does not suffer from the so-called curse of dimensionality. To choose the bandwidths we use a plug-in method developed in Kristensen (2012) and Ang and Kristensen (2012). In Appendix A.2, we provide more details on the implementation of the bandwidth selection.

Given these first-stage estimates, the feasible estimator of Λ is then

$$\mathrm{vec}\left(\hat{\Lambda}_{ols}^{tv}\right)=\left(\sum_{t=0}^{T-1}\left(\tilde{F}_t\tilde{F}'_t\otimes\hat{B}'_t\hat{B}_t\right)+\rho_T\right)^{-1}\sum_{t=0}^{T-1}\left(\tilde{F}_t\otimes\hat{B}'_t\right)\times\left(R_{t+1}-\hat{B}_t\hat{u}_{t+1}\right), \tag{36}$$

where ρ_T is a positive sequence that satisfies $\rho_T\to 0$. This additional term guarantees the stability of the estimator by ensuring that the matrix is always invertible. It is straightforward to show that when the betas and VAR coefficients no longer time vary and $\rho_T=0$, then $\hat{\Lambda}_{ols}^{tv}$ is analytically equivalent to $\hat{\Lambda}_{ols}$ from Section 4. We then have Theorem 3.

Theorem 3. Under our assumptions,

$$\sqrt{T}\,\mathrm{vec}\left(\hat{\Lambda}_{ols}^{tv}-\Lambda\right)\xrightarrow{d}\mathcal{N}(0,\mathcal{V}_\Lambda^{tv}),$$

as $T\to\infty$, where

$$\mathcal{V}_\Lambda^{tv}=\left(\int_0^1(\Omega_f(\tau)\otimes B(\tau)'B(\tau))\,d\tau\right)^{-1}\times\left[\int_0^1\left(\left(\Omega_f(\tau)\Lambda'D'_B\Omega_z(\tau)^{-1}D_B\Lambda\Omega_f(\tau)+\Omega_f(\tau)\right)\right.\right.$$

[7] See Robinson (1989). A number of other authors have used this assumption in conjunction with time-varying parameters. See Ang and Kristensen (2012) for a lucid discussion about this approach to modeling time-varying parameters.

$$\otimes B(\tau)'\Sigma_e(\tau)B(\tau)\big)\,\mathrm{d}\tau.$$
$$+\int_0^1 \big(\Omega_f(\tau)\otimes B(\tau)'B(\tau)\Sigma_u(\tau)B(\tau)'B(\tau)\big)\,\mathrm{d}\tau\Big]$$
$$\times\left(\int_0^1 \big(\Omega_f(\tau)\otimes B(\tau)'B(\tau)\big)\,\mathrm{d}\tau\right)^{-1}$$

and $\Omega_z(\cdot)$, $\Omega_f(\cdot)$, $\Sigma_u(\cdot)$, and D_B are defined in Appendix D.2.

Despite the fact that $\hat{\Lambda}^{\text{tv}}_{\text{ols}}$ is based on estimates of $\beta_{i,t}$, μ_t, and Φ_t, which converge at a rate slower than the parametric rate, our estimator of the price of risk achieves the parametric rate. This is an appealing feature as it means that the additional flexibility we introduce in modeling the time variation in the betas and VAR coefficients does not come at the cost of asymptotic efficiency. The intuition behind this result is that the additional averaging over time to estimate Λ accelerates the rate of convergence. Furthermore, in the spirit of the comment in Remark 1, we can then reestimate the $\beta_{i,t}$ from a kernel regression of $R_{i,t+1}$ on the sum $(\hat{\Lambda}^{\text{tv}}_{\text{ols}}\tilde{F}_t+\hat{u}_{t+1})$. Finally, in Appendix B, we discuss how to carry out restricted estimation and inference for the price of risk parameter Λ.

The time variation in μ_t and Φ_t implies that the mean of the factors is also shifting over time. The definition of $\overline{\lambda}$ must be changed accordingly:

$$\overline{\lambda}=\lambda_0+\Lambda_1\cdot\lim_{T\to\infty}T^{-1}\sum_{T}^{t=1}\mathbb{E}[F_t]. \tag{37}$$

We then have Theorem 4, an analogous result to Theorem 2,

Theorem 4. Under our assumptions,

$$\sqrt{T}\left(\hat{\overline{\lambda}}^{\text{tv}}_{\text{ols}}-\overline{\lambda}\right)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{V}^{\text{tv}}_{\overline{\lambda}}\right),$$

where $\mathcal{V}^{\text{tv}}_{\overline{\lambda}}$ is defined in Appendix D.2.

The asymptotic variance, $\mathcal{V}^{\text{tv}}_{\overline{\lambda}}$, can be estimated in a straightforward manner, and so inference on whether a factor is priced on average can be conducted easily.

5.1. Comparison with estimators using rolling regressions

In Section 6, we compare our results with the estimators proposed by Fama and MacBeth (1973) and Ferson and Harvey (1991) (hereafter "FM" and "FH"), both implemented using rolling regressions to obtain time-varying betas in the first estimation stage. To properly account for the persistent nature of (some) factors, we implement these procedures using the estimated innovations $\hat{u}_{t+1}$ as pricing factors. This is in contrast to much of the empirical literature, which estimates betas using the level of the pricing factors without controlling for lagged observations. When the pricing factors are persistent, these estimates are not consistent.[8] Rolling regressions yield estimates $\{\hat{\beta}^{\text{rr}}_{i,t}: i=1,\ldots,N, t=1,\ldots,T\}$, which we stack as $\{\hat{B}^{\text{rr}}_t: t=1,\ldots,T\}$. We then obtain the FM estimator of the constant price of risk parameter λ_0 from

$$\hat{\lambda}^{\text{FM}}_0=T^{-1}\sum_{t=1}^{T}\hat{\gamma}_t,\quad \hat{\gamma}_t=\left(\hat{B}^{\text{rr}\prime}_t\hat{B}^{\text{rr}}_t\right)^{-1}\hat{B}^{\text{rr}\prime}_t\hat{A}^{\text{rr}}_{0,t}, \tag{38}$$

where $\hat{A}^{\text{rr}}_{0,t}$ is the (stacked) estimated intercept from the rolling regressions, the analog to the constant beta case in Eq. (25). In practice, a five-year burn-in period is necessary to construct these estimators. For ease of notation, the equations presented in this subsection ignore this distinction. As in the constant beta case, the estimator in Eq. (38) is derived from the asset pricing restriction that the intercept satisfies $A_{0,t}=B_t\lambda_0$.

Ferson and Harvey (1991) have proposed to estimate time-varying prices of risk in conditional factor pricing models by first running Fama-MacBeth two-pass regressions as above and, subsequently, in a third estimation step, regressing the obtained time series of market prices of risk ($\hat{\gamma}_t$) on one-month lagged predictor variables. This estimator is similar in spirit to our estimator in which market prices of risk are modeled as affine functions of a set of forecasting (X_2 and X_3) variables.

To implement the FH estimator, we again use the innovations $\hat{u}_{t+1}$ as pricing factors but also control for the lagged values of the price of risk factors, F_t [i.e., Eq. (29)] to estimate the betas. We then estimate the price of risk parameters Λ by regressing $\hat{\gamma}_t$ on a constant and the lagged price of risk factors, i.e.,

$$\hat{\Lambda}^{FH}=\left(\sum_{t=0}^{T-1}\hat{\gamma}_{t+1}\tilde{F}'_t\right)\left(\sum_{t=0}^{T-1}\tilde{F}_t\tilde{F}'_t\right)^{-1}. \tag{39}$$

We compare the two estimators with the ones derived in this paper in terms of model-implied mean squared pricing errors in Section 6.

6. Empirical application

In this section, we apply our estimation method to a dynamic asset pricing model for equity and Treasury returns. We choose test assets that have been studied extensively in the empirical asset pricing literature to illustrate the usefulness of the regression-based dynamic asset pricing approach. We show that a parsimonious model with two pricing factors, two price of risk factors, and one factor that is both fits the cross section of size-sorted equity portfolios and constant maturity Treasury portfolios very well on average while, at the same time, giving rise to strongly significant time variation in risk premiums. We further show that allowing for time variation in factor risk exposures substantially improves the precision of price of risk parameters. Finally, allowing for time variation in prices of risk is more important than modeling time variation in factor risk exposures for minimizing squared pricing errors of the model. Traditional estimation approaches such as the one by Fama and MacBeth (1973) and Ferson and Harvey (1991) imply substantially larger pricing errors than our estimator.

[8] While the results they report are based on simple rolling regressions without controlling for the potential persistence in the pricing factors, Ferson and Harvey (1991) mention in Footnote 7 that their results are robust to the estimation of rolling betas controlling for the lagged level of the pricing factors.

6.1. Data

We obtain ten size-sorted portfolios for US equities from Ken French's online data library. We further use constant maturity Treasury portfolios with maturities one, two, five, seven, ten, 20, and 30 years from the Center for Research in Securities Prices (CRSP). We compute excess returns over the one-month Treasury bill yield, which we also obtain from French's website. Our sample spans the period 1964:01–2012:12 for a total of 588 monthly observations.

We use the following set of factors to price the joint cross section of equities and Treasuries. The excess return on the value-weighted equity market portfolio (MKT) from CRSP and the small minus big (SMB) portfolio from Fama and French (1993), and the ten-year Treasury yield (TSY10) serve as cross sectional pricing factors. We obtain the first two factors from French's website, and the third from the Federal Reserve Statistical Release H.15. The first two factors explain a substantial share of the variance of the size decile portfolio returns. However, they are not usually considered to be return forecasting variables. We, therefore, treat them as cross-sectional pricing factors and do not attribute to them a role for explaining time variation in prices of risk. The ten-year Treasury yield can be considered a good proxy for the level of the term structure of Treasury yields, which has been shown to be a priced factor in the cross section of Treasury returns (see e.g., Cochrane and Piazzesi, 2008; Adrian, Crump, and Moench, 2013). We also allow this factor to determine time variation in factor risk premiums, as long-term Treasury yields have been shown to contain predictive information for bond and stock returns (see e.g., Keim and Stambaugh, 1986; Campbell, 1987; Fama and French, 1989; Campbell and Thompson, 2008). In addition to these three factors, we consider two price of risk factors: the term spread between the yield on a ten-year Treasury note and the three-month Treasury bill (TERM) (also obtained from the Federal Reserve Statistical Release H.15), and the log dividend yield (DY) of the Standard & Poor's (S&P) 500 index from Haver Analytics. Both factors have previously been shown to predict equity and bond returns (see e.g., Campbell and Shiller, 1988; Fama and French, 1989; Campbell and Thompson, 2008; Cochrane, 2008) and are, therefore, good proxies for time variation in risk premiums. In summary, in our model excess returns are determined by risk exposures to MKT, SMB, and TSY10, where the market prices of risk of these three pricing factors are assumed to vary over time as affine functions of TSY10, TERM, and DY.

Given this set of test assets and pricing factors, the total number of risk exposure parameters to estimate is $N \times K_C$ or $17 \times 3 = 51$. The number of market price of risk parameters is $K_C \times (K_F + 1)$ or $3 \times 4 = 12$.

6.2. Empirical results

We start by discussing the estimates of factor risk exposures assuming constant betas. Table 1 provides beta estimates for all size and Treasury portfolio returns related to the three risk factors, implied by the estimators in Section 4. The first panel reports the OLS estimates and the second the QML estimates. In each panel, we provide the estimated betas and associated standard errors for the three cross-sectional pricing factors MKT, SMB, and TSY10. Several results are worth highlighting. First, the coefficients and standard errors implied by the OLS and the QML estimator are very similar. Hence, any discussion of estimated risk premiums does not qualitatively depend on the choice of estimator of B. Second, while all size portfolios significantly load on MKT and SMB, the Treasury portfolios do not. That is, Treasury portfolio returns do not contemporaneously co-move with shocks to the two equity pricing factors in the constant beta specification. The market betas of the size portfolios have the expected magnitudes around one with relatively little dispersion. This is the well-known size effect: exposure to MKT does not explain the large spread between average excess returns on small versus large market cap stocks. In contrast, the risk exposures to SMB show a strong differential between the smallest and the largest size deciles. Finally, while the Treasury portfolios do not load on the two equity risk factors, the equity portfolios generally load significantly on the ten-year Treasury yield factor.

We now compare these estimates with those obtained assuming betas are time-varying. Fig. 1 provides plots of factor risk exposures of two test assets, size5 and cmt10, for all three pricing factors in our model: MKT, SMB, and TSY10. For each factor-asset pair we compare three different beta estimates. The constant one (dash-dotted line) is obtained using the estimator in Section 4, the time-varying one (solid line) is obtained using the Gaussian kernel estimator with data-driven bandwidth choice discussed in Section 5, and the five-year rolling window estimator (dashed line) often used in the empirical asset pricing literature and also represents the first-stage estimates in our implementation of the Fama and MacBeth (1973) and Ferson and Harvey (1991) estimators.

Several remarks are in order. First, in all cases, the time-varying beta estimates are centered around the constant estimates. Second, while the Gaussian kernel with data-driven bandwidth implies some variability in factor risk exposures, it features considerably less time variation in betas than the five-year rolling beta estimator. In particular, for all factor-asset pairs, the latter implies betas with signs flipping multiple times across the sample period. At low frequencies, however, the rolling beta estimates mimic the evolution of the Gaussian kernel-based betas. Moreover, despite the smooth nature of Gaussian kernel estimates, their evolution over time gives rise to some interesting observations. Most important, the size5 portfolio's beta on the Treasury factor switches from a negative to a positive sign in the mid-1990s. Around the same time, the cmt10 portfolio's beta on the equity market portfolio switches from a positive to a negative sign. Hence, our time-varying beta estimates replicate the empirical observation that the correlation between stock and bond returns has flipped signs sometime in the 1990s (see e.g. Baele, Bekaert, and Inghelbrecht, 2010; Campbell, Sunderam, and Viceira, 2013; David and Veronesi, 2013). Another interesting observation is that the beta of the ten-year constant maturity Treasury return (cmt10) onto the ten-year Treasury yield factor (TSY10) fluctuates quite substantially over time. Because

Table 1
Factor risk exposure estimates.
This table provides estimates of factor risk exposures from the constant-beta specification of the dynamic asset pricing model discussed in Section 6. It reports ordinary least squares (OLS) estimates and quasi-maximum likelihood (QML) estimates. Asymptotic standard errors are provided in parentheses. The pricing factors are MKT, the excess return on the Center for Research in Security Prices (CRSP) value-weighted equity market portfolio, SMB, the small minus big portfolio both obtained from Ken French's website, and TSY10, the constant maturity ten-year Treasury yield from the Federal Reserve Statistical Release H.15. The test assets are the ten size-sorted stock decile portfolios from Ken French's website (size1 … size10), as well as constant maturity Treasury returns for maturities ranging from one through 30 years (cmt1 … cmt30). We obtain the latter from CRSP. "Wald stats" denote Wald tests for the joint significance of all factor risk exposures associated with the respective pricing factor. "LR stat" is a likelihood ratio test for the joint significance of all factor risk exposures across test assets and pricing factors in the model of Eq. (16) (see Kleibergen and Zhan, 2013). The sample period is 1964:01–2012:12. *** denotes significance at 1%, **, significance at 5%, and *, significance at the 10% level.

	β_{MKT}	$s.e.(\beta_{MKT})$	β_{SMB}	$s.e.(\beta_{SMB})$	β_{TSY10}	$s.e.(\beta_{TSY10})$
OLS estimates						
size1	0.851***	(0.030)	1.171***	(0.020)	0.109***	(0.021)
size2	0.984***	(0.019)	1.066***	(0.018)	0.053***	(0.017)
size3	1.007***	(0.016)	0.898***	(0.015)	−0.111***	(0.014)
size4	0.996***	(0.007)	0.804***	(0.004)	−0.099***	(0.005)
size5	1.018***	(0.007)	0.662***	(0.008)	0.036***	(0.010)
size6	1.003***	(0.017)	0.480***	(0.021)	−0.277***	(0.045)
size7	1.031***	(0.029)	0.362***	(0.039)	−0.160***	(0.038)
size8	1.033***	(0.032)	0.264***	(0.035)	−0.293***	(0.029)
size9	0.995***	(0.028)	0.067***	(0.020)	−0.429***	(0.012)
size10	0.988***	(0.005)	−0.276***	(0.007)	0.188***	(0.009)
cmt1	0.004	(0.011)	0.009	(0.012)	−1.043***	(0.019)
cmt2	0.001	(0.025)	0.008	(0.320)	−2.049***	(0.204)
cmt5	−0.006	(0.201)	−0.003	(0.184)	−4.226***	(0.168)
cmt7	−0.000	(0.149)	−0.018	(0.160)	−5.164***	(0.138)
cmt10	0.007	(0.156)	−0.015	(0.073)	−6.165***	(0.085)
cmt20	0.003	(0.129)	−0.009	(0.111)	−7.893***	(0.117)
cmt30	−0.027	(0.184)	−0.004	(0.269)	−8.581***	(0.320)
Wald stat	182,785.308***	(0.000)	81,346.613***	(0.000)	30,021.782***	(0.000)
LR stat	6,297.355***	(0.000)				
QML estimates						
size1	0.852***	(0.030)	1.174***	(0.020)	0.091***	(0.021)
size2	0.980***	(0.020)	1.060***	(0.018)	0.053***	(0.017)
size3	1.007***	(0.016)	0.898***	(0.015)	−0.082***	(0.014)
size4	0.997***	(0.007)	0.804***	(0.004)	−0.089***	(0.005)
size5	1.019***	(0.007)	0.664***	(0.008)	0.039***	(0.010)
size6	1.005***	(0.017)	0.485***	(0.020)	−0.291***	(0.045)
size7	1.031***	(0.029)	0.363***	(0.038)	−0.185***	(0.037)
size8	1.032***	(0.032)	0.263***	(0.035)	−0.288***	(0.029)
size9	0.994***	(0.027)	0.066***	(0.020)	−0.417***	(0.011)
size10	0.987***	(0.005)	−0.277***	(0.007)	0.186***	(0.009)
cmt1	0.004	(0.010)	0.009	(0.012)	−1.048***	(0.019)
cmt2	0.001	(0.025)	0.008	(0.318)	−2.047***	(0.204)
cmt5	−0.006	(0.201)	−0.003	(0.182)	−4.217***	(0.169)
cmt7	−0.000	(0.147)	−0.018	(0.158)	−5.158***	(0.137)
cmt10	0.007	(0.154)	−0.015	(0.071)	−6.151***	(0.080)
cmt20	0.002	(0.121)	−0.009	(0.106)	−7.916***	(0.114)
cmt30	−0.027	(0.180)	−0.004	(0.268)	−8.579***	(0.319)
Wald stat	184,171.026***	(0.000)	80,463.712***	(0.000)	31,121.825***	(0.000)

the return on a bond is, to a first-order approximation, equal to minus its duration times the yield change, this time variation reflects the fact that the duration of longer-dated Treasury securities has changed substantially over the 50 year sample that we consider. In fact, duration was low in the late 1970s and early 1980s, when rates were high, and has since experienced a secular upward trend against the backdrop of falling rates. These dynamics are well captured by the time-varying beta estimates. Moreover, the five-year rolling regression-based estimates mimic the evolution of time-varying betas from the Gaussian kernel-based estimates with data-driven bandwidth choice well. For other asset-factor pairs, they appear too noisy.

We next turn to a discussion of the estimated market prices of risk. Table 2 provides estimates of the market price of risk parameters λ_0 and Λ_1 implied by three different estimators. The second-to-last column provides the average price of risk estimates $\overline{\lambda}$ for each factor and its asymptotic standard error as provided in Theorems 2 and 4, respectively. These statistics allow us to test whether a given factor is priced on average in the cross section of test assets. Finally, the last column provides a Wald statistic for a test of whether the coefficients in a particular row of Λ_1 are jointly equal to zero. This statistic thus indicates whether there is time variation in each of the factor risk prices.

