Federal Reserve Bank of New York
Staff Reports

Efficient, Regression-Based Estimation of Dynamic Asset Pricing Models

Tobias Adrian
Richard K. Crump
Emanuel Moench

Staff Report no. 493
May 2011
Revised March 2012

Efficient, Regression-Based Estimation of Dynamic Asset Pricing Models
Tobias Adrian, Richard K. Crump, and Emanuel Moench
Federal Reserve Bank of New York Staff Reports, no. 493
May 2011; revised March 2012
JEL classification: E58, G28

Abstract

We study regression-based estimators for beta representations of dynamic asset pricing models with an affine pricing kernel specification. These estimators extend static cross-sectional asset pricing estimators to settings where prices of risk vary with observed state variables. We identify conditions under which four-stage regression-based estimators are efficient and also present alternative, closed-form linearized maximum likelihood (LML) estimators. We provide multi-stage standard errors necessary to conduct inference for asset pricing tests. In empirical applications, we find that time-varying prices of risk are pervasive, thus favoring dynamic cross-sectional asset pricing models over standard unconditional specifications.

Key words: dynamic asset pricing, Fama-MacBeth regressions, financial econometrics

---

Adrian, Crump, Moench: Federal Reserve Bank of New York (e-mail: tobias.adrian@ny.frb.org, richard.crump@ny.frb.org, emanuel.moench@ny.frb.org). The authors thank Matias Cattaneo, Darrell Duffie, Robert Engle, Arturo Estrella, Eric Ghysels, Monika Piazzesi, Michael Sockin, Jonathan Wright, and seminar participants at the Federal Reserve Bank of New York for helpful comments and discussions. Daniel Green and Ariel Zucker provided excellent research assistance.

1 Introduction

This paper proposes efficient, regression based estimators for dynamic asset pricing models with time varying prices of risk. The estimators and associated standard errors that we propose are computationally as simple as commonly used static cross sectional asset pricing regressions, yet explicitly provide estimates of time varying prices of risk, as well as estimates of the associated state variable dynamics. Our approach thus allows computationally efficient and robust estimation of dynamic asset pricing models. We present estimation results for a typical asset pricing application that allows the comparison of various estimators and provides evidence that dynamic price of risk specifications are highly significant relative to constant price of risk specifications. In addition, our dynamic asset pricing approach can be used to improve inference for predictive regressions due to the presence of cross sectional constraints.

Throughout the paper, we assume that prices of risk are affine functions of lagged state variables. 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 which makes our approach particularly well suited for applications across asset classes. In this paper, we study regression based estimators for the affine pricing kernel specification[1]. We introduce a simple three stage estimator and show that it is consistent and asymptotically normal under mild conditions. The estimator can be described as follows. In the first stage, shocks to the state variables are obtained from a time series vector autoregression. In the second stage, asset returns are regressed in the time series on lagged state variables and their contemporaneous innovations, generating predictive slopes and risk betas for each test asset. In the third stage, prices of risk are obtained by running a cross sectional regression of the stacked predictive slopes onto the stacked betas. We show that steps two and three of our estimator coincide computationally with the two stage Fama and MacBeth (1973) estimator when two conditions are met. First, the state variables have to be uncorrelated across time so that the first step of the four stage estimator is not necessary. 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 next show that a fourth estimation step in which betas are recomputed by regressing asset returns on the sum of (time varying) prices of risk and state variable shocks delivers efficient estimates of betas. This is somewhat surprising as the time series estimates of betas that are conducted in the second stage are inefficient. We show that the re-estimation of betas yields efficient betas also in the Fama-MacBeth setting. The three and four stage estimators described so far are based on OLS regressions which are efficient when the variances of test assets are equal. We show that both have a straightforward GLS generalization which is efficient in more general cases.

In addition to the OLS and GLS regression based estimators, we also provide a linearized maximum likelihood (LML) estimator. The LML estimator does not reduce to a direct regression

[1] For a regression-based approach to models featuring an exponentially affine pricing kernel, see Adrian, Crump, and Moench (2012).

based approach, but is based on output from regression methods. It is therefore easily programmed and does not involve numerical optimization. We show that the LML estimator is asymptotically equivalent to the four stage GLS regression based estimator for all parameters. However, from the point of view of intuitive understanding the four stage regression based estimator is preferable.

Contributions to the 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. We preserve the simplicity of the multistep regression based asset pricing set up, but add the dynamics of the state variables and the dynamics of the prices of risk to our estimation. The empirical applications of the static Fama-MacBeth approach are too numerous to list, but some of the seminal work includes Chen, Roll, and Ross (1986) and Fama and French (1992).