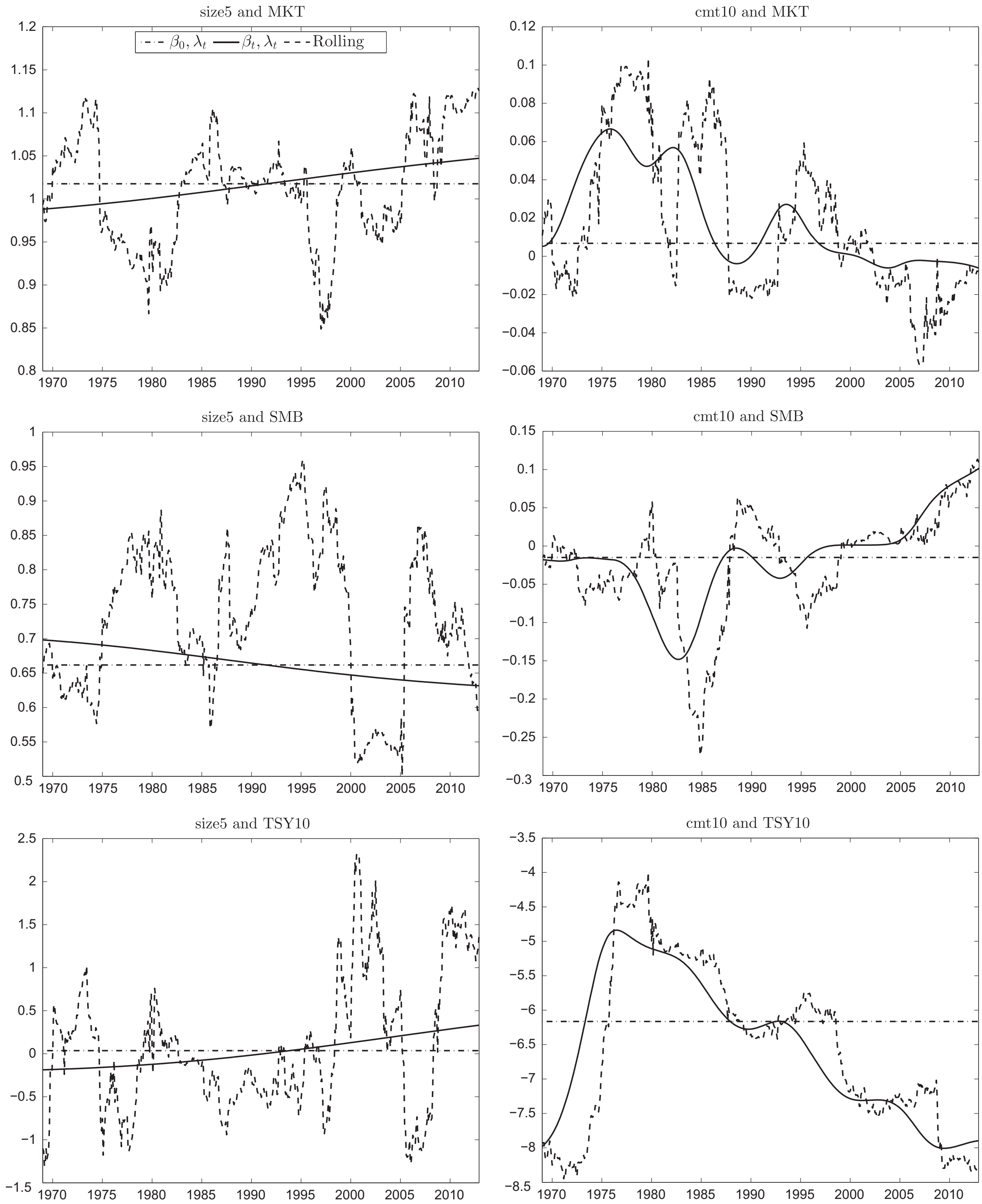

Fig. 1. Comparison of beta estimates. This figure provides plots of beta estimates obtained for different pairs of test assets and cross-sectional pricing factors. β_t, λ_t shows time-varying betas estimated using the kernel regression approach presented in Section 5. β_0, λ_t denotes the constant beta estimate obtained using the ordinary least squares (OLS) estimator described in Section 4. Rolling refers to the five-year rolling window estimate. size5 denotes the fifth decile portfolio from the set of size-sorted stock portfolios from Ken French's website. cmt10 refers to the constant maturity Treasury returns for the ten-year maturity, obtained from the Center for Research in Security Prices (CRSP). MKT, SMB, and TSY10 denote the value-weighted stock market portfolio from CRSP, the small minus big portfolio from Fama and French (1993), and the ten-year Treasury yield from the Federal Reserve Statistical Release H.15. The sample period is 1964:01–2012:12.

Table 2 reports estimates based on time-varying betas implied by a Gaussian kernel and Panels B and C show them for the three-step OLS and QML estimators under constant betas, respectively. The asymptotic standard errors (and p-values in the case of the W_{Λ_1} statistic) are shown in parentheses. We make the following

Table 2
Price of risk estimates.

This table provides estimates of market price of risk parameters from the dynamic asset pricing model discussed in Section 6. Panel A reports ordinary least squares (OLS) estimates for the specification with time-varying betas, and Panels B and C provide OLS and quasi-maximum likelihood (QML) estimates for the specification with constant betas. The pricing factors are MKT, the excess return on the Center for Research in Security Prices (CRSP) value-weighted equity market portfolio, SMB, the small minus big portfolio both obtained from Ken French's website, and TSY10, the constant maturity ten-year Treasury yield from the Federal Reserve Statistical Release H.15. The price of risk factors are TSY10, TERM, the spread between the constant maturity ten-year Treasury yield and the three-month Treasury bill, both obtained from the H.15 release, as well as DY, the log dividend yield obtained from Haver Analytics. The first column, λ_0, gives the estimated constant in the affine price of risk specification for each pricing factor. The second through fourth columns provide the estimated coefficients in the matrix Λ_1, which determine loadings of prices of risk on the price of risk factors. The column $\overline{\lambda}$ provides an estimate of the average price of risk as given in Eq. (22). The last column provides the Wald test statistic of the null hypothesis that the associated row is all zeros. The sample period is 1964:01–2012:12. *** denotes significance at 1%, **, significance at 5%, and *, significance at the 10% level.

	λ_0	TSY10	TERM	DY	$\overline{\lambda}$	W_{Λ_1}
Panel A: Time-varying betas						
MKT	0.062***	−0.184***	0.302***	0.014***	6.797**	25.328***
	(0.017)	(0.058)	(0.088)	(0.004)	(2.785)	(0.000)
SMB	0.054***	−0.194***	0.099	0.011***	3.565	23.190***
	(0.013)	(0.044)	(0.066)	(0.003)	(2.690)	(0.000)
TSY10	0.004***	−0.014***	−0.046***	0.001**	−0.359	48.636***
	(0.001)	(0.005)	(0.007)	(0.000)	(0.229)	(0.000)
Panel B: Constant betas (OLS)						
MKT	0.063**	−0.187*	0.301**	0.014**	6.067***	8.975**
	(0.028)	(0.098)	(0.147)	(0.006)	(1.487)	(0.030)
SMB	0.054**	−0.192***	0.093	0.011**	3.023	7.599*
	(0.022)	(0.073)	(0.108)	(0.005)	(2.336)	(0.055)
TSY10	0.004	−0.013	−0.050***	0.001	−0.386***	21.237***
	(0.002)	(0.008)	(0.012)	(0.001)	(0.085)	(0.000)
Panel C: Constant betas (QMLE)						
MKT	0.063**	−0.187*	0.301**	0.014**	6.066***	8.975**
	(0.028)	(0.098)	(0.147)	(0.006)	(1.424)	(0.030)
SMB	0.054**	−0.192***	0.093	0.011**	3.025	7.598*
	(0.022)	(0.073)	(0.108)	(0.005)	(2.037)	(0.055)
TSY10	0.004	−0.013	−0.050***	0.001	−0.386***	21.237***
	(0.002)	(0.008)	(0.012)	(0.001)	(0.103)	(0.000)

observations. First, the estimated market price of risk parameters are strikingly similar across the three estimators. This reinforces the above observation that the Gaussian kernel-based beta estimators with data-driven bandwidth choice do not move sharply over time. Second, the price of risk parameters are estimated with much greater precision in the time-varying beta case, as the standard errors of most elements of λ_0 and Λ_1 are substantially smaller in Panel A. Hence, because the price of risk parameters are identified based on cross-sectional variation of the betas, allowing for time-varying risk exposures more precisely captures the dynamics of the price of risk. Third, while the constant coefficients in the market prices of risk are all individually significant at the 1% level across the three estimators, the average price of risk statistic $\overline{\lambda}$ discussed in Theorems 2 and 4 is statistically different from zero only for MKT in the time-varying beta case and for MKT and TSY10 in the constant beta case. Hence, according to all three estimators exposure to SMB risk is not unconditionally priced in our cross section of test assets. This is consistent with other studies showing that SMB is not priced in the cross section of size- and book-to-market-sorted equity portfolios (see, for example, Lettau and Ludvigson, 2001). However, while the price of SMB risk is statistically not different from zero on average, it exhibits substantial time variation and fluctuates between positive and negative values that are significantly different from zero. This is also indicated by the Wald statistic W_{Λ_1} for the rows of Λ_1 being jointly equal to zero, provided in the last column. All three estimators suggest significant time variation in the prices of risk on all three cross-sectional pricing factors of our model, including SMB.

Looking at individual elements of Λ_1, we find strong evidence for time variation in the prices of risk of MKT, SMB, and TSY10 as all but one element of the coefficient matrix Λ_1 are individually significant at least at the 10% level. In particular, TSY10 affects the prices of risk of all three factors with a negative sign. That is, higher long-term interest rates drive down the price of risk for both equity and bond market factors. Third, while TERM does not significantly add to the variation in the price of SMB risk, a high term spread strongly raises the price of MKT risk and reduces the price of TSY10 risk. Because equity portfolios load positively on MKT, this implies that a positive term spread predicts higher expected excess returns on stocks, in line with, e.g., Campbell (1987) and Fama and French (1989). Moreover, noting that the factor risk exposures of bond returns on TSY10 are negative, the latter finding is consistent with the evidence in, e.g.,

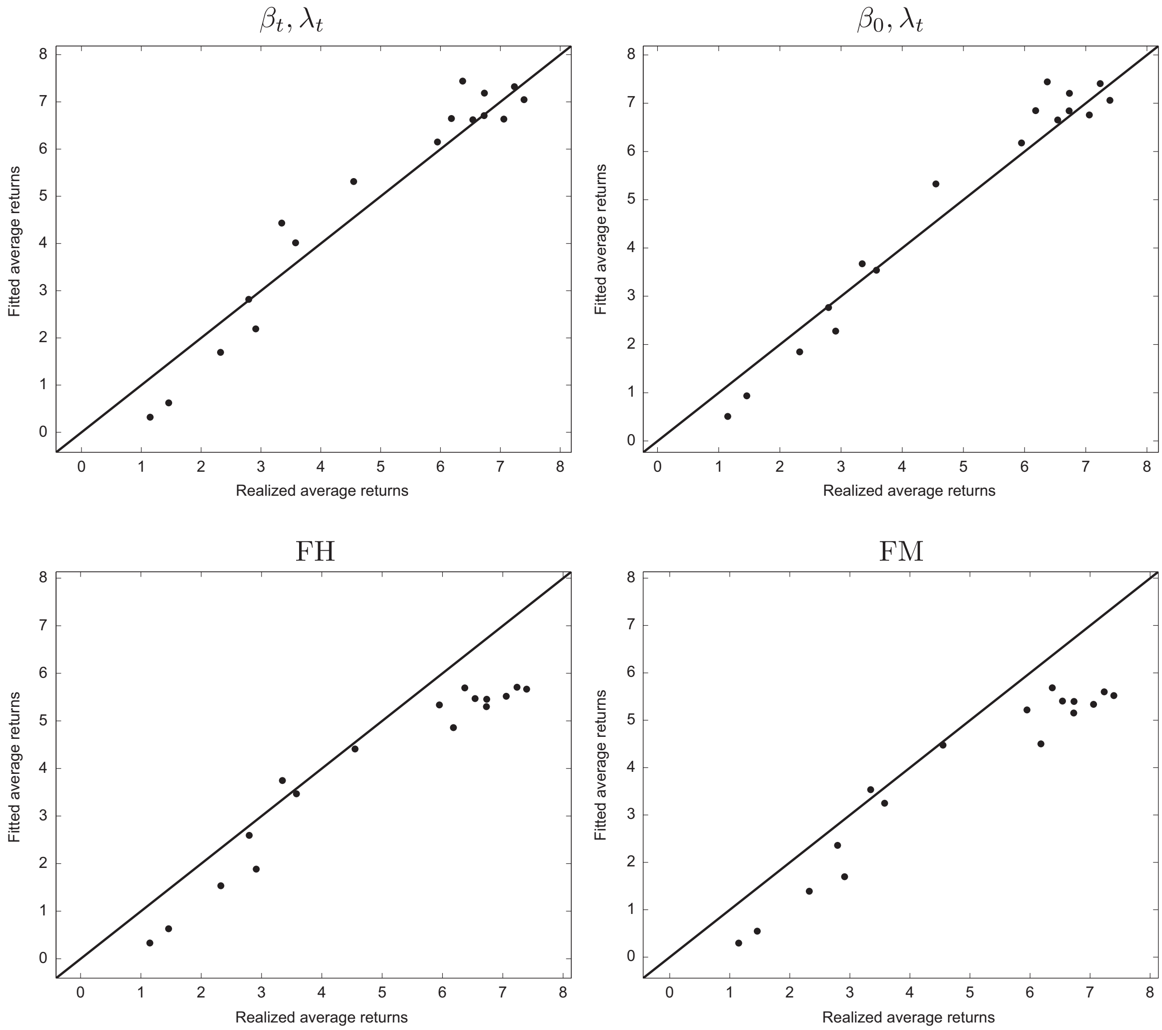

Fig. 2. Comparison of cross-sectional pricing properties. This figure provides plots of observed versus model-implied average excess returns on the set of test assets estimated using four different approaches as discussed in Section 6. The upper-left graph reports results based on our benchmark specification (β_t, λ_t) with time-varying betas and time-varying prices of risk, estimated using the approach presented in Section 5. The upper-right graph shows the unconditional fit of the specification with constant betas but time-varying prices of risk, estimated using the three-stage ordinary least squares (OLS) estimator discussed in Section 4. The lower-left graph shows the average fit of the model estimated using the approach suggested in Ferson and Harvey (1991), designated FH, which is based on time-varying betas estimated using five-year rolling window regressions. The lower-right graph presents results for the Fama and MacBeth (1973), designated, FM, two-pass estimator that is also based on time-varying betas estimated using five-year rolling window regressions but features constant prices of risk. We implement FM by treating the ten-year Treasury yield as a $X1$-type pricing factor and omitting the dividend yield and the term spread as factors. All excess returns are stated in annualized percentage terms. The test assets are the ten size-sorted stock decile portfolios from Ken French's website (size1 … size10), as well as constant maturity Treasury returns for maturities ranging from one through 30 years (cmt1 … cmt30), obtained from the Center for Research in Security Prices (CRSP). The plots are based on the OLS estimates of the model. The sample period is 1964:01–2012:12.

Campbell and Shiller (1991), that a positive slope of the yield curve predicts higher future Treasury returns. Finally, the log dividend yield DY has a positive impact on the prices of risk of all three factors. This confirms previous evidence, e.g. in Fama and French (1989), that the dividend yield predicts excess returns on stocks and bonds.

Before diving into a more specific analysis of time variation in risk premiums, we show the good performance of our dynamic asset pricing model in explaining average excess returns on size and Treasury portfolios. Fig. 2 shows average model-implied excess returns against average observed excess returns, as implied by four different model specifications and corresponding estimators. The upper-left graph shows the average model fit for the specification with both betas and market prices of risk time-varying, estimated with the Gaussian kernel-based estimator discussed in Section 5. The upper-right graph displays the model fit for the specification with constant betas, estimated using the three-step OLS regression approach outlined in Section 4. The lower two graphs show average pricing errors implied by the Ferson and Harvey (1991) and Fama and MacBeth (1973) estimation approaches. While the former features time-varying and the latter constant prices of risk, both are based on betas

Table 3
Mean squared pricing error comparison.

This table compares mean squared pricing errors across various model estimation approaches for the asset pricing model discussed in Section 6. Panel A reports, for each test asset, the mean squared pricing error implied by the various estimation approaches. β_t, λ_t denotes our benchmark specification with both time-varying betas and market prices of risk and the betas being estimated using the approach discussed in Section 5. β_0, λ_t is a specification with constant betas but time-varying prices of risk estimated using the ordinary least squares (OLS) estimator discussed in Section 4. Columns 3 (β_t, λ_0) and 4 (β_0, λ_0) denote specifications with time-varying and constant risk exposures, respectively, and constant prices of risk. "FH" refers to the Ferson and Harvey (1991) estimator discussed in Section 5, which is based on time-varying betas estimated using five-year rolling window regressions. "FM" denotes the Fama and MacBeth (1973) two-pass estimator also based on time-varying betas estimated using five-year rolling window regressions. Mean squared pricing errors are stated in percentage terms. Panel B shows the mean squared pricing errors of all model specifications relative to the benchmark estimation. The test assets are the ten size-sorted stock decile portfolios from Ken French's website (size1 … size10), as well as constant maturity Treasury returns for maturities ranging from one through 30 years (cmt1 … cmt30), obtained from the Center for Research in Security Prices (CRSP). The sample period is 1964:01–2012:12.

	β_t, λ_t	β_0, λ_t	β_t, λ_0	β_0, λ_0	FH	FM
Panel A: Mean squared pricing errors						
size1	5.87	6.13	7.07	7.06	6.35	6.34
size2	2.77	2.80	3.49	3.49	3.25	3.31
size3	1.96	2.00	2.80	2.80	2.45	2.53
size4	1.90	1.92	2.75	2.75	2.37	2.49
size5	1.69	1.72	2.49	2.49	2.16	2.26
size6	1.78	1.88	2.67	2.67	2.38	2.38
size7	1.74	1.78	2.32	2.32	2.08	2.12
size8	1.51	1.52	1.96	1.96	1.79	1.81
size9	1.27	1.27	1.60	1.60	1.44	1.47
size10	0.33	0.33	0.58	0.58	0.48	0.54
cmt1	0.08	0.10	0.10	0.11	0.08	0.09
cmt2	0.17	0.21	0.22	0.23	0.18	0.19
cmt5	0.35	0.38	0.44	0.44	0.38	0.40
cmt7	0.44	0.49	0.58	0.57	0.46	0.50
cmt10	0.43	0.61	0.71	0.74	0.55	0.58
cmt20	1.24	1.72	1.85	2.03	1.48	1.52
cmt30	1.80	2.82	2.85	3.14	2.08	2.08
Average	1.49	1.63	2.03	2.06	1.76	1.80
Panel B: Mean squared pricing errors relative to β_t, λ_t						
size1		1.04	1.20	1.20	1.08	1.08
size2		1.01	1.26	1.26	1.17	1.20
size3		1.02	1.43	1.43	1.25	1.29
size4		1.01	1.45	1.45	1.24	1.31
size5		1.02	1.47	1.47	1.28	1.34
size6		1.06	1.49	1.50	1.33	1.33
size7		1.02	1.33	1.33	1.20	1.22
size8		1.01	1.29	1.30	1.19	1.19
size9		1.00	1.26	1.26	1.13	1.15
size10		1.00	1.73	1.74	1.45	1.62
cmt1		1.27	1.26	1.32	1.03	1.05
cmt2		1.27	1.28	1.35	1.08	1.11
cmt5		1.09	1.26	1.26	1.09	1.15
cmt7		1.10	1.31	1.30	1.05	1.13
cmt10		1.43	1.65	1.73	1.30	1.36
cmt20		1.39	1.50	1.64	1.20	1.23
cmt30		1.56	1.58	1.74	1.15	1.15
Average		1.14	1.40	1.43	1.19	1.23

estimated via five-year rolling regressions. The graphs show that our joint dynamic asset pricing model fits the cross section of average excess returns very well, in both the constant and the time-varying beta specification. In contrast, both the Ferson and Harvey (1991) and Fama and MacBeth (1973) estimators imply average fitted excess returns for the equity portfolios in our cross section that are all lower than the observed average excess returns.

The model's ability to fit returns should be assessed not only on average but also at each point in time. Panel A of Table 3 reports mean squared pricing errors for our model as implied by the different specifications and estimation approaches. Specifically, for each test asset i, we report the quantity[9]

$$MSE_i = \frac{1}{T}\sum_{t=0}^{T-1} \hat{e}_{i,t+1}^2. \tag{40}$$

The first column (β_t, λ_t) shows our benchmark specification with both time-varying betas and market prices of risk and the betas being estimated using the approach discussed in Section 5. The second column (β_0, λ_t) is a specification with constant betas but time-varying prices of risk estimated using the OLS estimator discussed in Section 4. Columns 3 (β_t, λ_0) and 4 (β_0, λ_0) denote specifications with time-varying and constant risk exposures, respectively, and constant prices of risk. The fifth column ("FH") provides the Ferson and Harvey (1991) estimator discussed in Section 5, which is based on time-varying betas estimated using five-year rolling window regressions. Finally, the last column ("FM") shows the Fama and MacBeth (1973) two-pass estimator based also on time-varying betas estimated using five-year rolling window regressions. All mean squared pricing errors are stated in percent.

The main result of the table is that none of the alternative estimation approaches generates mean squared pricing errors that are smaller than those implied by the benchmark (β_t, λ_t) specification for any of the test assets. In particular, the specifications with constant prices of risk imply substantially larger pricing errors. The FH estimator, which features time-varying prices of risk but betas estimated using five-year rolling window regressions, also produces pricing errors that substantially exceed those implied by our benchmark estimator. The relative performance of the various estimation approaches can best be seen from mean squared error (MSE) ratios with respect to our benchmark estimation specification (β_t, λ_t), provided in Panel B of Table 3. These ratios show that the benchmark specification outperforms the specification with time-varying prices of risk but constant betas, substantially for the Treasury portfolios but, at most by a few percentage points for the size-sorted equity portfolios. This implies that in our model allowing for time variation in betas is relatively more important for Treasury returns. In contrast, allowing for time variation in prices of risk dramatically reduces the mean squared pricing error, as evidenced by the fact that both specifications with constant prices of risk $(\beta_t, \lambda_0$ and $\beta_0, \lambda_0)$ imply MSEs that

[9] To ensure a fair comparison across estimators, we report mean squared errors taken over the same sample period, thus taking into account the trimming of data in the time-varying beta case.

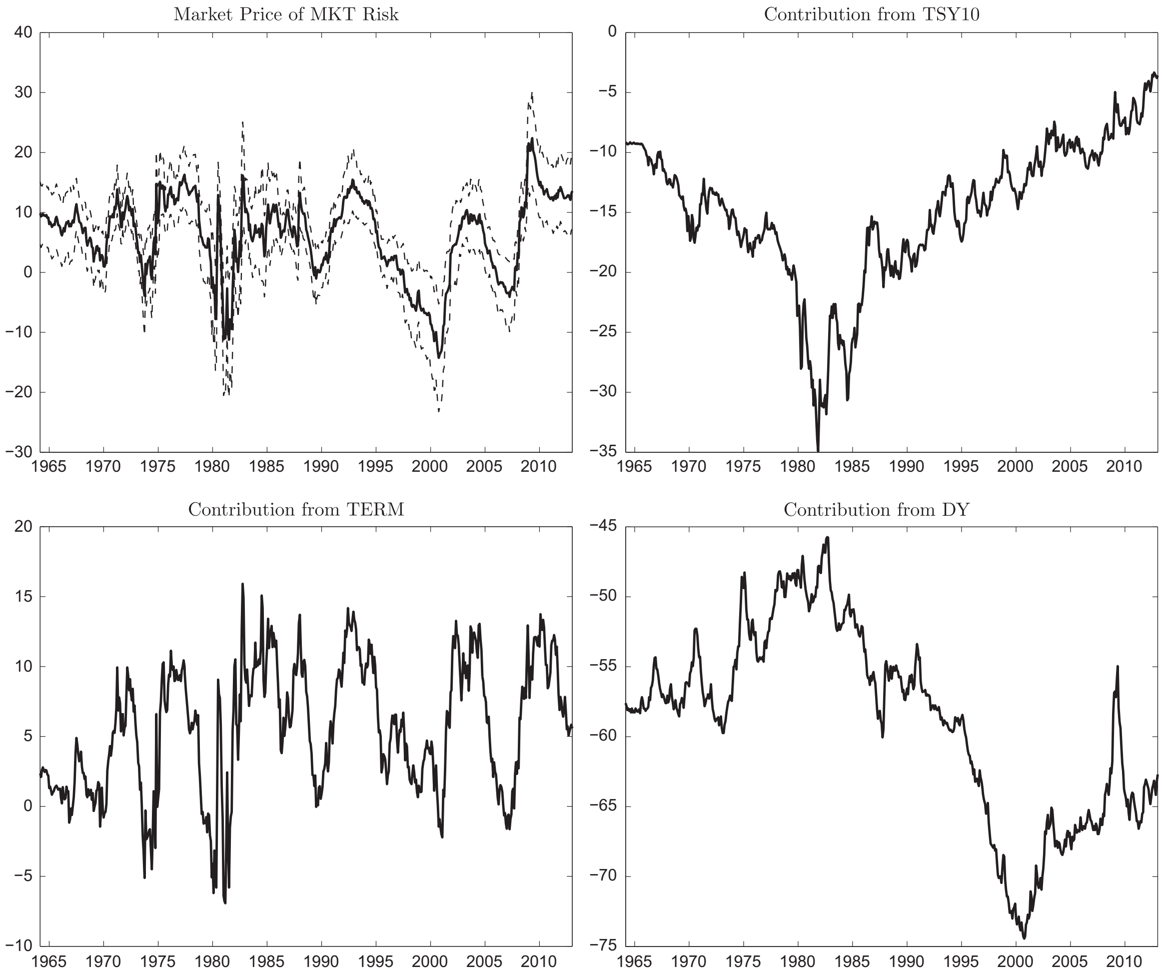

Fig. 3. Price of MKT risk dynamics. This figure provides plots of the estimated time series of the price of MKT risk implied by the dynamic asset pricing model with time-varying betas and prices of risk estimated using the approach in Section 5 and discussed in Section 6. The upper-left graph plots the price of market risk along with its conditional 95% confidence interval. The remaining graphs provide the contributions of the three price of risk factors TSY10, TERM, and DY to the dynamics of the price of market risk. All quantities are stated in annualized percentage terms. The sample period is 1964:01–2012:12.

exceed the benchmark specification between 20% and 74%. Turning to the last two columns, when betas are estimated using five-year rolling regressions, allowing for prices of risk to vary over time, as in the Ferson-Harvey estimator, improves the model fit with respect to the Fama-MacBeth estimator, but the difference is substantially smaller than when betas are estimated using Gaussian kernels. More important, both the FH and FM estimators imply an average MSE of 19% and 23% larger than that of our benchmark specification. Hence, estimators using rolling five-year window regressions perform substantially less well than our estimator using time-varying betas obtained from Gaussian kernel regressions.[10]

In sum, our results show that the time variation of excess returns on stock and bond portfolios is mainly driven by time-varying prices of risk and, to a much smaller extent, by changes in the factor risk betas. This finding is consistent with the results of Ferson and Harvey (1991) and highlights the importance of using a dynamic framework and an estimation approach consistent with such a framework when testing asset pricing models.

We now turn to a characterization of the dynamics of the price of risk. Fig. 3 provides a plot of the estimated price of MKT risk implied by the model, as given by our benchmark estimation approach with time-varying betas and time-varying prices of risk. The upper-left graph shows the time series evolution of the estimated price of risk along with its conditional 95% confidence interval. The plot shows that the price of MKT risk is strongly time-varying. While it has on average amounted to about 6% over the past 50 years, there have been a few episodes in which the estimated price of market risk has been markedly negative. In particular, during the final two years of the dotcom bubble as well as in the two years before the

[10] For comparison, we consider estimators of the price of risk parameters based on estimated betas using the level of price of risk factors instead of innovations. However, not surprisingly, the results are very poor and so we omit them from the presented results.

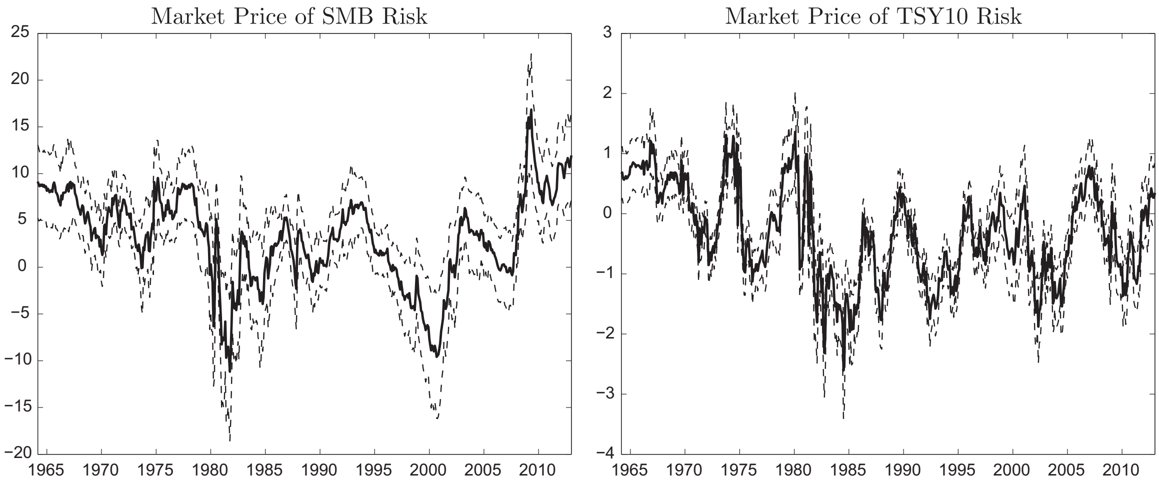

Fig. 4. Time variation in the price of SMB and TSY10 risk. This figure provides plots of the estimated time series of the price of SMB and TSY10 risk implied by the dynamic asset pricing model estimated using the method outlined in Section 5 and discussed in Section 6. The left panel plots the price of SMB risk along with its conditional 95% confidence interval, and the right panel reports the price of TSY10 risk along with its conditional 95% confidence interval. All quantities are stated in annualized percentage terms. The sample period is 1964:01–2012:12.

recent financial crisis, the estimated market risk premiums fell below zero, indicating that, according to our model, equity investors would have anticipated negative excess returns on equity in these periods.

The remaining graphs in Fig. 3 show the contribution of the three price of risk factors to these dynamics. Recall that, in our model, $\lambda_t = \lambda_0 + \Lambda_1 F_t$, where F_t is the vector of price of risk factors. Accordingly, the three graphs show the quantities $\lambda_{1j} F_{jt}$, where λ_{1j} is the $(1, j)$ element of Λ_1 and F_{jt} is the jth factor in F_t. These graphs thus allow one to attribute the dynamics of the price of market risk to its various components. As an example, our model implies that the equity risk premium was at an all-time high in the spring of 2009. Looking at the individual contributions of the three price of risk factors, this period was characterized by a combination of a very low ten-year Treasury yield, a relatively high term spread, and a fairly elevated dividend yield.

Fig. 4 shows the estimated time series of annualized prices of risk for the SMB and TSY10 factors along with their conditional 95% confidence intervals. Both series exhibit substantial time variation. The price of SMB risk largely mimics the dynamics of the price of MKT risk, but it has a somewhat lower average level. As shown in Table 2, the average price of SMB risk is not significantly different from zero in our sample. However, as shown by its conditional 95% confidence interval, the price of SMB risk has been significantly different from zero over various subperiods in our sample. Turning to the evolution of the market price of TSY10 risk, shown in the right graphs of Fig. 4, exposure to long-term Treasury risk was associated with a positive price of risk for much of the period from the beginning of the sample in 1963 through the early 1980s. However, around the time of the Volcker disinflation period, the price of TSY10 risk switched sign and has since fluctuated around mostly negative values. The exposure of equity portfolios to the Treasury factor switched signs from negative to positive sometime in the mid-1990s. Combined, these results imply that exposure to long-term Treasury risk generated strongly fluctuating risk prices for stock portfolios over the last 50 years. While the price of risk was mostly negative in the early part of the sample, it flipped sign in the early 1980s and became negative again around the mid-1990s.

An important aspect of our modeling framework is that we can use the dynamics of the pricing factors to predict expected excess returns further out than one month ahead. This is useful as it facilitates a quantitative analysis of risk premiums at longer-term investment horizons. Fig. 5 shows the model-implied expected excess return on the fifth size portfolio and the ten-year constant maturity Treasury portfolio one year and five years into the future, as implied by our benchmark specification with time-varying betas and prices of risk. The charts indicate that the model-implied risk premiums feature sizable time variation. For the fifth size portfolio, they varied in a range from minus 15% to 30% at a one-year-ahead horizon and between 2% and 12% at a five-year-ahead horizon. For the ten-year Treasury portfolio, the time variation of risk premiums is in a narrower range of around minus four to slightly over 10% at the one-year horizon and between slightly below zero and around 5% at the five-year horizon. Hence, our model and estimation approach predict meaningful variation of longer-term risk premiums, consistent with the persistence of actual excess returns over long horizons. For comparison, we superimpose the corresponding long-horizon risk premiums implied by the specification with time-varying betas but constant prices of risk. Not surprisingly, this specification implies only minor time variation in risk premiums.

7. Conclusion

Dynamic asset pricing models constitute the core of modern finance theory. Virtually all of the macro-finance literature of recent decades is cast in dynamic terms, often giving rise to time-varying risk prices. Empirically,

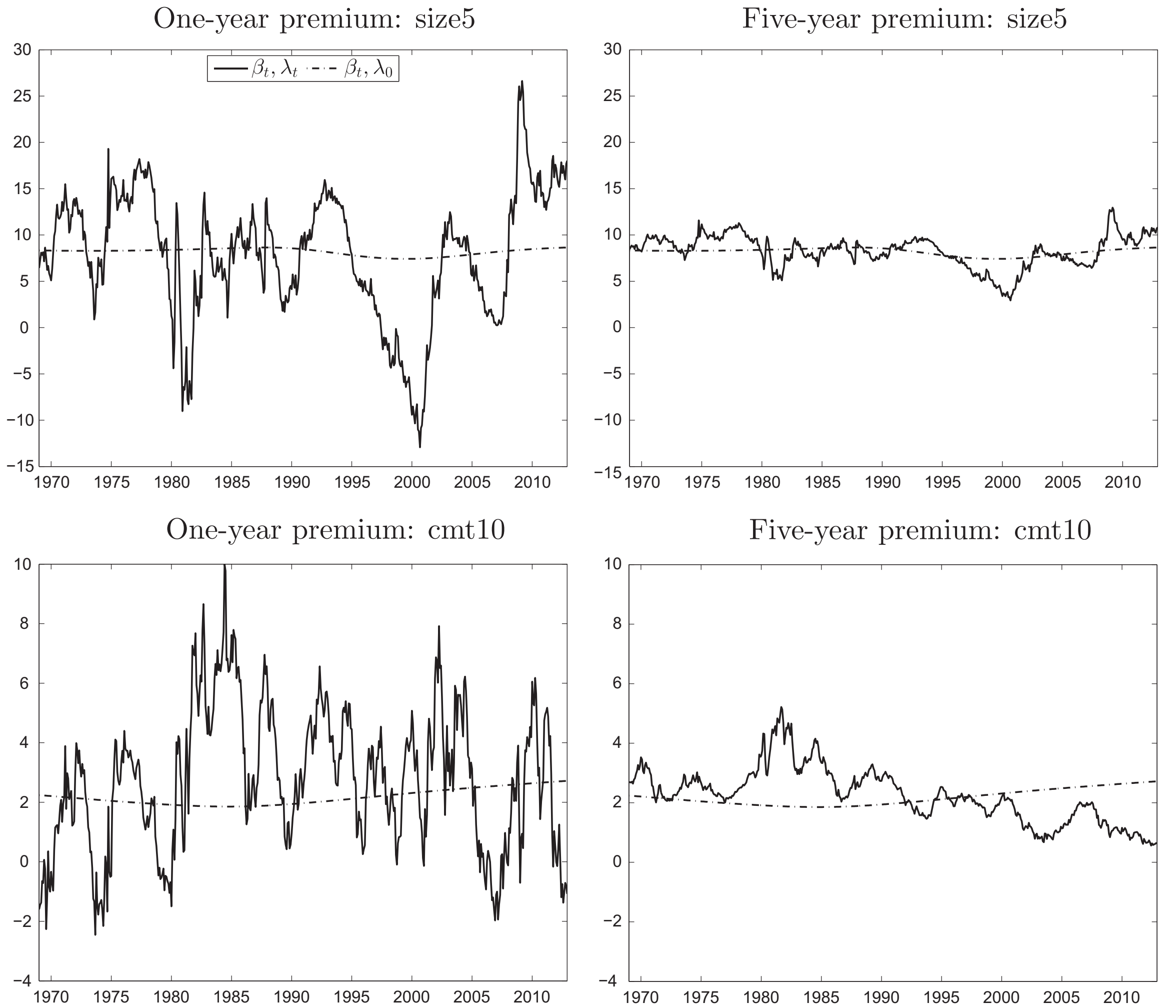

Fig. 5. One-year and five-year risk premium dynamics. This figure provides plots of the estimated one- and five-year ahead expected excess returns for two test assets implied by the dynamic asset pricing model with time-varying betas and prices of risk estimated using the approach in Section 5 and discussed in Section 6. size5 denotes the fifth decile portfolio from the set of size-sorted stock portfolios from Ken French's website. cmt10 refers to the constant maturity Treasury returns for the ten year maturity, obtained from the Center for Research in Security Prices (CRSP). All quantities are stated in annualized percentage terms. The sample period is 1964:01–2012:12.

the time variation in prices of risk has been shown robustly (see, e.g., Campbell and Shiller, 1988; Cochrane, 2011).

In this paper, we provide a unifying framework for estimating beta representations of generic dynamic asset pricing models that impose cross-sectional no-arbitrage restrictions and allow for betas to vary smoothly over time and for prices of risk to vary with observable state variables. We allow for state variables that are cross-sectional pricing factors or forecasting variables for the price of risk, or both. Our estimation results show that all three types of variables are empirically relevant.

Our regression-based estimation approach can be explained as a three-step estimator. First, shocks to the state variables are obtained from a time series vector autoregression. Second, asset returns are regressed on lagged state variables and their contemporaneous innovations, generating predictive slopes and risk betas for each test asset. In the third step, prices of risk are obtained by either regressing the predictive slopes on the betas cross-sectionally or by an eigenvalue decomposition of the predictive slopes and betas. The three-step regression estimator coincides with the estimator of Fama and MacBeth (1973) when state variables are uncorrelated across time and prices of risk are constant. Our approach thus nests the popular Fama-MacBeth two-pass estimator.

All of the estimators presented in this paper are either directly or indirectly based on standard regression outputs. As a result, our estimation approach is computationally efficient and robust. We provide an application to the joint pricing of stocks and bonds, which features very good cross-sectional pricing properties with small average pricing errors as well as strongly significant time variation of risk premiums. We find that the time variation in risk prices is more important than the time variation in betas for achieving good model fits.

Appendix A. Implementing the estimators

A.1. Constant betas

More concretely $\hat{\lambda}_{0,\text{ols}}$, $\hat{\lambda}_{0,\text{qmle}}$, $\hat{\Lambda}_{1,\text{ols}}$, and $\hat{\Lambda}_{1,\text{qmle}}$ can be obtained by the following three steps.