The Fama-MacBeth approach has been extended to conditional asset pricing models. Ferson and Harvey (1991) use Fama-MacBeth regressions to obtain estimates of time varying market prices of risk which they then regress on lagged conditioning variables. They find strong evidence for predictable variation in prices of risk and associate most of the predictable variation in stock returns to time variation in risk compensation rather than time variation in betas. Our estimation approach generalizes the one used in Ferson and Harvey (1991) by explicitly taking the time variation of prices of risk into account in the estimation of betas. Jagannathan and Wang (1996), Lettau and Ludvigson (2001) and others use the Fama-MacBeth technology to estimate scaled factor models. The regression coefficients of such conditional asset pricing models can in principal be used to recover some of the deeper price of risk parameters that we are estimating with our fully fledged dynamic approach. However, the scaled factor approaches typically do not take the dynamic properties of the conditioning variables into account in making inference, which can potentially lead to inefficient standard errors. Furthermore, they typically do not explicitly provide estimates for the parameters governing the dynamics of prices of risk. Moreover, the beta representations of such models are nested in our more general framework.

We are adding a number of results relative to the financial econometrics literature, also in the static case with constant risk prices. The seminal work of Shanken (1992) showed that the two stage GLS estimator of the constant prices of risk is efficient under normality. We confirm Shanken's efficiency result, and show how to extend the regression based estimator to an efficient one in the case of time varying prices of risk. In addition, we show that the betas obtained by regressing returns on the time series of risk factor shocks is inefficient, but that efficient betas can be obtained by recomputing betas once prices of risk are obtained. This recomputation is efficient both in the case where prices of risk are constant and when they are time varying.

In the Fama-MacBeth framework, Jagannathan and Wang (1998) provide standard errors in a setting where factor and return innovations are not assumed to be conditionally homoskedastic. As is well known, the Gaussian assumption ensures that conditionally uncorrelated shocks are also independent. Jagannathan and Wang (1998) show that by relaxing this assumption the asymptotic standard errors have additional terms that arise due to the potential dependence of the shocks of the state variables and the asset return innovations. We also extend the results of Jagannathan and Wang (1998) to a dynamic setting.

We illustrate the usefulness of our approach and associated estimators by applying them to the conditional CAPM suggested by Lettau and Ludvigson (2001). This model implicitly assumes time variation in risk premia, but has previously been estimated using methods developed for constant price of risk specifications. We demonstrate that the beta representation of the Lettau and Ludvigson (2001) model can be obtained as a special case of our affine pricing kernel specification. Following the original paper, we estimate the Lettau and Ludvigson (2001) model for a cross section of size and book-to-market sorted equity portfolios. We compare OLS, feasible GLS, and the LML estimators. Independently of the estimator used, our results indicate that the time variation of prices of risk is highly significant, both statistically and economically. We also document that the conditional model with time varying prices of risk gives rise to economically important reductions in conditional pricing errors relative to models with constant prices of risk. Finally, even though the sample size is moderate, we find that the asymptotically efficient four stage feasible GLS and LML estimators generally provide smaller conditional pricing errors than alternative estimators.

Our paper is organized as follows. In Section 2, we introduce and discuss the class of dynamic asset pricing models used in the paper. In Section 3, we study the large sample properties for the multistage regression based OLS and GLS estimators with the affine pricing kernel specification. In Section 4, we provide the conditions under which the regression based estimator is efficient, and provide an alternative LML estimator. We provide our empirical illustration in Section 5, and Section 6 concludes. All proofs and some additional results are relegated to the Appendix.

Notation: It is convenient to introduce the following notation that will be frequently used throughout. The symbols $\otimes$ and $\odot$ represent the Kronecker and Hadamard products, respectively. Let a lower case letter denote the $\operatorname{vec}(\cdot)$ operator applied to a matrix (e.g., $\gamma = \operatorname{vec}(\Gamma)$). For an $m \times n$ matrix A, define the $mn \times mn$ commutation matrix κ_{mn} which satisfies $\operatorname{vec}(A') = \kappa_{mn} \operatorname{vec}(A)$. $\operatorname{bdiag}_{[pq],[mn]}(A_i)$ will denote a $p \times q$ block-diagonal matrix (not necessarily square) with ith diagonal element equal to the $m \times n$ matrix A_i. Finally, let I_m be the $m \times m$ identity matrix and let ι_m be a $m \times 1$ vector of ones.

2 The Affine Model

2.1 Pricing Kernel Assumptions and Return Generation

We assume that the dynamics of a $K \times 1$ vector of state variables X_t evolves according to the following vector autoregressive process:

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

with initial condition X_0. This specification can be interpreted as a discrete time analog to the state variable dynamics of Merton (1973)'s ICAPM or Cox, Ingersoll, and Ross (1985)'s general equilibrium setup. Initially, we will not necessarily assume that the shocks v_{t+1} are conditionally Gaussian, identical, or independent. Later we will introduce further stochastic assumptions. For now we only assume that:

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

where $\mathcal{F}_t$ denotes the information set at time t.