1. Estimate the joint VAR in Eq. (1) via $\hat{V} = X - \hat{\Psi}_{\text{ols}}\tilde{X}_-$ where $\hat{\Psi}_{\text{ols}} = X\tilde{X}'_-\left(\tilde{X}_-\tilde{X}'_-\right)^{-1}$ and $\tilde{X}_- = [\iota_T \,\vert\, X'_-]'$. Form $\hat{U}$ as the $K_C \times T$ matrix extracted from the first K_C rows of $\hat{V}$. Finally, construct the estimators $\hat{\Sigma}_u = \hat{U}\hat{U}'/T$ and $\hat{\Upsilon}_{FF} = \tilde{F}_-\tilde{F}'_-/T$.
2. Estimate $\hat{A}_{\text{ols}} = R\hat{Z}'\left(\hat{Z}\hat{Z}'\right)^{-1}$ and then form the heteroskedasticity robust standard errors

$$\hat{\mathcal{V}}_{\text{rob}} = T \cdot \left(\left(\hat{Z}\hat{Z}'\right)^{-1} \otimes I_N\right)\left(\sum_{t=1}^{T}(\hat{z}_t\hat{z}'_t \otimes \hat{e}_t\hat{e}'_t)\right)\left(\left(\hat{Z}\hat{Z}'\right)^{-1} \otimes I_N\right),$$

where $\hat{z}_t = (1, F'_{t-1}, \hat{u}'_t)'$ and $\hat{e}_t = R_t - \hat{A}_{\text{ols}}\hat{z}_t$.
3. Estimate

$$\hat{\lambda}_{0,\text{ols}} = \left(\hat{B}'_{\text{ols}}\hat{B}_{\text{ols}}\right)^{-1}\hat{B}'_{\text{ols}}\hat{A}_{0,\text{ols}}, \quad \hat{\Lambda}_{1,\text{ols}} = \left(\hat{B}'_{\text{ols}}\hat{B}_{\text{ols}}\right)^{-1}\hat{B}'_{\text{ols}}\hat{A}_{1,\text{ols}}.$$

Next, let $L = [\zeta_1 \cdots \zeta_{K_C}]$, where ζ_i is the eigenvector associated with the ith largest eigenvalue of the matrix $\hat{A}_{\text{ols}}\hat{Z}\hat{Z}'\hat{A}'_{\text{ols}}$. Then let

$$\hat{B}_{\text{qmle},0} = L, \quad \hat{D}_{\text{qmle},0} = L'\hat{A}_{\text{ols}}.$$

Define $\hat{\Delta}_{\text{qmle},0}$ as the last K_C columns of the matrix $\hat{D}_{\text{qmle},0}$. Then,

$$\hat{B}_{\text{qmle}} = \hat{B}_{\text{qmle},0}\hat{\Delta}_{\text{qmle},0}, \quad \hat{D}_{\text{qmle}} = \hat{\Delta}^{-1}_{\text{qmle},0}\hat{D}_{\text{qmle},0},$$

and $\hat{\Lambda}_{\text{qmle}}$ is the matrix formed from the first $K_F + 1$ columns of $\hat{D}_{\text{qmle}}$. Finally, construct the variance estimators

$$\hat{\mathcal{V}}_{\Lambda,\text{ols}} = \left(\hat{\Upsilon}^{-1}_{FF} \otimes \hat{\Sigma}_u\right) + \mathcal{H}_\Lambda\left(\hat{B}_{\text{ols}}, \hat{\Lambda}_{\text{ols}}\right)\hat{\mathcal{V}}_{\text{rob}}\mathcal{H}_\Lambda\left(\hat{B}_{\text{ols}}, \hat{\Lambda}_{\text{ols}}\right)'$$

and

$$\hat{\mathcal{V}}_{\Lambda,\text{qmle}} = \left(\hat{\Upsilon}^{-1}_{FF} \otimes \hat{\Sigma}_u\right) + \mathcal{H}_\Lambda\left(\hat{B}_{\text{qmle}}, \hat{\Lambda}_{\text{qmle}}\right)\hat{\mathcal{V}}_{\text{rob}}\mathcal{H}_\Lambda\left(\hat{B}_{\text{qmle}}, \hat{\Lambda}_{\text{qmle}}\right)'.$$

A.2. Time-varying betas

$\hat{\Lambda}^{\text{tv}}$ can be obtained in three steps. The implementation requires choices of the trimming parameter ρ_T. In our empirical application, we choose $\rho_T = 10^{-6}$. In addition, to avoid boundary bias issues we drop the first and last 12 monthly observations in our empirical application, following Ang and Kristensen (2012).

1. Estimate the time-varying joint VAR in Eq. (30). Assume Ψ_t follows a polynomial of order P in t, i.e., regress $X_{i,t+1}$ on $(\pi(t) \otimes X_{t-1})$, where $\pi(t) = (1, t, \ldots, t^P)$ for $i = 1, \ldots, K$. Combine these coefficient estimates to form $\hat{\Psi}^0_t$. In our application, we choose P=6 following Ang and Kristensen (2012). Next, follow the steps in Ang and Kristensen (2009) and Kristensen (2012) to obtain the short-run and long-run bandwidth choices b_i^{sr} and b_i^{lr} for $i = 1, \ldots, K$. Then construct the estimator of the ith row of Ψ_t via

$$\left[\hat{\Psi}_{t-1}\right]_{i,\cdot} = \sum_{s=1}^{T}\mathcal{K}_b\left(\frac{s-t}{T}\right)X_{i,s}\tilde{X}'_{s-1}\left(\sum_{s=1}^{T}\mathcal{K}_b\left(\frac{s-t}{T}\right)\tilde{X}_{s-1}\tilde{X}'_{s-1}\right)^{-1},$$

where $b \in \left\{b_i^{\text{sr}}, b_i^{\text{lr}}\right\}$, $X_{i,s}$ is the ith element of X_s and $\tilde{X}_{s-1} = (1, X'_{s-1})'$. Here, $\mathcal{K}(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. Then form $\hat{v}_t$ by $\hat{v}_t = X_t - \hat{\Psi}_{t-1}\tilde{X}_{t-1}$. Finally, construct

$$\hat{\Omega}_{x,t} = T^{-1}\sum_{s=1}^{T}\mathcal{K}_b\left(\frac{s-t}{T}\right)\tilde{X}_{s-1}\tilde{X}'_{s-1}, \quad \hat{\Sigma}_{v,t} = T^{-1}\sum_{s=1}^{T}\mathcal{K}_b\left(\frac{s-t}{T}\right)\hat{v}_s\hat{v}'_s,$$

where $b = b_c$ is a common bandwidth choice. In our application, we use the average bandwidth chosen across the K equations.

2. Estimate the time-varying reduced-form return generating Eq. (29). Assume A_t follows a polynomial of order P in t, i.e., regress $R_{i,t+1}$ on $(\pi(t)\otimes z_t^{\text{tv}})$, where $\pi(t)=(1,t,\dots,t^P)$ for $i=1,\dots,N$. Combine these coefficient estimates to form $\hat{A}_{i,t}^0$. In our application we choose P=6 following Ang and Kristensen (2012). Next, follow the steps in Ang and Kristensen (2009) and Kristensen (2012) to obtain the short-run and long-run bandwidth choices h_i^{sr} and h_i^{lr} for $i=1,\dots,N$. Then, construct the estimator of $A_{i,t}$ via
$$\hat{A}_{i,t-1}=\left(\sum_{s=1}^{T}\mathcal{K}_h\left(\frac{s-t}{T}\right)z_s^{\text{tv}}z_s^{\text{tv}\prime}\right)^{-1}\sum_{s=1}^{T}\mathcal{K}_h\left(\frac{s-t}{T}\right)z_s^{\text{tv}}R_{i,s},$$
where $h\in\left\{h_i^{\text{sr}},h_i^{\text{lr}}\right\}$ and $z_s^{\text{tv}}=\left(\tilde{X}_{s-1}',C_s'\right)'$. Here, $\mathcal{K}(x)=(2\pi)^{-1/2}\exp(-x^2/2)$. Then, form $\hat{e}_{i,t}=R_{i,t}-\hat{A}_{i,t-1}z_t^{\text{tv}}$. Finally, construct
$$\hat{\Omega}_{f,t}=T^{-1}\sum_{s=1}^{T}\mathcal{K}_h\left(\frac{s-t}{T}\right)\tilde{F}_{s-1}\tilde{F}_{s-1}',\quad \hat{\Sigma}_{e,t}=T^{-1}\sum_{s=1}^{T}\mathcal{K}_h\left(\frac{s-t}{T}\right)\hat{e}_s\hat{e}_s',$$
where $h=h_c$ is a common bandwidth choice. In our application, we use the average bandwidth chosen across the N equations.
3. Estimate
$$\text{vec}\left(\hat{\Lambda}_{\text{ols}}^{\text{tv}}\right)=\left(\sum_{t=0}^{T-1}\left(\tilde{F}_t\tilde{F}_t'\otimes\hat{B}_t'\hat{B}_t\right)+\rho_T\cdot I_{K_C(K_F+1)}\right)^{-1}\sum_{t=0}^{T-1}\left(\tilde{F}_t\otimes\hat{B}_t'\right)\left(R_{t+1}-\hat{B}_t\hat{u}_{t+1}\right),$$
where $\hat{B}_t=\left[\hat{\beta}_{1,t}\cdots\hat{\beta}_{N,t}\right]'$ and $\hat{\beta}_{i,t}$ are the last K_C elements of $\hat{A}_{i,t}$ from Step 2 using the long-run bandwidths h_i^{lr} for $i=1,\dots,N$. Finally, construct the variance estimators
$$\hat{\mathcal{V}}_{\Lambda}^{\text{tv}}=\hat{\mathcal{V}}_{\Lambda,1}^{\text{tv}}+\hat{\mathcal{V}}_{\Lambda,2}^{\text{tv}},$$
where
$$\begin{aligned}\hat{\mathcal{V}}_{\Lambda,1}^{\text{tv}}=T\cdot&\left[\sum_{t=1}^{T}\left(\hat{\Omega}_{f,t}\otimes\hat{B}_{t-1}'\hat{B}_{t-1}\right)\right]^{-1}\\&\times\left[\sum_{t=1}^{T}\left(\left(\hat{\Omega}_{f,t}\hat{\Lambda}_{\text{ols}}^{\text{tv}\prime}D_B'\Omega_{z,t}^{-1}D_B\hat{\Lambda}_{\text{ols}}^{\text{tv}}\hat{\Omega}_{f,t}+\hat{\Omega}_{f,t}\right)\otimes\hat{B}_{t-1}'\hat{\Sigma}_{e,t}\hat{B}_{t-1}\right)\right]\\&\times\left[\sum_{t=1}^{T}\left(\hat{\Omega}_{f,t}\otimes\hat{B}_{t-1}'\hat{B}_{t-1}\right)\right]^{-1}\end{aligned}$$
and
$$\begin{aligned}\hat{\mathcal{V}}_{\Lambda,2}^{\text{tv}}=T\cdot&\left[\sum_{t=1}^{T}\left(\hat{\Omega}_{f,t}\otimes\hat{B}_{t-1}'\hat{B}_{t-1}\right)\right]^{-1}\left[\sum_{t=1}^{T}\left(\hat{\Omega}_{f,t}\otimes\hat{B}_{t-1}'\hat{B}_{t-1}\hat{\Sigma}_{u,t}\hat{B}_{t-1}'\hat{B}_{t-1}\right)\right]\\&\times\left[\sum_{t=1}^{T}\left(\hat{\Omega}_{f,t}\otimes\hat{B}_{t-1}'\hat{B}_{t-1}\right)\right]^{-1}.\end{aligned}$$

Appendix B. Imposing restrictions on parameters

Although the classification of state variables into risk and price of risk factors allows for the specification of more parsimonious models, situations still could exist in which one would like to impose zero (or other linear) restrictions to the parameter of interest Λ (or possibly to B). These restrictions could be most easily imposed by the following steps. Suppose, without loss of generality, the restrictions are of the form $H\,\text{vec}(\theta)=0$, where H is a known $q\times K_C(K_F+1)$ matrix with $\text{rank}(H)=q$, $\theta=(\text{vec}(B)',\text{vec}(\Lambda)')'$, and the restrictions do not violate that $\text{rank}(B'B)=K_C$. For example, if one wanted to impose the restriction that the second element of λ_0 is equal to zero, then $H=\left(0_{NK_C\times 1}',(0,1,0,\dots0)'\right)'$.

Let $\hat{B}$ and $\hat{\Lambda}$, and the corresponding $\hat{\theta}$, stand in for either the OLS or QMLE estimators introduced in this paper. Then the restricted estimator can be found by
$$\hat{\theta}_r=\min_{\theta\text{ s.t. }H\,\text{vec}(\theta)=0}\hat{\theta}'W_T\hat{\theta}=\hat{\theta}-W_T^{-1}H'\left(HW_T^{-1}H'\right)^{-1}H\hat{\theta}.\tag{41}$$

The optimal weight matrix is one that satisfies $W_T \to_p \mathcal{V}_\theta^{-1}$, as $T\to\infty$, where $\mathcal{V}_\theta$ is the asymptotic variance of $\hat{\theta}$. Under this choice of weighting matrix with $H\,\text{vec}(\theta)=0$,

$$\sqrt{T}\left(\text{vec}\left(\hat{\theta}-\theta\right)\right)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{V}_\theta-\mathcal{V}_\theta H'(H\mathcal{V}_\theta H')^{-1}H\mathcal{V}_\theta\right), \tag{42}$$

as $T\to\infty$. In the case of $\hat{B}_{\text{ols}}$ and $\hat{\Lambda}_{\text{ols}}$, $\mathcal{V}_\theta$ is

$$\mathcal{V}_\theta=\begin{bmatrix}\mathcal{V}_{B,\text{ols}} & \mathcal{C}_{\text{ols}}\\ \mathcal{C}'_{\text{ols}} & \mathcal{V}_\Lambda\end{bmatrix},\quad \mathcal{C}_{\text{ols}}=\left[0_{N(K_F+1)}\,\vdots\, I_{NK_C}\right]\mathcal{V}_{\text{rob}}\mathcal{H}_\Lambda(B,\Lambda)', \tag{43}$$

and $\mathcal{V}_{B,\text{ols}}$ is the $(NK_C\times NK_C)$ bottom right sub-matrix of $\mathcal{V}_{\text{rob}}$. In the case of $\hat{B}_{\text{qmle}}$ (or $\hat{B}_{\text{4ols}}$) and $\hat{\Lambda}_{\text{qmle}}$, $\mathcal{V}_\theta$ is

$$\mathcal{V}_\theta=\begin{bmatrix}\mathcal{V}_{B,\text{qmle}} & \mathcal{C}_{\text{qmle}}\\ \mathcal{C}'_{\text{qmle}} & \mathcal{V}_\Lambda\end{bmatrix}, \tag{44}$$

$$\mathcal{C}_{\text{qmle}}=\mathcal{H}_B(B,\Lambda)\mathcal{V}_{\text{rob}}\mathcal{H}_\Lambda(B,\Lambda)',\quad \mathcal{V}_{B,\text{qmle}}=\mathcal{H}_B(B,\Lambda)\mathcal{V}_{\text{rob}}\mathcal{H}_B(B,\Lambda)', \tag{45}$$

and

$$\begin{aligned}\mathcal{H}_B(B,\Lambda)=&\left(\left[\Lambda\Upsilon_{FF}\Lambda'+\Sigma_u\right]^{-1}\left[\Lambda\,\vdots\, I_{K_C}\right]\Upsilon_{ZZ}\otimes I_N\right)\\ &-\left(\left[\Lambda\Upsilon_{FF}\Lambda'+\Sigma_u\right]^{-1}\Lambda\Upsilon_{FF}\otimes B\right)\mathcal{H}_\Lambda(B,\Lambda),\end{aligned} \tag{46}$$

where $\Upsilon_{ZZ}=\text{p}\lim_{T\to\infty}\hat{Z}\hat{Z}'/T$. Further details are provided in Appendix D.

In the case in which betas are time-varying we can follow similar steps. For the linear restriction $H\theta=0$, where $\theta=\text{vec}(\Lambda)$, the restricted estimator can be written as

$$\hat{\theta}_r^{\text{tv}}=\text{vec}\left(\hat{\Lambda}_{\text{ols}}^{\text{tv}}\right)-W_T^{-1}H'\left(HW_T^{-1}H'\right)H\,\text{vec}\left(\hat{\Lambda}_{\text{ols}}^{\text{tv}}\right). \tag{47}$$

The optimal weight matrix is one that satisfies $W_T\to_p(\mathcal{V}_\Lambda^{\text{tv}})^{-1}$ as $T\to\infty$. Under this choice of weighting matrix with $H\theta=0$,

$$\sqrt{T}\left(\hat{\theta}_r^{\text{tv}}-\theta\right)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{V}_\Lambda^{\text{tv}}-\mathcal{V}_\Lambda^{\text{tv}}H'(H\mathcal{V}_\Lambda^{\text{tv}}H')^{-1}H\mathcal{V}_\Lambda^{\text{tv}}\right). \tag{48}$$

Appendix C. Preliminary results

Before proving Theorem 1, we provide some useful results on reduced rank regressions that are used throughout the Appendix. We work in the generality of the model

$$\mathcal{Y}_t=A\mathcal{X}_t+F\mathcal{Z}_t+\varepsilon_t,\quad t=1,\ldots,T, \tag{49}$$

where $A=BD$, B is $n\times k$, D is $k\times m$, $k<\min(n,m)$, $\mathcal{X}_t$ is $m\times 1$, $\mathcal{Z}_t$ is $p\times 1$, and F is full rank. Let $G=[A\,\vdots\, F]$ and $\tilde{\mathcal{Z}}_t=(\mathcal{X}'_t,\mathcal{Z}'_t)'$. If we stack the model, we have $\mathcal{Y}=A\mathcal{X}+F\mathcal{Z}+\mathcal{E}=G\tilde{\mathcal{Z}}+\mathcal{E}$ and $\hat{G}_{\text{ols}}=\mathcal{Y}\tilde{\mathcal{Z}}'\left(\tilde{\mathcal{Z}}\tilde{\mathcal{Z}}'\right)^{-1}$. Under the population moment condition $\mathbb{E}\left[(\tilde{\mathcal{Z}}_t\otimes\varepsilon_t)\right]=0$, the GMM objective function can be written as

$$\begin{aligned}&T\cdot\left(T^{-1}\sum_{t=1}^{T}(\tilde{\mathcal{Z}}_t\otimes\varepsilon_t)\right)'W^{\text{gmm}}\left(T^{-1}\sum_{t=1}^{T}(\tilde{\mathcal{Z}}_t\otimes\varepsilon_t)\right)\\ &=T^{-1}\cdot\text{vec}\left(\left(\mathcal{Y}-\hat{G}_{\text{ols}}\tilde{\mathcal{Z}}+\left(\hat{G}_{\text{ols}}-G\right)\tilde{\mathcal{Z}}\right)\tilde{\mathcal{Z}}'\right)'W^{\text{gmm}}\text{vec}\left(\left(\mathcal{Y}-\hat{G}_{\text{ols}}\tilde{\mathcal{Z}}+\left(\hat{G}_{\text{ols}}-G\right)\tilde{\mathcal{Z}}\right)\tilde{\mathcal{Z}}'\right)\\ &=T\cdot\text{vec}\left(\hat{G}_{\text{ols}}-G\right)'\left(\left(\tilde{\mathcal{Z}}\tilde{\mathcal{Z}}'/T\right)\otimes I_n\right)W^{\text{gmm}}\left(\left(\tilde{\mathcal{Z}}\tilde{\mathcal{Z}}'/T\right)\otimes I_n\right)\text{vec}\left(\hat{G}_{\text{ols}}-G\right),\end{aligned} \tag{50}$$

which is the MD criterion function with $W^{\text{md}}=\left(\left(\tilde{\mathcal{Z}}\tilde{\mathcal{Z}}'/T\right)\otimes I_n\right)W^{\text{gmm}}\left(\left(\tilde{\mathcal{Z}}\tilde{\mathcal{Z}}'/T\right)\otimes I_n\right)$. Thus, the GMM and MD criterion functions are one-to-one. To show that ML is a special case of MD/GMM, note that under the assumption $\text{vec}(\mathcal{E})\sim\mathcal{N}(0,(I_T\otimes\sigma^2 I_n))$ the log-likelihood is $\ell(B,D_0,\sigma^2)=-\frac{nT}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\,\text{tr}\,(\mathcal{E}'\mathcal{E})$. However,

$$\text{tr}(\mathcal{E}'\mathcal{E})=\text{tr}\left(\left(\mathcal{Y}-\hat{G}_{\text{ols}}\tilde{\mathcal{Z}}\right)'\left(\mathcal{Y}-\hat{G}_{\text{ols}}\tilde{\mathcal{Z}}\right)\right)+\text{tr}\left(\left(\hat{G}_{\text{ols}}-G\right)'\left(\hat{G}_{\text{ols}}-G\right)\tilde{\mathcal{Z}}\tilde{\mathcal{Z}}'\right), \tag{51}$$

so that the ML, which solves $\min_G\text{tr}(\mathcal{E}'\mathcal{E})$ is the same as the MD estimator with weight matrix $\left(\left(\tilde{\mathcal{Z}}\tilde{\mathcal{Z}}'/T\right)\otimes I_n\right)$. Under the general symmetric weight function $W^{\text{md}}=\left(W_1^{\text{md}}\otimes W_2^{\text{md}}\right)$ with

$$W_1^{\text{md}}=\begin{bmatrix}W_{1,11}^{\text{md}} & W_{1,12}^{\text{md}}\\ W_{1,12}^{\text{md}\prime} & W_{1,22}^{\text{md}}\end{bmatrix}, \tag{52}$$

and the normalization $B'W_2^{\mathrm{md}}B = I_k$ (because B and D are not separately identified without further assumption), it can be shown that the MD estimators are

$$\hat{B}_{\mathrm{md}} = \left(W_2^{\mathrm{md}}\right)^{-1/2}\mathcal{L}, \quad \hat{D}_{\mathrm{md}} = \hat{B}_{\mathrm{md}}'W_2^{\mathrm{md}}\hat{A}_{\mathrm{ols}}, \quad \hat{F}_{\mathrm{md}} = \hat{F}_{\mathrm{ols}} + \left(\hat{A}_{\mathrm{ols}} - \hat{B}_{\mathrm{md}}\hat{D}_{\mathrm{md}}\right)W_{1,12}^{\mathrm{md}}\left(W_{1,22}^{\mathrm{md}}\right)^{-1}, \tag{53}$$

where $\mathcal{L} = [\zeta_1 \cdots \zeta_k]$ and ζ_i is eigenvector associated with the ith largest eigenvalue of the matrix

$$\left(W_2^{\mathrm{md}}\right)^{1/2}\hat{A}_{\mathrm{ols}}\left(W_{1,11}^{\mathrm{md}} - W_{1,12}^{\mathrm{md}}\left(W_{1,22}^{\mathrm{md}}\right)^{-1}W_{1,21}^{\mathrm{md}}\right)\hat{A}_{\mathrm{ols}}'\left(W_2^{\mathrm{md}}\right)^{1/2}. \tag{54}$$