We denote holding period returns in excess of the risk free rate R_t^F of asset i by $R_{i,t+1}^e$. We assume the existence of a pricing kernel M_{t+1} such that:

$$\mathbb{E}\left[M_{t+1}R_{i,t+1}^e\middle|\mathcal{F}_t\right]=0.$$

We assume that the pricing kernel is of the following form:

$$M_{t+1}=\frac{1}{R_t^F}\left(1-\lambda_t'\Sigma_{v,t}^{-1/2}v_{t+1}\right), \tag{2}$$

where λ_t is a $K\times 1$ vector assumed to be an affine function of the state variables X_t:

$$\lambda_t=\Sigma_{v,t}^{-1/2}\left(\lambda_0+\Lambda_1 X_t\right).$$

With these two elements, we find the following beta representation of expected returns:

$$\mathbb{E}\left[R_{i,t+1}^e\middle|\mathcal{F}_t\right]=-\frac{\mathbb{C}[M_{t+1},R_{i,t+1}^e|\mathcal{F}_t]}{\mathbb{E}\left[M_{t+1}\middle|\mathcal{F}_t\right]}=\beta_{i,t}'\left(\lambda_0+\Lambda_1 X_t\right), \tag{3}$$

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

$$\beta_{i,t}=\Sigma_{v,t}^{-1}\mathbb{C}\left[X_{t+1},R_{i,t+1}^e\middle|\mathcal{F}_t\right], \tag{4}$$

and $(\lambda_0+\Lambda_1 X_t)$ is the K-dimensional vector of prices of risk. We can then decompose excess returns into an expected and an unexpected component:

$$R_{i,t+1}^e=\beta_{i,t}'\left(\lambda_0+\Lambda_1 X_t\right)+\left(R_{i,t+1}^e-\mathbb{E}\left[R_{i,t+1}^e\middle|\mathcal{F}_t\right]\right).$$

The unexpected excess return $R_{i,t+1}^e-\mathbb{E}\left[R_{i,t+1}^e|\mathcal{F}_t\right]$ can be further decomposed into a component that is conditionally correlated with the innovations of the states, $v_{t+1}=X_{t+1}-\mathbb{E}\left[X_{t+1}|\mathcal{F}_t\right]$, and a return pricing error $e_{i,t+1}$ that is conditionally orthogonal to the state innovations:

$$R_{i,t+1}^e-\mathbb{E}\left[R_{i,t+1}^e|\mathcal{F}_t\right]=\gamma_{i,t}'\left(X_{t+1}-\mathbb{E}\left[X_{t+1}|\mathcal{F}_t\right]\right)+e_{i,t+1}.$$

It is easy to show that $\gamma_{i,t}=\beta_{i,t}$ using equation (4). It then follows that excess returns are a function of lagged state variables X_t, state variable innovations v_{t+1}, and return pricing errors $e_{i,t+1}$:

$$R^e_{i,t+1} = \beta'_{i,t}\left(\lambda_0 + \Lambda_1 X_t\right) + \beta'_{i,t} v_{t+1} + e_{i,t+1}, \qquad t = 1, \ldots, T. \tag{5}$$

The excess return thus depends on the expected return, $\beta'_{i,t}\left(\lambda_0 + \Lambda_1 X_t\right)$, a component that is conditionally correlated with the innovations of the states, $\beta'_{i,t} v_{t+1}$, and a return pricing error $e_{i,t+1}$ that is conditionally orthogonal to the state innovations. Therefore, the innovations to the state variables are cross sectional pricing factors, and the levels of the states are forecasting variables. This is in line with 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. Note that a similar return generating process was studied in the context of foreign exchange return predictability in Hansen and Hodrick (1983) who estimate the model using GMM. In the equity literature, time variation in risk premia has been modeled in similar ways e.g. in Gibbons and Ferson (1985) and Campbell (1987). 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 (5) for $i = 1, .., N$ embeds the no arbitrage restrictions which were derived from the assumption about the form of the pricing kernel introduced in equation (2). Relative to a SUR model where $R^e_{i,t+1} = a_{i,t} + c_{i,t} X_t + \beta'_{i,t} v_{t+1} + e_{i,t+1}$, the assumption of no arbitrage implies $a_{i,t} = \beta'_{i,t}\lambda_0$ and $c_{i,t} = \beta'_{i,t}\Lambda_1$. For fixed t, these cross equation constraints reduce the number of parameters to be estimated by $(N - K)(K + 1)$.