This follows because

$$\begin{aligned}&\min_{B,D,F}\mathrm{vec}\left(\hat{G}_{\mathrm{ols}} - G\right)'\left(W_1^{\mathrm{md}} \otimes W_2^{\mathrm{md}}\right)\mathrm{vec}\left(\hat{G}_{\mathrm{ols}} - G\right) \\ &\quad = \mathrm{tr}\left(\hat{G}_{\mathrm{ols}}'W_2^{\mathrm{md}}\hat{G}_{\mathrm{ols}}W_1^{\mathrm{md}}\right) + \mathrm{tr}\left(G'W_2^{\mathrm{md}}GW_1^{\mathrm{md}}\right) - 2\,\mathrm{tr}\left(\hat{G}_{\mathrm{ols}}'W_2^{\mathrm{md}}GW_1^{\mathrm{md}}\right).\end{aligned} \tag{55}$$

We can ignore the first term as it is not a function of B, D, or F. If we fix A (i.e., B and D) and solve for F, then

$$\hat{F}_{\mathrm{md}} = \left(\left(\hat{A}_{\mathrm{ols}} - A\right)W_{1,12}^{\mathrm{md}} + \hat{F}_{\mathrm{ols}}W_{1,22}^{\mathrm{md}}\right)\left(W_{1,22}^{\mathrm{md}}\right)^{-1} = \hat{F}_{\mathrm{ols}} + \left(\hat{A}_{\mathrm{ols}} - BD\right)W_{1,12}^{\mathrm{md}}\left(W_{1,22}^{\mathrm{md}}\right)^{-1}, \tag{56}$$

and plugging this back in and using the normalization $B'W_2^{\mathrm{md}}B = I_k$ we can obtain

$$\begin{aligned}&\min_{B,D,F}\mathrm{vec}\left(\hat{G}_{\mathrm{ols}} - G\right)'\left(W_1^{\mathrm{md}} \otimes W_2^{\mathrm{md}}\right)\mathrm{vec}\left(\hat{G}_{\mathrm{ols}} - G\right) \\ &\quad = \mathrm{tr}\left(D\left(W_{1,11}^{\mathrm{md}} - W_{1,12}^{\mathrm{md}}\left(W_{1,22}^{\mathrm{md}}\right)^{-1}W_{1,12}^{\mathrm{md}\prime}\right)D'\right) \\ &\quad\quad - 2\cdot\mathrm{tr}\left(D\left(W_{1,11}^{\mathrm{md}} - W_{1,12}^{\mathrm{md}}\left(W_{1,22}^{\mathrm{md}}\right)^{-1}W_{1,12}^{\mathrm{md}\prime}\right)\hat{A}_{\mathrm{ols}}'W_2^{\mathrm{md}}B\right).\end{aligned} \tag{57}$$

Given B we can solve for D, which yields $\hat{D}_{\mathrm{md}} = \hat{B}_{\mathrm{md}}'W_2^{\mathrm{md}}\hat{A}_{\mathrm{ols}}$. Plugging this back in yields the following maximization problem:

$$\max_{\tilde{B}}\mathrm{tr}\left(\tilde{B}'\left(W_2^{\mathrm{md}}\right)^{1/2}\hat{A}_{\mathrm{ols}}\left(W_{1,11}^{\mathrm{md}} - W_{1,12}^{\mathrm{md}}\left(W_{1,22}^{\mathrm{md}}\right)^{-1}W_{1,12}^{\mathrm{md}\prime}\right)\hat{A}_{\mathrm{ols}}'\left(W_2^{\mathrm{md}}\right)^{1/2}\tilde{B}\right) \quad \text{s.t. } \tilde{B}'\tilde{B} = I_k, \tag{58}$$

where $\tilde{B} = \left(W_2^{\mathrm{md}}\right)^{1/2}B$ and the result follows.

Using these derivations, it is straightforward to form a bias-corrected estimator of Λ for the bias induced by replacing u_{t+1} by $\hat{u}_{t+1}$. In particular, this bias arises because $\hat{u}_{t+1}$ is a function of $X_{1,t}$, which does not show up in the formulation for returns in Eq. (13). The prescription to deal with this bias is simply to include $X_{1,t}$ in the first-step regression (associated with a full-rank coefficient matrix). The degree of the bias is affected by a subset of elements of Φ, namely, those parameters that designate how strong the predictive power of X_1-type variables is for X_1- and X_2-type variables. The proofs of Theorems 1 and 2 can then be straightforwardly modified to provide appropriate limiting distributions for these estimators using the results in this section and Appendix D.

Appendix D. Proofs

D.1. Constant betas

For the results in the constant-beta case, we make the following assumptions (in addition to those made in the main text): (i) all eigenvalues of Φ have modulus less than one; (ii) $\Sigma_{v,t} = \Sigma_v$ for all t and Σ_v is positive definite; (iii) the initial condition X_0 is fixed; (iv) $(R_t', v_t)'$ is a stationary ergodic sequence with $\mathbb{E}\|(R_t', v_t)'\|^4 < \infty$; (v) the matrix $B'B$ has minimum eigenvalue bounded away from zero; and (vi) $\mathbb{E}[e_{i,t}v_t v_t' | \mathcal{F}_{t-1}] = 0\ \forall t$ and $i = 1, \ldots, N$.

All of these assumptions are standard except perhaps assumption (vi). Assumption (i) ensures that the dynamics of X_t are stationary. From an economic perspective, this restriction rules out phenomena such as rational bubbles that would be associated with exploding risk premiums. From a statistical point of view, the assumption allows us to avoid nonstandard asymptotic arguments. Assumption (ii) is natural given that B does not time-vary in this case. Assumption (iii) ensures that the influence of the initial condition is asymptotically negligible. Assumption (v) guarantees that the matrix $B'B$ satisfies $\mathrm{rank}(B'B) = K_C$. Intuitively, we are assuming away the presence of redundant, uninformative or unspanned factors. Assumption (vi) limits the degree of dependence between $e_{i,t}$ and v_t and consequently simplifies our asymptotic variance formulas. To provide intuition for this assumption note that it would hold in the case that we assumed that $(R_t', v_t)'$ are jointly independent and identically distributed, conditional on $\mathcal{F}_{t-1}$, from an elliptically symmetric distribution. Under these assumptions we have the following results

$$\begin{bmatrix} T^{-1/2}\cdot \mathrm{vec}\left(V\tilde{X}'_{-}\right) \\ T^{-1/2}\cdot \mathrm{vec}\left(E\tilde{X}'_{-}\right) \\ T^{-1/2}\cdot \mathrm{vec}(EV') \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} (\Upsilon_{XX}\otimes\Sigma_v) & 0 \\ 0 & \mathcal{V}_{\mathrm{rob}} \end{bmatrix}\right), \tag{59}$$

and $\hat{\mathcal{V}}_{\mathrm{rob}} \xrightarrow{p} \mathcal{V}_{\mathrm{rob}}$, where $\Upsilon_{XX} = \mathrm{p}\lim_{T\to\infty} \tilde{X}_{-}\tilde{X}'_{-}/T$.

D.1.1. Proof of Theorem 1

We first show the result for $\hat{\Lambda}_{\mathrm{ols}}$. Let $\hat{A}_{01,\mathrm{ols}} = \left[\hat{A}_{1,\mathrm{ols}} \,\vdots\, \hat{A}_{2,\mathrm{ols}}\right]$ so that $\hat{\Lambda}_{\mathrm{ols}} = \left(\hat{B}'_{\mathrm{ols}}\hat{B}_{\mathrm{ols}}\right)^{-1}\hat{B}'_{\mathrm{ols}}\hat{A}_{01,\mathrm{ols}}$. Then,

$$\begin{aligned} &\hat{\Lambda}_{\mathrm{ols}} \\ &\quad= \left(\hat{B}'_{\mathrm{ols}}\hat{B}_{\mathrm{ols}}\right)^{-1}\hat{B}'_{\mathrm{ols}} R M_{\hat{U}}\tilde{F}'_{-}\left(\tilde{F}_{-}M_{\hat{U}}\tilde{F}'_{-}\right)^{-1} \\ &\quad= \Lambda + U M_{\hat{U}}\tilde{F}'_{-}\left(\tilde{F}_{-}M_{\hat{U}}\tilde{F}'_{-}\right)^{-1} + \left(\hat{B}'_{\mathrm{ols}}\hat{B}_{\mathrm{ols}}\right)^{-1}\hat{B}'_{\mathrm{ols}} E M_{\hat{U}}\tilde{F}'_{-}\left(\tilde{F}_{-}M_{\hat{U}}\tilde{F}'_{-}\right)^{-1} \\ &\qquad - \left(\hat{B}'_{\mathrm{ols}}\hat{B}_{\mathrm{ols}}\right)^{-1}\hat{B}'_{\mathrm{ols}}\left(\hat{B}_{\mathrm{ols}}-B\right)\Lambda - \left(\hat{B}'_{\mathrm{ols}}\hat{B}_{\mathrm{ols}}\right)^{-1}\hat{B}'_{\mathrm{ols}}\left(\hat{B}_{\mathrm{ols}}-B\right) U M_{\hat{U}}\tilde{F}'_{-}\left(\tilde{F}_{-}M_{\hat{U}}\tilde{F}'_{-}\right)^{-1}, \end{aligned} \tag{60}$$

where $M_{\hat{U}} = I_T - \hat{U}'\left(\hat{U}\hat{U}'\right)^{-1}\hat{U}$. Under our assumptions, the last term is $\mathrm{o_p}\left(T^{-1/2}\right)$ so that

$$\sqrt{T}\,\mathrm{vec}\left(\hat{\Lambda}_{\mathrm{ols}}-\Lambda\right) = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3 + \mathrm{o_p}(1), \tag{61}$$

where

$$\mathcal{T}_1 = \left(\left(\tilde{F}_{-}\tilde{F}'_{-}/T\right)^{-1}\otimes I_{K_C}\right)\mathrm{vec}\left(T^{-1/2}U\tilde{F}'\right), \tag{62}$$

$$\mathcal{T}_2 = \left(I_{(K_F+1)}\otimes (B'B)^{-1}B'\right)\mathrm{vec}\left(\sqrt{T}\left(\hat{A}_{01,\mathrm{ols}}-A\right)\right), \tag{63}$$

and

$$\mathcal{T}_3 = -\left(\Lambda'\otimes (B'B)^{-1}B'\right)\mathrm{vec}\left(\sqrt{T}\left(\hat{B}_{\mathrm{ols}}-B\right)\right). \tag{64}$$

Under our assumptions, $\sqrt{T}\left(\mathrm{vec}\left(\hat{A}_{\mathrm{ols}}-A\right)\right)\to_d \mathcal{N}(0,\mathcal{V}_{\mathrm{rob}})$ and is asymptotically uncorrelated with $\mathrm{vec}\left(T^{-1/2}U\tilde{F}'\right)\to_d \mathcal{N}(0,(\Upsilon_{FF}\otimes\Sigma_u))$ and so the result follows.

Now let us consider $\hat{\Lambda}_{\mathrm{qmle}}$. By the derivations above (when $F=0$) and with weight matrix $W_1^{\mathrm{md}} = \hat{Z}\hat{Z}'/T$ and $W_2^{\mathrm{md}} = I_N$, then we can find $\hat{B}_{\mathrm{qmle}}$ and $\hat{\Lambda}_{\mathrm{qmle}}$ as transformations of the K_C eigenvectors associated with the largest eigenvalues of the matrix $\hat{A}_{\mathrm{ols}}\left(\hat{Z}\hat{Z}'/T\right)\hat{A}'_{\mathrm{ols}}$ (see Appendix A). By standard properties of MD estimators, we know that the asymptotic variance of $\left(\mathrm{vec}\left(\hat{B}_{\mathrm{qmle}}\right),\mathrm{vec}\left(\hat{\Lambda}_{\mathrm{qmle}}\right)\right)'$ is

$$\mathcal{V}_{B\Lambda,\mathrm{qmle}} = \left(\mathcal{J}'_{\mathrm{md}}W^{\mathrm{md}}\mathcal{J}_{\mathrm{md}}\right)^{-1}\left(\mathcal{J}'_{\mathrm{md}}W^{\mathrm{md}}\mathcal{V}_{\mathrm{rob}}W^{\mathrm{md}}\mathcal{J}_{\mathrm{md}}\right)\left(\mathcal{J}'_{\mathrm{md}}W^{\mathrm{md}}\mathcal{J}_{\mathrm{md}}\right)^{-1}, \tag{65}$$

where

$$\begin{aligned} \mathcal{J}_{\mathrm{md}} &= \left[\frac{\partial\,\mathrm{vec}\left(\hat{A}_{\mathrm{ols}}-A\right)}{\partial\,\mathrm{vec}(B)'} \,\vdots\, \frac{\partial\,\mathrm{vec}\left(\hat{A}_{\mathrm{ols}}-A\right)}{\partial\,\mathrm{vec}(\Lambda)'}\right] \\ &= \left[-\left([\Lambda \,\vdots\, I_{K_C}]'\otimes I_N\right) \,\vdots\, -\left(I_{(K_C+K_F+1)}\otimes B\right)\left[I_{K_C(K_F+1)} \,\vdots\, 0_{K_C(K_F+1)\times K_C^2}\right]'\right]. \end{aligned} \tag{66}$$

After incorporating the uncertainty from replacing U by $\hat{U}$, it can then be shown that this yields

$$\mathcal{V}_{B\Lambda,\mathrm{qmle}} = \begin{bmatrix} \mathcal{H}_B(B,\Lambda)\mathcal{V}_{\mathrm{rob}}\mathcal{H}_B(B,\Lambda)' & \mathcal{H}_B(B,\Lambda)\mathcal{V}_{\mathrm{rob}}\mathcal{H}_\Lambda(B,\Lambda)' \\ \mathcal{H}_\Lambda(B,\Lambda)\mathcal{V}_{\mathrm{rob}}\mathcal{H}_B(B,\Lambda)' & \left(\Upsilon_{FF}^{-1}\otimes\Sigma_u\right)+\mathcal{H}_\Lambda(B,\Lambda)\mathcal{V}_{\mathrm{rob}}\mathcal{H}_\Lambda(B,\Lambda)' \end{bmatrix}, \tag{67}$$

where

$$\begin{aligned} \mathcal{H}_B(B,\Lambda) &= \left(\left[\Lambda\Upsilon_{FF}\Lambda'+\Sigma_u\right]^{-1}\left[\Lambda \,\vdots\, I_{K_C}\right]\Upsilon_{ZZ}\otimes I_N\right) \\ &\quad - \left(\left[\Lambda\Upsilon_{FF}\Lambda'+\Sigma_u\right]^{-1}\Lambda\Upsilon_{FF}\otimes B\right)\mathcal{H}_\Lambda(B,\Lambda), \end{aligned} \tag{68}$$

and $\Upsilon_{ZZ} = \mathrm{plim}_{T\to\infty}\hat{Z}\hat{Z}'/T = \mathrm{plim}_{T\to\infty}\tilde{Z}\tilde{Z}'/T$.

D.1.2. Proof of Theorem 2

Let $\mu_F=\mathbb{E}[F_t]$ and $\hat{\mu}_F=T^{-1}\sum_{t=1}^{T}F_t$. Here we derive the asymptotic distribution of the estimator $\hat{\overline{\lambda}}=\hat{\lambda}_0+\hat{\Lambda}_1\hat{\mu}_F$. We could also estimate μ_F by the last K_F elements of $\left(I_K-\hat{\Phi}\right)^{-1}\hat{\mu}$. These two approaches are asymptotically equivalent. Then,

$$\hat{\overline{\lambda}}-\overline{\lambda}=\left(\hat{\lambda}_0-\lambda_0\right)+\left(\hat{\Lambda}_1-\Lambda_1\right)\mu_F+\Lambda_1(\hat{\mu}_F-\mu_F)+o_p\left(T^{-1/2}\right). \tag{69}$$

Define $\tilde{\mu}_F=(1,\mu_F')'$ and $\tilde{\Lambda}_1=[0_{K_C\times K_1}\Lambda_1]$ so that

$$\hat{\overline{\lambda}}-\overline{\lambda}=(\tilde{\mu}_F'\otimes I_{K_C})\text{vec}\left(\hat{\Lambda}-\Lambda\right)+\tilde{\Lambda}_1(\hat{\mu}-\mu)+o_p\left(T^{-1/2}\right), \tag{70}$$

where $\hat{\mu}_X=T^{-1}\sum_{t=1}^{T}X_t$. Note that

$$\sqrt{T}(\hat{\mu}_X-\mu_X)=(I_K-\Phi)^{-1}T^{-1/2}V\iota_T+o_p(1) \tag{71}$$

and, from the proof of Theorem 1,

$$\text{vec}\left(\sqrt{T}\left(\hat{\Lambda}_{\text{ols}}-\Lambda\right)\right)=\left(\Upsilon_{FF}^{-1}\otimes I_{K_C}\right)\text{vec}\left(U\tilde{F}'_-\right)+\mathcal{H}_\Lambda(B,\Lambda)\sqrt{T}\,\text{vec}\left(\hat{A}_{\text{ols}}-A\right)+o_p(1). \tag{72}$$

Under our assumptions, the only covariance term arises from

$$T^{-1}\,\text{vec}\left(V\iota_T/\sqrt{T}\right)\text{vec}\left(U\tilde{F}'/\sqrt{T}\right)'=T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\tilde{F}'_{s-1}\otimes v_tu_s'\right). \tag{73}$$

For $s\neq t$, the sum converges in probability to zero under our assumptions so that

$$\sqrt{T}\left(\hat{\overline{\lambda}}-\overline{\lambda}\right)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{V}_{\overline{\lambda}}\right), \tag{74}$$

where

$$\mathcal{V}_{\overline{\lambda}}=(\tilde{\mu}_F'\otimes I_{K_C})\mathcal{V}_\Lambda(\tilde{\mu}_F'\otimes I_{K_C})'+\tilde{\Lambda}_1(I_K-\Phi)^{-1}\Sigma_v\left[(I_K-\Phi)^{-1}\right]'\tilde{\Lambda}_1'+C_{\overline{\lambda}}+C_{\overline{\lambda}}', \tag{75}$$

$C_{\overline{\lambda}}=\tilde{\Lambda}_1(I_K-\Phi)^{-1}\Sigma_{vu}$, and Σ_{vu} is formed from the first K_C columns of the matrix Σ_v.

D.1.3. Derivations for Appendix B

First we derive the asymptotic covariance matrix $\mathcal{C}_{\text{ols}}$. The asymptotic variance of $\sqrt{T}\,\text{vec}\left(\hat{B}_{\text{ols}}-B\right)$ is the bottom-right block element of the matrix $\mathcal{V}_{\text{rob}}$ and, from the proof of Theorem 1,

$$\text{vec}\left(\sqrt{T}\left(\hat{\Lambda}_{\text{ols}}-\Lambda\right)\right)=\left(\Upsilon_{FF}^{-1}\otimes I_{K_C}\right)\text{vec}\left(U\tilde{F}'_-\right)+\mathcal{H}_\Lambda(B,\Lambda)\sqrt{T}\,\text{vec}\left(\hat{A}_{\text{ols}}-A\right)+o_p(1). \tag{76}$$

Thus, $\mathcal{C}_{\text{ols}}=\left[0_{N(K_F+1)}\,\vdots\,I_{NK_C}\right]\mathcal{V}_{\text{rob}}\mathcal{H}_\Lambda(B,\Lambda)'$. Next, we derive the asymptotic covariance matrix $\mathcal{C}_{\text{qmle}}$. From the proof of Theorem 1,

$$\text{vec}\left(\sqrt{T}\left(\hat{B}_{\text{qmle}}-B\right)\right)=\mathcal{H}_B(B,\Lambda)\sqrt{T}\,\text{vec}\left(\hat{A}_{\text{ols}}-A\right)+o_p(1),$$

so that under our assumptions $\mathcal{C}_{\text{qmle}}=\mathcal{H}_B(B,\Lambda)\mathcal{V}_{\text{rob}}\mathcal{H}_\Lambda(B,\Lambda)'$.

D.2. Time-varying betas

For the results in the time-varying beta case, we proceed conditional on the realizations of the random processes $\Psi(\cdot)$ and $\beta_i(\cdot)$ for $i=1,\ldots,N$. However, we suppress these arguments in the expectation operator to simplify notation. To simplify the notation in this Appendix we map, without loss of generality, $\beta_{i,t}\mapsto\beta_{i,t+1}$, $\mu_t\mapsto\mu_{t+1}$, and $\Phi_t\mapsto\Phi_{t+1}$. For the case in which betas are time-varying it is more convenient to state the assumptions in terms of the linear model $\mathcal{Y}_{t,T}=\mathcal{A}_{t,T}\mathcal{X}_{t,T}+\mathcal{E}_{t,T}$, $t=1,\ldots,T$, which nests both Eqs. (29) and (30). Although this is a triangular array of models, we suppress the dependence on T for simplicity of notation. Finally, define $\Omega_{z,t}=\mathbb{E}\left[z_t^{\text{tv}}z_t^{\text{tv}\prime}\right]$, $\Omega_{f,t}=\mathbb{E}\left[\tilde{F}_{t-1}\tilde{F}'_{t-1}\right]$, $\Omega_{x,t}=\mathbb{E}\left[\tilde{X}_{t-1}\tilde{X}'_{t-1}\right]$, $\Sigma_{e,t}=\mathbb{E}\left[e_te_t'\right]$, and $\Sigma_{v,t}=\mathbb{E}\left[v_tv_t'\right]$, where $\Sigma_{u,t}$ is the matrix formed from the first K_C rows and columns of $\Sigma_{v,t}$. We make the following assumptions.

(i) For all $t\geq1$, $\sup_{T\geq1}\sup_{t\leq T}\mathbb{E}\left[\|(\mathcal{X}_t',\mathcal{E}_t')\|^{8+4\delta}\right]<\infty$ for some $\delta>0$ and is mixing where the mixing coefficients

$$m_T(i)=\sup_{-T\leq\ell\leq T}\;\sup_{A\in\mathcal{F}_{-\infty}^{\ell},B\in\mathcal{F}_{T+i}^{\infty}}\left|\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B)\right|$$

satisfy $m_T(i)\leq m(i)$, $T\geq0$, and the dominating sequence $m(i)$ is geometrically decreasing. $\mathcal{E}_t$ is a martingale difference sequence with respect to $\mathcal{F}_t=\sigma(\mathcal{X}_t,\mathcal{E}_{t-1},\mathcal{X}_{t-1},\mathcal{E}_{t-2},\ldots)$.

(ii) The observed data $\{(\mathcal{Y}_t, \mathcal{X}_t'): t = 0, \ldots, T\}$ have been symmetrically trimmed with positive trimming sequence a_T, which satisfies $a_T/b_T \to 0$, $a_T/h_T \to 0$, and $\sqrt{T}a_T \to 0$.

(iii) The sequence $\beta_{i,t}$ satisfies $\beta_{i,t} = \beta_i(t/T) + o(1)$ for $i = 1, \ldots, N$ and similarly for Ψ_t, $\Omega_{x,t}$, $\Sigma_{e,t}$, and $\Sigma_{v,t}$ for some functions $\beta_i(\cdot)$, $\Psi(\cdot) \equiv [\mu(\cdot)\ \Phi(\cdot)]$, $\Omega_x(\cdot)$, $\Sigma_e(\cdot)$ and $\Sigma_v(\cdot)$, respectively. The elements of these functions are in $\mathcal{C}^r[0,1]$, the space of r times continuously differentiable functions, for some $r \geq 2$. For all $\tau \in [0,1]$, $\Omega_x(\tau)$, $\Sigma_e(\tau)$, and $\Sigma_v(\tau)$ are positive definite with eigenvalues bounded and bounded away from zero. Finally, $\sup_{0 \leq \tau \leq 1} |\gamma_{\max}(\Phi(\tau))|$ is bounded below one, where $\gamma_{\max}(\cdot)$ is the maximum eigenvalue of a matrix.

(iv) $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} (\Omega_{f,t} \otimes B_t'B_t) = \int_0^1 (\Omega_f(\tau) \otimes B(\tau)'B(\tau))\, d\tau$ exists and is positive definite with all eigenvalues bounded and bounded away from zero. Also,

$$\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\left(\left(\Omega_{f,t}\Lambda' D_B'\Omega_{z,t}^{-1}D_B\Lambda\Omega_{f,t} + \Omega_{f,t}\right) \otimes B_t'\Sigma_{e,t}B_t\right)$$
$$= \int_0^1 \left(\left(\Omega_f(\tau)\Lambda' D_B'\Omega_z(\tau)^{-1}D_B\Lambda\Omega_f(\tau) + \Omega_f(\tau)\right) \otimes B(\tau)'\Sigma_e(\tau)B(\tau)\right) d\tau \tag{77}$$

and

$$\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} (\Omega_{f,t} \otimes B_t'B_t\Sigma_{u,t}B_t'B_t)$$
$$= \int_0^1 (\Omega_f(\tau) \otimes B(\tau)'B(\tau)\Sigma_u(\tau)B(\tau)'B(\tau))\, d\tau, \tag{78}$$

and $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \mathbb{E}[X_t]$ exist.

(v) X_0 is fixed, and for $1 \leq i \leq N$, $\mathbb{E}[e_{i,t}v_tv_t' | \mathcal{F}_{t-1}] = 0$.

(vi) The kernel $\mathcal{K}$ satisfies the following conditions. There exists $B, L < \infty$ such that either $\mathcal{K}(w) = 0$ for $\|w\| > L$ and $|\mathcal{K}(w) - \mathcal{K}(w')| \leq B\|w - w'\|$, or $\mathcal{K}(w)$ is differentiable with $|\partial\mathcal{K}(w)/\partial w| \leq B$ and, for some $\varrho > 1$, $|\partial\mathcal{K}(w)/\partial w| \leq \|w\|^{-\varrho}$ for $\|w\| \geq L$. Also, $|\mathcal{K}(w)| \leq B\|w\|^{-\varrho}$ for $\|w\| \geq L$. $\int \mathcal{K}(w)\, dw = 1$ and for some $r \geq 2$, $\int w^i\mathcal{K}(w)\, dw = 0$ for $i = 1, \ldots, r-1$ and $\int |w|^r\mathcal{K}(w)\, dw < \infty$.

(vii) The sequence ρ_T satisfies $\sqrt{T}\rho_T \to 0$. The bandwidth sequence h_T satisfies $Th_T^{2r} \to 0$, $\log(T)^2 / Th_T^2 \to 0$, and $T^{(\epsilon-1)/2}h_T^{-(1+\delta)/(2+\delta)} \to 0$ for some $\epsilon > 0$. The bandwidth sequence b_T satisfies $Tb_T^{2r} \to 0$, $\log(T)^2 / Tb_T^2 \to 0$, and $T^{(\epsilon-1)/2}b_T^{-(1+\delta)/(2+\delta)} \to 0$ for some $\epsilon > 0$.

Assumptions (i)–(iii) and (vi)–(vii) are essentially the same as those of Kristensen (2012). The remaining assumptions are tailored to our model specification. Following similar steps as in Section 3, the martingale difference assumption implies that $\mathbb{E}[M_{t+1}R_{t+1}|(e_s, v_s: s \leq t)] = 0$. Thus, these assumptions are consistent with the asset pricing restrictions discussed in Section 3. When implementing the estimators introduced in this paper, different bandwidths should be used for each series (see Appendix A.2), however, without loss of generality, the derivations rely on a common bandwidth choice h_T and b_T to simplify the presentation. In addition, we suppress the dependence of the bandwidth sequences on T. Finally, define $\prod_{i=1}^{m} A_i = A_1A_2\cdots A_m$ for a sequence of square matrices and $\|A\| = \sqrt{\mathrm{tr}(A'A)}$ the matrix Euclidean norm.

To proceed, we make repeated use of the following two lemmas. The second lemma is Lemma B.11 of Kristensen (2012). We restate it for convenience.

Lemma D.1. Under our assumptions,

(a) $$\hat{\Psi}_t - \Psi_t = T^{-1}\sum_{s=1}^{T}\mathcal{K}_b\left(\frac{s-t}{hT}\right)\cdot\left[(\Psi_s - \Psi_t)\tilde{X}_{s-1}\tilde{X}_{s-1}' + v_s\tilde{X}_{s-1}'\right]\Omega_{x,t}^{-1} + O_p\left(b^{2r}\right) + O_p\left(\frac{\log T}{bT}\right),$$

(b) $$\hat{A}_t - A_t = T^{-1}\sum_{s=1}^{T}\mathcal{K}_h\left(\frac{s-t}{T}\right)\cdot[(A_s - A_t)z_sz_s' + e_sz_s']\Omega_{z,t}^{-1} + O_p\left(h^{2r}\right) + O_p\left(\frac{\log T}{hT}\right),$$

(c) $$\sup_{1 \leq t \leq T}\left\|\hat{A}_t - A_t\right\| = O_p(h^r) + O_p\left(\sqrt{\frac{\log(T)}{hT}}\right),$$

(d) $$\sup_{1 \leq t \leq T}\left\|\hat{\Psi}_t - \Psi_t\right\| = O_p(b^r) + O_p\left(\sqrt{\frac{\log(T)}{bT}}\right),$$

uniformly over $1 \leq t \leq T$.

Proof of Lemma D.1. Parts (a) and (b) follow by the same steps as in the proof of Theorem 2 in Ang and Kristensen (2009). Parts (c) and (d) follow Kristensen (2009).□

Lemma D.2. Assume that $m(t)^{\delta/(2+\delta)} = o(t^{-2+\epsilon})$ *for some* $\delta, \epsilon > 0$. *Then, for any symmetric function* $\phi_T(Y_s, Y_t)$, *the following decomposition holds*:

$$\binom{T}{2}^{-1} \sum_{s<t} \phi_T(Y_s, Y_t) = \theta_T + \frac{2}{T} \sum_{t=1}^{T} \left[\overline{\phi}_T(Y_t) - \theta_T \right] + \Re_T,$$

where

$$\theta_T = \binom{T}{2}^{-1} \sum_{s<t} \mathbb{E}[\phi_T(Y_s, Y_t)], \quad \overline{\phi}_T(y) = \mathbb{E}[\phi_T(y, Y_t)],$$

and

$$\left(\mathbb{E}\left[\Re_T^2 \right] \right)^{1/2} = O\left(\mathfrak{s}_{T,\delta} \cdot T^{-1+\epsilon/2} \right), \quad \mathfrak{s}_{T,\delta} = \sup_{s \neq t} \left(\mathbb{E}\left[\left| \phi_T(Y_s, Y_t) \right|^{2+\delta} \right] \right)^{1/(2+\delta).}$$

D.2.1. Proof of Theorem3

Throughout we use z_t instead of z_t^{tv} for simplicity of notation. We first find the asymptotic linear representation of

$$\sqrt{T}\,\text{vec}\left(\hat{\Lambda}_{\text{ols}}^{\text{tv}} \right) = \left[T^{-1} \sum_{t=1}^{T} \left(\tilde{F}_{t-1} \tilde{F}'_{t-1} \otimes \hat{B}'_t \hat{B}_t \right) + \rho_T \cdot I_{K_C(K_F+1)} \right]^{-1} \times T^{-1/2} \sum_{t=1}^{T} \left(\tilde{F}_{t-1} \otimes \hat{B}'_t \right) \text{vec}\left(R_t - \hat{B}_t \hat{u}_t \right). \tag{79}$$

The first factor satisfies

$$\left[T^{-1} \sum_{t=1}^{T} \left(\tilde{F}_{t-1} \tilde{F}'_{t-1} \otimes \hat{B}'_t \hat{B}_t \right) + \rho_T \right]^{-1} = \left[\int \left(\Omega_f(\tau) \otimes B(\tau)' B(\tau) \right) d\tau \right]^{-1} + o_p(1). \tag{80}$$

This follows because

$$\begin{aligned} &\left\| T^{-1} \sum_{t=1}^{T} \left(\tilde{F}_{t-1} \tilde{F}'_{t-1} \otimes \left(\hat{B}_t - B_t \right)' B_t \right) \right\| \\ &\leq T^{-1} \sum_{t=1}^{T} \left\| \tilde{F}_{t-1} \tilde{F}'_{t-1} \right\| \left\| \left(\hat{B}_t - B_t \right)' B_t \right\| \\ &\leq C \sup_{1 \leq t \leq T} \left\| \left(\hat{B}_t - B_t \right) \right\| \sup_{1 \leq t \leq T} \|B_t\| \cdot T^{-1} \textstyle\sum_{t=1}^{T} \left\| \tilde{F}_{t-1} \tilde{F}'_{t-1} \right\| \\ &= C \sup_{1 \leq t \leq T} \left\| \left(\hat{B}_t - B_t \right) \right\| \cdot T^{-1} \sum_{t=1}^{T} \text{tr}\left(\tilde{F}_{t-1} \tilde{F}'_{t-1} \right) \\ &= O_p(h^r) + O_p\left(\sqrt{\frac{\log(T)}{hT}} \right), \end{aligned} \tag{81}$$

and, by similar steps,

$$T^{-1} \sum_{t=1}^{T} \left(\tilde{F}_{t-1} \tilde{F}'_{t-1} \otimes \left(\hat{B}_t - B_t \right)' \left(\hat{B}_t - B_t \right) \right) = O_p\left(h^{2r} \right) + O_p\left(\frac{\log(T)}{hT} \right). \tag{82}$$

Thus we just need to deal with the term $T^{-1/2} \sum_{t=1}^{T} \hat{B}'_t \left(R_t - \hat{B}_t \hat{u}_t \right)$. Combining

$$R_t - \hat{B}_t \hat{u}_t = \hat{B}_t \Lambda \tilde{F}_{t-1} - \left(\hat{B}_t - B_t \right) \left(\Lambda \tilde{F}_{t-1} + u_t \right) - B_t (\hat{u}_t - u_t) - \left(\hat{B}_t - B_t \right) (\hat{u}_t - u_t) + e_t, \tag{83}$$

and, by similar steps as above, that

$$\left\| T^{-1/2} \sum_{t=1}^{T} \hat{B}'_t \left(\hat{B}_t - B_t \right) (\hat{u}_t - u_t) \tilde{F}_{t-1} \right\| = o_p(1) \tag{84}$$

yields

$$\text{vec}\left(\hat{\Lambda}_{\text{ols}}^{\text{tv}} - \Lambda \right) = -\rho_T \left[\int_0^1 \left(\Omega_f(\tau) \otimes B(\tau)' B(\tau) \right) d\tau \right]^{-1} \text{vec}(\Lambda) + \left[\int_0^1 \left(\Omega_f(\tau) \otimes B(\tau)' B(\tau) \right) d\tau \right]^{-1}$$

$$\times T^{-1}\sum_{t=1}^{T}\left(\tilde{F}_{t-1}\otimes\hat{B}_t'\right)\left[-\left(\hat{B}_t-B_t\right)\left(\Lambda\tilde{F}_{t-1}+u_t\right)+e_t-B_t(\hat{u}_t-u_t)\right]+o_p\left(T^{-1/2}\right). \quad (85)$$

The first term is $o_p\left(T^{-1/2}\right)$ under our assumptions and so we only need deal with

$$\begin{aligned}&T^{-1/2}\sum_{t=1}^{T}\hat{B}_t'\left(-\left(\hat{B}_t-B_t\right)\left(\Lambda\tilde{F}_{t-1}+u_t\right)+e_t-B_t(\hat{u}_t-u_t)\right)\tilde{F}_{t-1}'\\&=T^{-1/2}\sum_{t=1}^{T}B_t'\left(-\left(\hat{B}_t-B_t\right)\left(\Lambda\tilde{F}_{t-1}+u_t\right)+e_t-B_t(\hat{u}_t-u_t)\right)\tilde{F}_{t-1}'+o_p(1),\end{aligned} \quad (86)$$

where the equality follows because

$$\left\|T^{-1/2}\sum_{t=1}^{T}\left(\hat{B}_t-B_t\right)'\left(\hat{B}_t-B_t\right)\Lambda\tilde{F}_{t-1}\tilde{F}_{t-1}'\right\|=o_p(1), \quad (87)$$

$$\left\|T^{-1/2}\sum_{t=1}^{T}\left(\hat{B}_t-B_t\right)'\left(\hat{B}_t-B_t\right)u_t\tilde{F}_{t-1}'\right\|=o_p(1), \quad (88)$$

$$\left\|T^{-1/2}\sum_{t=1}^{T}\left(\hat{B}_t-B_t\right)'e_t\tilde{F}_{t-1}'\right\|=o_p(1), \quad (89)$$

and

$$\left\|T^{-1/2}\sum_{t=1}^{T}\left(\hat{B}_t-B_t\right)'B_t(\hat{u}_t-u_t)\tilde{F}_{t-1}'\right\|=o_p(1), \quad (90)$$

under our assumptions. Eqs. (87), (88), and (90) follow by similar steps as in Eq. (81) and Eq. (82). Equation (89) is

$$\begin{aligned}&T^{-1/2}\sum_{t=1}^{T}\left(\hat{B}_t-B_t\right)'e_t\tilde{F}_{t-1}'\\&=T^{-1/2}\sum_{t=1}^{T}D_B'\left(\hat{A}_t-A_t\right)'e_t\tilde{F}_{t-1}'\\&=T^{-3/2}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathcal{K}_h\left(\frac{s-t}{T}\right)\cdot D_B'\Omega_{z,t}^{-1}\{z_sz_s'(A_s-A_t)'+z_se_s'\}e_t\tilde{F}_{t-1}'+o_p(1),\end{aligned} \quad (91)$$

where D_B is the $(K+1+K_C)\times K_C$ matrix, which satisfies $A_tD_B=B_t$ and the second equality follows by Lemma D.1. To find the order of this term, we follow similar steps as in Ang and Kristensen (2009). Thus, we need to find the order of two terms:

$$T^{-3/2}\sum_{t=1}^{T}\mathcal{K}_h(0)\cdot D_B'\Omega_{z,t}^{-1}z_te_t'e_t\tilde{F}_{t-1}' \quad (92)$$

and

$$T^{-3/2}\sum_{s\neq t}\mathcal{K}_h\left(\frac{s-t}{T}\right)\cdot D_B'\Omega_{z,t}^{-1}\{z_sz_s'(A_s-A_t)'+z_se_s'\}e_t\tilde{F}_{t-1}'. \quad (93)$$

For Eq. (92), note that

$$\begin{aligned}&\left\|T^{-3/2}\sum_{t=1}^{T}\mathcal{K}_h(0)\cdot D_B'\Omega_{z,t}^{-1}z_te_t'e_t\tilde{F}_{t-1}'\right\|\le C\cdot T^{-3/2}\sum_{t=1}^{T}|\mathcal{K}_h(0)|\cdot\left\|z_te_t'e_t\tilde{F}_{t-1}'\right\|\\&\le C\cdot T^{-1/2}h^{-1}\cdot T^{-1}\sum_{t=1}^{T}\|z_t\|^2\|e_t\|^2\\&=O_p\left(T^{-1/2}h^{-1}\right),\end{aligned} \quad (94)$$

which is $o_p(1)$ under our assumptions. For Eq. (93) note that

$$2\cdot T^{-3/2}\sum_{s<t}\mathcal{K}_h\left(\frac{s-t}{T}\right)\cdot D_B'\Omega_{z,t}^{-1}\{z_sz_s'(A_s-A_t)'+z_se_s'\}e_t\tilde{F}_{t-1}'=T^{-3/2}\sum_{s<t}\phi_{0,T}(Y_s,Y_t), \quad (95)$$

where

$$\phi_{0,T}(Y_s,Y_t)=\varphi_{0,T}(Y_s,Y_t)+\varphi_{0,T}(Y_t,Y_s) \quad (96)$$

and

$$\varphi_{0,T}(Y_s, Y_t) = \mathcal{K}_h\left(\frac{s-t}{T}\right) \cdot D_B' \Omega_{z,t}^{-1}\{z_s z_s'(A_s - A_t)' + z_s e_s'\} e_t \tilde{F}_{t-1}', \tag{97}$$

where $Y_t = (e_t, z_t, \varsigma_t)$ with $\varsigma_t = \frac{t}{T} \in [0,1]$. We can, without loss of generality, proceed under the assumption that $\varsigma_t \sim_{iid} \mathcal{U}[0,1]$. Note first that $\mathbb{E}\left[\varphi_{0,T}(Y_s, Y_t)\right] = \mathrm{O}(h^r)$. Next, define $y = (e, z, \tau)$,

$$\mathbb{E}\left[\varphi_{0,T}(y, Y_t)\right] = \mathbb{E}\left[K_h(\tau - \varsigma_t) \cdot D_B' \Omega_z(\varsigma_t)^{-1}\{zz'(A(\tau) - A(\varsigma_t))' + ze'\} e_t \tilde{F}_{t-1}'\right] = 0, \tag{98}$$

and

$$\begin{aligned}\mathbb{E}\left[\varphi_{0,T}(Y_t, y)\right] &= \mathbb{E}\left[K_h(\tau - \varsigma_t) \cdot D_B' \Omega_z(\tau)^{-1}\{z_t z_t'(A(\varsigma_t) - A(\tau))' + z_t e_t'\} e \tilde{F}'\right] \\ &= e\tilde{F}' \cdot \mathrm{O}(h^r) + \mathrm{o}(h^r),\end{aligned} \tag{99}$$

so that, by Lemma D.2,

$$T^{-3/2} \sum_{s<t} \phi_{0,T}(Y_s, Y_t) = \mathrm{O_p}(h^r) + \sqrt{T} \cdot \mathfrak{R}_{0,T}. \tag{100}$$

The remainder term of the Hoeffding decomposition is $\mathfrak{R}_{0,T} = \mathrm{O_p}\left(T^{-1+\epsilon/2} \sup_{s \neq t} \mathbb{E}\left[\|\varphi_{0,T}(Y_s, Y_t)\|^{2+\delta}\right]^{1/(2+\delta)}\right)$ and, under our assumptions,

$$\sup_{s \neq t} \mathbb{E}\left[\|\varphi_{0,T}(Y_s, Y_t)\|^{2+\delta}\right]^{1/(2+\delta)} = \mathrm{O}\left(h^{-(1+\delta)/(2+\delta)}\right). \tag{101}$$

Then we have

$$T^{-1/2} \sum_{t=1}^{T} B_t'\left(-\left(\hat{B}_t - B_t\right)\left(\Lambda \tilde{F}_{t-1} + u_t\right) + e_t - B_t(\hat{u}_t - u_t)\right)\tilde{F}_{t-1}' = \mathcal{T}_1^{\mathrm{tv}} + \mathcal{T}_2^{\mathrm{tv}} + \mathcal{T}_3^{\mathrm{tv}}, \tag{102}$$

where

$$\mathcal{T}_1^{\mathrm{tv}} = -T^{-1/2} \sum_{t=1}^{T} B_t'\left(\hat{B}_t - B_t\right)\left(\Lambda \tilde{F}_{t-1} + u_t\right)\tilde{F}_{t-1}', \tag{103}$$

$$\mathcal{T}_2^{\mathrm{tv}} = -T^{-1/2} \sum_{t=1}^{T} B_t' B_t (\hat{u}_t - u_t) \tilde{F}_{t-1}', \tag{104}$$

and

$$\mathcal{T}_3^{\mathrm{tv}} = T^{-1/2} \sum_{t=1}^{T} B_t' e_t \tilde{F}_{t-1}'. \tag{105}$$

$\mathcal{T}_3^{\mathrm{tv}}$ is already simplified, so we need only simplify $\mathcal{T}_1^{\mathrm{tv}}$ and $\mathcal{T}_2^{\mathrm{tv}}$. First consider $\mathcal{T}_1^{\mathrm{tv}}$:

$$\begin{aligned}\mathcal{T}_1^{\mathrm{tv}} &= -T^{-1/2} \sum_{t=1}^{T} B_t'\left(\hat{B}_t - B_t\right)\left(\Lambda \tilde{F}_{t-1} + u_t\right)\tilde{F}_{t-1}' \\ &= -T^{-3/2} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathcal{K}_h\left(\frac{s-t}{T}\right) \times \\ &\quad B_t'\left[(A_s - A_t) z_s z_s' + e_s z_s'\right] \Omega_{z,t}^{-1} D_B (\Lambda D_{F,z} z_t + D_u v_t) z_t' D_{F,z}' + \mathrm{o_p}\left(T^{-1/2}\right) \\ &= \mathcal{T}_{1,1}^{\mathrm{tv}} + \mathcal{T}_{1,2}^{\mathrm{tv}} + \mathrm{o_p}\left(T^{-1/2}\right),\end{aligned} \tag{106}$$

where the first equality follows by Lemma D.1, $D_{F,z}$ is the $(K_F+1) \times (K+1+K_C)$ matrix such that $D_{F,z} z_t = \tilde{F}_{t-1}$, D_u is the $K_C \times K$ matrix such that $D_u v_t = u_t$, and

$$\mathcal{T}_{1,1}^{\mathrm{tv}} = -T^{-3/2} \sum_{t=1}^{T} \mathcal{K}_h(0) \cdot B_t' e_t z_t' \Omega_{z,t}^{-1} D_B (\Lambda D_{F,z} z_t + D_u v_t) z_t' D_{F,z}' \tag{107}$$

and

$$\mathcal{T}_{1,2}^{\mathrm{tv}} = -T^{-3/2} \sum_{s \neq t} \mathcal{K}_h\left(\frac{s-t}{T}\right) \cdot B_t'\left[(A_s - A_t) z_s z_s' + e_s z_s'\right] \Omega_{z,t}^{-1} D_B (\Lambda D_{F,z} z_t + D_u v_t) z_t' D_{F,z}'. \tag{108}$$

$\|\mathcal{T}^{\mathrm{tv}}_{1,1}\| = O_p\left(h^{-1}T^{-1/2}\right)$ by similar steps as above, and $\mathcal{T}^{\mathrm{tv}}_{1,2}$ is

$$\mathcal{T}^{\mathrm{tv}}_{1,2} = T^{-3/2}\sum_{s<t}\phi_{1,T}(Y_s, Y_t), \tag{109}$$

where $\phi_{1,T}(Y_s, Y_t) = \varphi_{1,T}(Y_s, Y_t) + \varphi_{1,T}(Y_t, Y_s)$,

$$\varphi_{1,T}(Y_s, Y_t) = -\mathcal{K}_h\left(\frac{s-t}{T}\right)\cdot B_t'\left[(A_s - A_t)z_s z_s' + e_s z_s'\right]\Omega_{z,t}^{-1}D_B\left(\Lambda\tilde{F}_{t-1} + u_t\right)\tilde{F}_{t-1}', \tag{110}$$

$Y_t = (e_t, z_t, \varsigma_t)$, and $\varsigma_t = \frac{t}{T}$. Then, by Lemma D.2,

$$\mathcal{T}^{\mathrm{tv}}_{1,2} = \sqrt{T}\cdot\left[T^{-1}\sum_{t=1}^{T}\overline{\phi}_{1,T}(Y_t) + \Re_{1,T}\right] + o_p(1), \tag{111}$$

where $\overline{\phi}_{1,T}(y) = \mathbb{E}\left[\varphi_{1,T}(y, Y_t)\right] + \mathbb{E}\left[\varphi_{1,T}(Y_t, y)\right]$. We have

$$\begin{aligned}
&\mathbb{E}\left[\phi_{1,T}(y, Y_t)\right]\\
&\quad= \mathbb{E}\left[-\frac{1}{h}\mathcal{K}_h(\tau-\varsigma_t)B(t/T)'\left[(A(\tau)-A(\varsigma_t))zz' + ez'\right]\Omega_z(\varsigma_t)^{-1}D_B\Lambda D_{F,z}\Omega_z(\varsigma_t)D_{F,z}'\right]\\
&\quad= \int_0^1 -\mathcal{K}_h(\tau-\varsigma)B(\varsigma)'[(A(\tau)-A(\varsigma))zz' + ez']\Omega_z(\varsigma)^{-1}D_B\Lambda D_{F,z}\Omega_z(\varsigma)D_{F,z}'\,\mathrm{d}\varsigma\\
&\quad= \int_{(\tau-1)h^{-1}}^{\tau h^{-1}} -\mathcal{K}(\varpi)B(\tau-\varpi h)'\left[(A(\tau)-A(\tau-\varpi h))zz' + ez'\right]\Omega_z(\tau-\varpi h)^{-1}D_B\Lambda D_{F,z}\Omega_z(\tau-\varpi h)D_{F,z}'\,\mathrm{d}\varpi\\
&\quad= -B(\tau)'ez'\Omega_z(\tau)^{-1}D_B\Lambda D_{F,z}\Omega_z(\tau)D_{F,z}' + O(h^r) + o(h^r).
\end{aligned} \tag{112}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}\left[\varphi_{1,T}(Y_t, y)\right]\\
&\quad= \mathbb{E}\left[-\mathcal{K}_h(\tau-\varsigma_t)\cdot B(\tau)'(A(\varsigma_t)-A(\tau))\Omega_z(\varsigma_t)\Omega_z(\tau)^{-1}D_B(\Lambda D_{F,z}z + D_u v)z'D_{F,z}'\right]\\
&\quad= \int_0^1 -\mathcal{K}_h(\tau-\varsigma)\cdot B(\tau)'(A(\varsigma)-A(\tau))\Omega_z(\varsigma)\Omega_z(\tau)^{-1}D_B(\Lambda D_{F,z}z + D_u v)z'D_{F,z}'\,\mathrm{d}\varsigma\\
&\quad= \int_{(\tau-1)h^{-1}}^{\tau h^{-1}} -\mathcal{K}_h(\varpi)\cdot B(\tau)'(A(\tau+\varpi h)-A(\tau))\Omega_z(\tau+\varpi h)\Omega_z(\tau)^{-1}D_B(\Lambda D_{F,z}z + D_u v)z'D_{F,z}'\,\mathrm{d}\varpi\\
&\quad= O(h^r) + o(h^r).
\end{aligned} \tag{113}$$

Thus the contribution from this term is

$$\mathcal{T}^{\mathrm{tv}}_{1,2} = -T^{-1/2}\sum_{t=1}^{T}B_t'e_t z_t'\Omega_{z,t}^{-1}D_B\Lambda D_{F,z}\Omega_{z,t}D_{F,z}' + o_p(1), \tag{114}$$

because $\sqrt{T}\Re_{1,T} = o_p(1)$ under our assumptions by similar steps as for $\Re_{0,T}$. Next consider $\mathcal{T}^{\mathrm{tv}}_2$:

$$\mathcal{T}^{\mathrm{tv}}_2 = -T^{-1/2}\sum_{t=1}^{T}B_t'B_t(\hat{u}_t - u_t)\tilde{F}_{t-1}' = T^{-1/2}\sum_{t=1}^{T}B_t'B_t D_U\left(\hat{\Psi}_t - \Psi_t\right)\tilde{X}_{t-1}\tilde{X}_{t-1}'D_{F,x}, \tag{115}$$

where $D_{F,x}$ is the $(K_F+1)\times(K+1)$ matrix such that $D_{F,x}\tilde{X}_{t-1} = \tilde{F}_{t-1}$. Then by Lemma D.1,

$$\begin{aligned}
\mathcal{T}^{\mathrm{tv}}_2 &= T^{-3/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{K}_b\left(\frac{s-t}{T}\right)B_t'B_t D_U\left[(\Psi_s - \Psi_t)\tilde{X}_{s-1}\tilde{X}_{s-1}' + v_s\tilde{X}_{s-1}'\right]\Omega_{x,t}^{-1}\tilde{X}_{t-1}\tilde{X}_{t-1}'D_{F,x}' + o_p(1)\\
&= \mathcal{T}^{\mathrm{tv}}_{2,1} + \mathcal{T}^{\mathrm{tv}}_{2,2} + o_p(1),
\end{aligned} \tag{116}$$

where

$$\mathcal{T}^{\mathrm{tv}}_{2,1} = T^{-3/2}\sum_{t=1}^{T}\mathcal{K}_b(0)\cdot B_t'B_t D_U v_t\tilde{X}_{t-1}'\Omega_{x,t}^{-1}\tilde{X}_{t-1}\tilde{X}_{t-1}'D_{F,x}', \tag{117}$$

and

$$\mathcal{T}^{\mathrm{tv}}_{2,2} = T^{-3/2}\sum_{s\neq t}\mathcal{K}_b\left(\frac{s-t}{T}\right)B_t'B_t D_U\left[(\Psi_s - \Psi_t)\tilde{X}_{s-1}\tilde{X}_{s-1}' + v_s\tilde{X}_{s-1}'\right]\Omega_{x,t}^{-1}\tilde{X}_{t-1}\tilde{X}_{t-1}'D_{F,x}'. \tag{118}$$

$\|\mathcal{T}_{2,1}^{\text{tv}}\| = \mathrm{O_p}\left(T^{-1/2}b^{-1}\right)$ by similar steps as above, while

$$\mathcal{T}_{2,2}^{\text{tv}} = T^{-3/2}\sum_{s<t}\phi_{2,T}(Y_{2,s}, Y_{2,t}), \tag{119}$$

where

$$\phi_{2,T}(Y_{2,s}, Y_{2,t}) = \varphi_{2,T}(Y_{2,s}, Y_{2,t}) + \varphi_{2,T}(Y_{2,t}, Y_{2,s}), \tag{120}$$

$$\varphi_{2,T}(Y_{2,s}, Y_{2,t}) = \mathcal{K}_b\left(\frac{s-t}{T}\right)B_t'B_tD_U\left[(\Psi_s-\Psi_t)\tilde{X}_{s-1}\tilde{X}_{s-1}'+v_s\tilde{X}_{s-1}'\right]\Omega_{x,t}^{-1}\tilde{X}_{t-1}\tilde{X}_{t-1}'D_{F,x}', \tag{121}$$

and $Y_{2,t} = \left(v_t, \tilde{X}_{t-1}, \varsigma_t\right)$. Then, by Lemma D.2,

$$\mathcal{T}_{2,2}^{\text{tv}} = \sqrt{T}\cdot\left[T^{-1}\sum_{t=1}^{T}\overline{\phi}_{2,T}(Y_{2,t}) + \mathfrak{R}_{2,t}\right] + \mathrm{o_p}(1), \tag{122}$$

where $\overline{\phi}_{2,T}(y_2) = \mathbb{E}\left[\varphi_{2,T}(y_2, Y_{2,t})\right] + \mathbb{E}\left[\varphi_{2,T}(Y_{2,t}, y_2)\right]$ and $y_2 = \left(v, \tilde{X}, \tau\right)$. First,

$$\begin{aligned}\mathbb{E}\left[\varphi_{2,T}(y_2, Y_{2,t})\right] &= \mathbb{E}\left[\mathcal{K}_b(\tau-\varsigma_t)B(\varsigma_t)'B(\varsigma_t)D_U\left[(\Psi(\tau)-\Psi(\varsigma_t))\tilde{X}\tilde{X}'+v\tilde{X}'\right]\Omega_x(\varsigma_t)^{-1}\tilde{X}_{t-1}\tilde{X}_{t-1}'D_{F,x}'\right]\\ &= \int_0^1 \mathcal{K}_b(\tau-\varsigma)B(\varsigma)'B(\varsigma)D_U\left[(\Psi(\tau)-\Psi(\varsigma))\tilde{X}\tilde{X}'+v\tilde{X}'\right]D_{F,x}'\,\mathrm{d}\varsigma\\ &= \int_{(\tau-1)b^{-1}}^{\tau b^{-1}} \mathcal{K}(\varpi)B(\tau-\varpi h)'B(\tau-\varpi h)D_U\left[(\Psi(\tau)-\Psi(\tau-\varpi h))\tilde{X}\tilde{X}'+v\tilde{X}'\right]D_{F,x}'\,\mathrm{d}\varpi\\ &= B(\tau)'B(\tau)D_Uv\tilde{X}'D_{F,x}' + \mathrm{O}(b^r) + \mathrm{o}(b^r).\end{aligned} \tag{123}$$

Similarly,

$$\begin{aligned}\mathbb{E}\left[\varphi_{2,T}(Y_{2,t}, y_2)\right] &= \mathbb{E}\left[\mathcal{K}_b(\tau-\varsigma_t)B(\tau)'B(\tau)D_U\left[(\Psi(\varsigma_t)-\Psi(\tau))\tilde{X}_{t-1}\tilde{X}_{t-1}'+v_t\tilde{X}_{t-1}'\right]\Omega_x(\tau)^{-1}\tilde{X}\tilde{X}'D_{F,x}'\right]\\ &= \left[\int_0^1 \mathcal{K}_b(\tau-\varsigma)B(\tau)'B(\tau)D_U(\Psi(\varsigma)-\Psi(\tau))\Omega_x(\varsigma)\Omega_x(\tau)^{-1}\,\mathrm{d}\varsigma\right]\tilde{X}\tilde{X}'D_{F,x}'\\ &= \left[\int_{(\tau-1)b^{-1}}^{\tau b^{-1}} \mathcal{K}(\varpi)\cdot B(\tau)'B(\tau)D_U(\Psi(\tau-\varpi h)-\Psi(\tau))\Omega_x(\tau-\varpi h)\Omega_x(\tau)^{-1}\,\mathrm{d}\varpi\right]\tilde{X}\tilde{X}'D_{F,x}'\\ &= \mathrm{O}(b^r) + \mathrm{o}(b^r).\end{aligned} \tag{124}$$

Thus, the contribution from this term is

$$\mathcal{T}_{2,2}^{\text{tv}} = T^{-1/2}\sum_{t=1}^{T}B_t'B_tD_Uv_t\tilde{X}_{t-1}'D_{F,x}' + \mathrm{O}(b^r) + \mathrm{o}(b^r), \tag{125}$$

because $\sqrt{T}\,.\mathfrak{R}_{2,t} = \mathrm{o_p}(1)$ under our assumptions by similar steps as for $\mathfrak{R}_{0,t}$. Thus,

$$\begin{aligned}\sqrt{T}\,\mathrm{vec}\left(\hat{\Lambda}_{\text{ols}}^{\text{tv}} - \Lambda\right) = &\left[\int (\Omega_f(\tau)\otimes B(\tau)'B(\tau))\,\mathrm{d}\tau\right]^{-1}\\ &\times\left[T^{-1/2}\sum_{t=1}^{T}\left(D_{F,z}\left(I_{(K+1+K_C)} - \Omega_{z,t}D_{F,z}'\Lambda'D_B'\Omega_{z,t}^{-1}\right)\otimes B_t'\right)\mathrm{vec}(e_tz_t')\right.\\ &\left.+T^{-1/2}\sum_{t=1}^{T}(D_{F,z}\otimes B_t'B_tD_u)\mathrm{vec}\left(v_t\tilde{X}_{t-1}'\right)\right] + \mathrm{o_p}(1),\end{aligned} \tag{126}$$

and the result follows by Wooldridge and White (1988) and $D_{F,z}D_B = 0$.

D.2.2. Proof of Theorem 4

By Theorem 3 and similar steps as in the proof of Theorem 2, we have that

$$\sqrt{T}\left(\hat{\bar{\lambda}}_{\text{ols}}^{\text{tv}} - \bar{\lambda}\right) = \sqrt{T}\left(\hat{\Lambda}_{\text{ols}}^{\text{tv}} - \Lambda\right)\left(T^{-1}\sum_{t=1}^{T}\tilde{\mu}_{F,t}\right) + \tilde{\Lambda}_1\left(T^{-1/2}\sum_{t=1}^{T}\left(\hat{\mu}_{X,t}-\mu_{X,t}\right)\right) + \mathrm{o_p}(1), \tag{127}$$

where $\tilde{\mu}_{F,t} = \mathbb{E}\left[\tilde{F}_t\right]$ and $\mu_{X,t} = \mathbb{E}[X_t]$. By Theorem 3, we only need focus on the expression $T^{-1/2}\sum_{t=1}^{T}\left(\hat{\mu}_{X,t}-\mu_{X,t}\right)$. Recursive substitution yields that

$$T^{-1}\sum_{t=1}^{T}\mathbb{E}[X_t] = T^{-1}\sum_{t=1}^{T}\mu_t + T^{-1}\sum_{t=1}^{T}\left[\sum_{s=1}^{t-1}\left(\prod_{i=s+1}^{t}\Phi_i'\right)'\mu_s\right] + \mathrm{o_p}\left(T^{-1/2}\right), \tag{128}$$

with associated plug-in estimator, $T^{-1}\sum_{t=1}^{T}\hat{\mu}_{X,t}$. We aim to write

$$T^{-1}\sum_{t=1}^{T}\left(\hat{\mu}_{X,t}-\mu_{X,t}\right)=T^{-1}\sum_{t=1}^{T}w_t^{\Phi}\mathrm{vec}\left(\hat{\Phi}_t-\Phi_t\right)+T^{-1}\sum_{t=1}^{T}w_t^{\mu}(\hat{\mu}_t-\mu_t), \tag{129}$$

where $w_t^{\Phi}=w_t^{\Phi}(\Phi_{-t},\mu_{-t})$, $w_t^{\mu}=w_t^{\mu}(\Phi_{-t})$, and $\Phi_{-t}=(\Phi_1,\ldots,\Phi_{t-1},\Phi_{t+1},\ldots,\Phi_T)$ and similarly for μ_{-t}. We now need to find the weights w_t^{μ} and w_t^{Φ}. It is more straightforward to deal with the weights w_t^{μ},

$$T^{-1}\sum_{t=1}^{T}w_t^{\mu}(\hat{\mu}_t-\mu_t)=T^{-1}\sum_{t=1}^{T}(\hat{\mu}_t-\mu_t)+T^{-1}\sum_{t=1}^{T}\tilde{w}_t^{\mu}(\hat{\mu}_t-\mu_t), \tag{130}$$

where

$$\tilde{w}_t^{\mu}=\tilde{w}_t^{\mu}(\Phi_{-t})=\sum_{\ell_1=t+1}^{T}\left(\prod_{\ell_2=t+1}^{\ell_1}\Phi'_{\ell_2}\right)', \tag{131}$$

so that

$$w_t^{\mu}=w_t^{\mu}(\Phi_{-t})=I_K+\sum_{\ell_1=t+1}^{T}\left(\prod_{\ell_2=t+1}^{\ell_1}\Phi'_{\ell_2}\right)' \tag{132}$$

with $w_T^{\mu}=I_K$. Next we need to find w_t^{Φ}.

$$w_t^{\Phi}(\Phi_{-t},\mu_{-t})=\left(\left(\sum_{\ell_1=1}^{t-1}\left(\prod_{\ell_2=\ell_1+1}^{t-1}\Phi'_{\ell_2}\right)'\mu_{\ell_1}\right)'\otimes\left(I_K+\sum_{\ell_1=t+1}^{T}\left(\prod_{\ell_2=t+1}^{\ell_1}\Phi'_{\ell_2}\right)'\right)\right), \tag{133}$$

with $w_1^{\Phi}=0$. Let

$$w_t=w_t(\Phi_{-t},\mu_{-t})=\left[w_t^{\mu}(\Phi_{-t})\ \ \vdots\ \ w_t^{\Phi}(\Phi_{-t},\mu_{-t})\right]. \tag{134}$$

Then by repeated applications of Lemma D.1, we have that

$$T^{-1/2}\sum_{t=1}^{T}\left(\hat{\mu}_{X,t}-\mu_{X,t}\right)=T^{-1/2}\sum_{t=1}^{T}w_t\mathrm{vec}\left(\hat{\Psi}_t-\Psi_t\right)+o_p(1). \tag{135}$$

By an additional application of Lemma D.1,

$$T^{-1/2}\sum_{t=1}^{T}w_t\mathrm{vec}\left(\hat{\Psi}_t-\Psi_t\right)=T^{-3/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{K}_b\left(\frac{s-t}{T}\right)w_t\mathrm{vec}\left(\left[(\Psi_s-\Psi_t)\tilde{X}_{s-1}\tilde{X}'_{s-1}+v_s\tilde{X}'_{s-1}\right]\Omega_{x,t}^{-1}\right)+o_p(1), \tag{136}$$

under our assumptions. Let $Y_{4,t}=\left(v_t,\tilde{X}_{t-1},\varsigma_1,\ldots,\varsigma_T\right)$ and, following steps in the proof of Theorem 3,

$$T^{-1/2}\sum_{t=1}^{T}w_t\mathrm{vec}\left(\hat{\Psi}_t-\Psi_t\right)=\mathcal{T}_{4,1}^{\mathrm{tv}}+\mathcal{T}_{4,2}^{\mathrm{tv}}+o_p(1), \tag{137}$$

where

$$\mathcal{T}_{4,1}^{\mathrm{tv}}=T^{-3/2}\sum_{t=1}^{T}\mathcal{K}_b(0)w_t\mathrm{vec}\left(v_s\tilde{X}'_{s-1}\Omega_{x,t}^{-1}\right) \tag{138}$$

and

$$\mathcal{T}_{4,2}^{\mathrm{tv}}=T^{-3/2}\sum_{s<t}\phi_{4,T}(Y_{4,s},Y_{4,t}), \tag{139}$$

where $\phi_{4,T}(Y_{4,s},Y_{4,t})=\varphi_{4,T}(Y_{4,s},Y_{4,t})+\varphi_{4,T}(Y_{4,t},Y_{4,s})$ and

$$\varphi_{4,T}(Y_{4,s},Y_{4,t})=\mathcal{K}_b\left(\frac{s-t}{T}\right)w_t\mathrm{vec}\left(\left[(\Psi_s-\Psi_t)\tilde{X}_{s-1}\tilde{X}'_{s-1}+v_s\tilde{X}'_{s-1}\right]\Omega_{x,t}^{-1}\right). \tag{140}$$

$\|\mathcal{T}_{4,1}^{\mathrm{tv}}\|=o_p(1)$ under our assumptions by similar steps as in the proof of Theorem 3. By Lemma D.2,

$$\mathcal{T}_{4,2}^{\mathrm{tv}}=\sqrt{T}\cdot\left[T^{-1}\sum_{t=1}^{T}\overline{\phi}_{4,T}(Y_{4,t})+\mathfrak{R}_{4,T}\right]+o_p(1), \tag{141}$$

where $\overline{\phi}_{4,T}(Y_{4,t})=\mathbb{E}\left[\varphi_{4,T}(y_4,Y_{4,t})\right]+\mathbb{E}\left[\varphi_{4,T}(Y_{4,t},y_4)\right]$ and $y_4=\left(v,\tilde{X},\tau_1,\ldots,\tau_T\right)$. If we define ς_{-t} to be $(\varsigma_1,\ldots,\varsigma_T)$ excluding ς_t and similarly for τ_{-t}, then

$$\mathbb{E}\left[\varphi_{4,T}(y_4,Y_{4,t})\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathcal{K}_b(\tau_t-\varsigma_t)w_t(\tau_{-t})\text{vec}\left(\left[(\Psi(\tau_t)-\Psi(\varsigma_t))\tilde{X}\tilde{X}'+v\tilde{X}'\right]\Omega_x(\varsigma_t)^{-1}\right)\right] \\
&= \int_0^1 \cdots \int_0^1 \mathcal{K}_b(\tau-\varsigma_t)w_t(\varsigma_{-t})\text{vec}\left(\left[(\Psi(\tau)-\Psi(\varsigma_t))\tilde{X}\tilde{X}'+v\tilde{X}'\right]\Omega_x(\varsigma_t)^{-1}\right)\,\mathrm{d}\varsigma_t\,\mathrm{d}\varsigma_{-t} \\
&= \left[\int_0^1 \cdots \int_0^1 w_t(\varsigma_{-t})\,\mathrm{d}\varsigma_{-t}\right]\int_0^1 \mathcal{K}_b(\tau-\varsigma_t)\text{vec}\left(\left[(\Psi(\tau)-\Psi(\varsigma_t))\tilde{X}\tilde{X}'+v\tilde{X}'\right]\Omega_x(\varsigma_t)^{-1}\right)\,\mathrm{d}\varsigma_t \\
&= \left[\int_0^1 \cdots \int_0^1 w_t(\varsigma_{-t})\,\mathrm{d}\varsigma_{-t}\right]\int_{(\tau-1)h^{-1}}^{\tau h^{-1}} \mathcal{K}(\varpi)\text{vec}\left(\left[(\Psi(\tau)-\Psi(\tau-\varpi h))\tilde{X}\tilde{X}'+v\tilde{X}'\right]\Omega_x(\tau-\varpi h)^{-1}\right)\,\mathrm{d}\varpi \\
&= \left[\int_0^1 \cdots \int_0^1 w_t(\varsigma_{-t})\,\mathrm{d}\varsigma_{-t}\right]\text{vec}\left(v\tilde{X}'\Omega_x(\tau)^{-1}\right)+\mathrm{O}(b^r)+\mathrm{o}(b^r).
\end{aligned} \tag{142}
$$

Similarly, it can be shown that

$$\mathbb{E}\left[\varphi_{4,T}(Y_{4,t},y_4)\right]=\mathrm{O}(b^r)+\mathrm{o}(b^r) \tag{143}$$

and that $\sqrt{T}\cdot\mathfrak{R}_{4,T}=\mathrm{o}_\mathrm{p}(1)$ under our assumptions, so that

$$\mathcal{T}_{4,2}^{\text{tv}}=T^{-1/2}\sum_{t=1}^{T}\overline{w}_t\text{vec}\left(v_t\tilde{X}_{t-1}'\Omega_{x,t}^{-1}\right)+\mathrm{o}_p(1),\quad \overline{w}_t=\int_0^1\cdots\int_0^1 w_t(\varsigma_{-t})\,\mathrm{d}\varsigma_{-t}. \tag{144}$$

$\overline{w}_t$ is a function of $\overline{\Phi}=\int_0^1\Phi(\tau)\,\mathrm{d}\tau$ and $\overline{\mu}=\int_0^1\mu(\tau)\,\mathrm{d}\tau$ and from Eqs. (132) and (133) we have that

$$\overline{w}_t^{\mu}=I_K+\sum_{\ell_1=t+1}^{T}\left(\prod_{\ell_2=t+1}^{\ell_1}\overline{\Phi}_{\ell_2}'\right)'=\sum_{m=0}^{T-t}\overline{\Phi}^m, \tag{145}$$

with $\overline{w}_T^{\mu}=I_K$ and

$$
\begin{aligned}
w_t^{\Phi}(\Phi_{-t},\mu_{-t})&=\left(\left(\sum_{\ell_1=1}^{t-1}\left(\prod_{\ell_2=\ell_1+1}^{t-1}\overline{\Phi}_{\ell_2}'\right)'\overline{\mu}_{\ell_1}\right)'\otimes\left(I_K+\sum_{\ell_1=t+1}^{T}\left(\prod_{\ell_2=t+1}^{\ell_1}\overline{\Phi}_{\ell_2}'\right)'\right)\right) \\
&=\left(\left(\sum_{\ell_1=1}^{t-1}\overline{\Phi}^{(t-\ell_1-1)}\overline{\mu}\right)'\otimes\left(I_K+\sum_{\ell_1=t+1}^{T}\overline{\Phi}^{(\ell_1-t)}\right)\right)
\end{aligned} \tag{146}
$$

with $\overline{w}_1^{\Phi}=0$. Thus, we have that

$$
\begin{aligned}
&\sqrt{T}\left(\hat{\overline{\lambda}}_{\text{ols}}^{\text{tv}}-\overline{\lambda}\right) \\
&=\sqrt{T}\left(\hat{\Lambda}_{\text{ols}}^{\text{tv}}-\Lambda\right)\left(T^{-1}\sum_{t=1}^{T}\tilde{\mu}_{F,t}\right)+\tilde{\Lambda}_1\left(T^{-1/2}\sum_{t=1}^{T}\overline{w}_t\left(\Omega_{x,t}^{-1}\otimes I_K\right)\text{vec}\left(v_t\tilde{x}_{t-1}'\right)\right)+\mathrm{o}_\mathrm{p}(1),
\end{aligned} \tag{147}
$$

and so using Theorem 3 the asymptotic variance is

$$
\begin{aligned}
\mathcal{V}_{\overline{\lambda}}^{\text{tv}}=&\left(\left(\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\tilde{\mu}_{F,t}\right)'\otimes I_{K_C}\right)\mathcal{V}_{\Lambda}^{\text{tv}}\left(\left(\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\tilde{\mu}_{F,t}\right)'\otimes I_{K_C}\right)' \\
&+\tilde{\Lambda}_1\left(\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\overline{w}_t\left(\Omega_{x,t}^{-1}\otimes\Sigma_{v,t}\right)\overline{w}_t'\right)\tilde{\Lambda}_1'+\mathcal{C}_{\overline{\lambda}}^{\text{tv}}+\mathcal{C}_{\overline{\lambda}}^{\text{tv}\prime},
\end{aligned} \tag{148}
$$

where

$$
\begin{aligned}
\mathcal{C}_{\overline{\lambda}}^{\text{tv}}=&\left[\left(\left(\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\tilde{\mu}_{F,t}\right)'\otimes I_{K_C}\right)\right]\left[\int_0^1(\Omega_f(\tau)\otimes B(\tau)'B(\tau))\,\mathrm{d}\tau\right]^{-1} \\
&\times\left[\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}(D_{F,x}\otimes B_t'B_tD_u\Sigma_{v,t})\overline{w}_t'\right]\tilde{\Lambda}_1'.
\end{aligned} \tag{149}
$$

References

Abrahams, M.G., Adrian, T., Crump, R.K., Moench, E., 2014. Decomposing real and nominal yield curves. Staff Report 570, Federal Reserve Bank of New York, New York, NY.

Adrian, T., Crump, R.K., Moench, E., 2013. Pricing the term structure with linear regressions. Journal of Financial Economics 110, 110–138.

Adrian, T., Moench, E., 2008. Pricing the term structure with linear regressions. Staff Report 340, Federal Reserve Bank of New York, New York, NY.

Ahn, S.C., Gadarowski, C., 1999. Two-pass cross-sectional regression of factor pricing models: minimum distance approach. Working paper.

Andrews, D.W.K., Lu, B., 2001. Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. Journal of Econometrics 101, 123–164.

Ang, A., Kristensen, D., 2009. Testing conditional factor models. Working paper.

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

Ang, A., Liu, J., Schwarz, K., 2010. Using stocks or portfolios in tests of factor models. Working paper.

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745–787.

Ang, A., Ulrich, M., 2012. Nominal bonds, real bonds, and equity. Working paper.

Baele, L., Bekaert, G., Inghelbrecht, K., 2010. The determinants of stock and bond return comovements. Review of Financial Studies 23, 2374–2428.

Balduzzi, P., Robotti, C., 2010. Asset pricing models and economic risk premia: a decomposition. Journal of Empirical Finance 17, 54–80.

Bekaert, G., Engstrom, E., Grenadier, S., 2010. Stock and bond returns with moody investors. Journal of Empirical Finance 17, 867–894.

Braun, P.A., Nelson, D.B., Sunier, A.M., 1995. Good news, bad news, volatility, and betas. Journal of Finance 50, 1575–1603.

Burnside, A.C., 2011. The cross-section of foreign currency risk premia and consumption growth risk: comment. American Economic Review 101, 3456–3476.

Burnside, C., 2010. Identification and inference in linear stochastic discount factor models. NBER Working paper 16634. National Bureau of Economic Research, Inc.

Campbell, J.Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.

Campbell, J.Y., 1996. Understanding risk and return. Journal of Political Economy 104, 572–621.

Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ.

Campbell, J.Y., Shiller, R.J., 1988. The dividend–price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–228.

Campbell, J.Y., Shiller, R.J., 1991. Yield spreads and interest rate movements: a bird's eye view. Review of Economic Studies 58, 495–514.

Campbell, J.Y., Sunderam, A., Viceira, L.M., 2013. Inflation bets or deflation hedges? the changing risks of nominal bonds. Working paper 09-088. Harvard Business School, Cambridge, MA.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: can anything beat the historical average?. Review of Financial Studies 21, 1509–1531.

Chen, N.-F., Roll, R., Ross, S.A., 1986. Economic forces and the stock market. Journal of Business 59, 383–403.

Chordia, T., Goyal, A., Shanken, J., 2013. Cross-sectional asset pricing with individual stocks: Betas versus characteristics. Working paper.

Cochrane, J., 2005. Asset Pricing, revised edition Princeton University Press, Princeton, NJ.

Cochrane, J., 2011. Discount rates. Journal of Finance 66, 1047–1109.

Cochrane, J., Piazzesi, M., 2008. Decomposing the yield curve. Working paper.

Cochrane, J.H., 2008. The dog that did not bark: a defense of return predictability. Review of Financial Studies 21, 1533–1575.

Dai, Q., Singleton, K., 2002. Expectations puzzles, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics 63, 415–441.

David, A., Veronesi, P., 2013. What ties return volatilities to price valuations and fundamentals? Journal of Political Economy 121, 682–746.

Duffee, G.R., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405–443.

Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23–49.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.

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

Fama, E.F., French, K.R., 1997. Industry costs of equity. Journal of Financial Economics 43, 153–193.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 113, 607–636.

Ferson, W.E., Harvey, C.R., 1991. The variation of economic risk premiums. Journal of Political Economy 99, 385–415.

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

Gagliardini, P., Ossola, E., Scaillet, O., 2014. Time-varying risk premium in large cross-sectional equity datasets. Working paper.

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

Gibbons, M.R., 1982. Multivariate tests of financial models: a new approach. Journal of Financial Economics 10, 3–27.

Gibbons, M.R., Ferson, W., 1985. Testing asset pricing models with changing expectations and an unobservable market portfolio. Journal of Financial Economics 14, 217–236.

Gomes, J., Kogan, L., Zhang, L., 2003. Equilibrium cross section of returns. Journal of Political Economy 111, 693–732.

Gospodinov, N., Kan, R., Robotti, C., 2012. Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. Review of Financial Studies 27, 2097–2138.

Hansen, L.-P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Harvey, C.R., 1989. Time-varying conditional covariances in tests of asset pricing models. Journal of Financial Economics 24, 289–317.

Harvey, C.R., 1991. The world price of covariance risk. Journal of Finance 46, 111–157.

Harvey, C.R., 2001. The specification of conditional expectations. Journal of Empirical Finance 8, 573–637.

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

Jagannathan, R., Wang, Z., 1998. An asymptotic theory for estimating beta-pricing models using cross-sectional regression. Journal of Finance 53, 1285–1309.

Joslin, S., Priebsch, M., Singleton, K.J., 2012. Risk premiums in dynamic term structure models with unspanned macro risks. Journal of Finance 69, 1197–1233.

Kan, R., Chen, R., 2005. Finite sample analysis of two-pass cross-sectional regressions. Working paper.

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

Kandel, S., 1984. The likelihood ratio test statistic of mean–variance efficiency without a riskless asset. Journal of Financial Economics 13, 575–592.

Keim, D.B., Stambaugh, R.F., 1986. Predicting returns in the stock and bond markets. Journal of Financial Economics 17, 357–390.

Kleibergen, F., 1998. Reduced rank regression using GMM. In: Matyas, L. (Ed.), Generalised Method of Moments Estimation, Cambridge University Press, Cambridge, UK, pp. 171–209.

Kleibergen, F., 2009. Tests of risk premia in linear factor models. Journal of Econometrics 149, 149–173.

Kleibergen, F., Zhan, Z., 2013. Unexplained factors and their effects on second pass r-squared's and t-tests. Working paper. Brown University, Providence, RI.

Koijen, R.S.J., Lustig, H.N., van Nieuwerburgh, S., 2013. The cross-section and time-series of stock and bond returns. Working paper.

Kristensen, D., 2009. Uniform convergence rates of kernel estimators with heterogeneous dependent data. Econometric Theory 25, 1433–1445.

Kristensen, D., 2012. Non-parametric detection and estimation of structural change. Econometrics Journal 15, 420–461.

Lettau, M., Ludvigson, S., 2001. Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying. Journal of Political Economy 109, 1238–1287.

Lettau, M., Wachter, J., 2011. The term structures of equity and interest rates. Journal of Financial Economics 101, 90–113.

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

Lewellen, J., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset pricing tests. Journal of Financial Economics 96, 175–194.

Mamaysky, H., 2002. Market prices of risk and return predictability in a joint stock-bond pricing model. Working paper.

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

Petkova, R., 2006. Do the Fama-French factors proxy for innovations in predictive variables? Journal of Finance 61, 581–612.
Petkova, R., Zhang, L., 2005. Is value riskier than growth? Journal of Financial Economics 78, 187–202.
Robinson, P., 1989. Nonparametric estimation of time-varying parameters. In: Hackl, P. (Ed.), Statistical Analysis and Forecasting of Economic Structural Change, Springer-Verlag, Berlin, Germany, pp. 253–264.
Roll, R., 1985. A note on the geometry of Shanken's CSR T^2 test for mean/variance efficiency. Journal of Financial Economics 14, 349–357.
Roussanov, N., 2014. Composition of wealth, conditioning information, and the cross-section of stock returns. Journal of Financial Economics 111, 352–380.
Shanken, J., 1985. Multivariate tests of the zero-beta CAPM. Journal of Financial Economics 14, 327–348.
Shanken, J., 1986. Testing portfolio efficiency when the zero-beta rate is unknown: a note. Journal of Finance 41, 269–276.
Shanken, J., 1990. Intertemporal asset pricing: an empirical investigation. Journal of Econometrics 45, 99–120.
Shanken, J., 1992. On the estimation of beta-pricing models. Review of Financial Studies 5, 1–33.
Shanken, J., Zhou, G., 2007. Estimating and testing beta pricing models: alternative methods and their performance in simulations. Journal of Financial Economics 84, 40–86.
Singleton, K., 2006. Empirical Dynamic Asset Pricing. Princeton University Press, Princeton, NJ.
Wooldridge, J.M., White, H., 1988. Some invariance principles and central limit theorems for dependent heterogeneous processes. Econometric Theory 4, 210–230.
Zhou, G., 1994. Analytical GMM tests: asset pricing with time-varying risk premiums. Review of Financial Studies 7, 687–709.