The standard static cross sectional asset pricing model as reviewed by e.g., Campbell, Lo, and MacKinley (1997) and Cochrane (2005) makes two additional assumptions: $\Lambda_1 = 0$ in equation (5), and $\Phi = 0$ in equation (1). We will consider these special cases in the following sections. However, the main contribution of this paper is to study the dynamic case where $\Phi \neq 0$ and $\Lambda_1 \neq 0$.

While the focus of this paper is the estimation of the beta representation of dynamic asset pricing models, there is an extensive literature that estimates the SDF representation using GMM. In that literature, the expression $\mathbb{E}\left[M_{t+1} R^e_{i,t+1} \middle| \mathcal{F}_t \right] = 0$ is estimated directly (see Harvey (1989) and Harvey (1991)). Singleton (2006) provides an overview of dynamic asset pricing estimators, and Nagel and Singleton (2010) provide a GMM estimator with an optimal weighting matrix.

2.2 Assumptions and Further Notation

In order to analyze the estimation of model (5), we introduce the following assumptions.

Assumption 1 *(a) We observe $\{X_t\}_{t=1}^T$ generated by equation (1) where $X_0 = x_0$ is fixed; (b) $\beta_{i,t} = \beta_i$ for all t and the matrix $B = [\beta_1 \cdots \beta_N]$ has full row rank; (c) All eigenvalues of Φ have modulus less than one.*

Assumption 2 *Put $\varepsilon_t = (v'_t, e'_t)'$ and define $\mathcal{G}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$. (a) We have $\mathbb{E}[\varepsilon_t | \mathcal{G}_{t-1}] = 0$, $\mathbb{E}[\varepsilon_t \varepsilon'_t | \mathcal{G}_{t-1}] = \Sigma$, a block diagonal matrix with $(1,1)$-element Σ_v and $(2,2)$-element Σ_e; (b) Σ_v is a* *positive definite matrix and* Σ_e *is both positive definite and diagonal; (c) We have* $\mathbb{E}[v_t v_t^{'}|e_t, \mathcal{G}_{t-1}] = \Sigma_v$ *and* $\mathbb{E}[e_t e_t^{'}|v_t, \mathcal{G}_{t-1}] = \Sigma_e$*; (d)* $\sup_t \mathbb{E}[\|\varepsilon_t\|^4] < \infty$ *where* $\|\cdot\|$ *is the Euclidean norm.*

Assumption 1 (a) states that the factors are observable. In typical pricing applications, such factors could be macroeconomic variables, aggregate valuation and accounting ratios, or yields. Assumption 1 (b) states that the risk exposures are the same across time. In principle, our setup could be extended to allow for time varying conditional betas which we leave for future research. The requirement that B has full row rank of Assumption 1 (b) ensures that we can identify Λ. This requirement is equivalent to the statement that the K columns of the matrix B' are linearly independent vectors. Intuitively, we are assuming away the presence of redundant, uninformative or unspanned factors. Finally, Assumption 1 (c) states 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 premia. From a statistical point of view, the assumption means that we may avoid non-standard asymptotic arguments. Relaxation of this assumption is beyond the scope of the paper and is left for future work.

Assumption 2 facilitates the asymptotic results presented in the next section. Assumption 2 (a) and (b) characterize the disturbance terms as a joint martingale difference sequence with associated variance matrices and zero contemporaneous correlation matrix. The diagonality in Assumption 2 (b) is not vital to our results and may be easily relaxed. We maintain the assumption as it is implied by the factor structure of our model. Assumption 2 (c) is a conditional homoskedasticity assumption that is analogous to Assumption 1 in Shanken (1992). Later in the paper we discuss how our results change by accommodating conditional heteroskedasticity (i.e., analogous to Assumption 1 in Jagannathan and Wang (1998)). Finally, Assumption 2 (d) ensures that the requisite moments exist for the appropriate central limit theorem to hold.

It will be notationally convenient to introduce the matrix versions of equations (1) and (5),

$$R_e = B'\Lambda Z_- + B'V + E \tag{6}$$

$$X = \Psi Z_- + V, \tag{7}$$

where $\Psi = [\zeta \;\; \Phi]$, $\Lambda = [\lambda_0 \;\; \Lambda_1]$, $X = [X_1 \; X_2 \; \cdots \; X_T]$, $X_- = [X_0 \; X_1 \; \cdots \; X_{T-1}]$ and $Z_- = [\iota_T \;\; X_-']'$. R_e, E and V are matrices which are formed by stacking $R^e_{i,t}$, $e_{i,t}$ and v_t in the corresponding manner.

3 Estimation and Inference in the Affine Model

We start by studying various regression estimators of the affine model. We compare these estimators to the Fama-MacBeth approach, and discuss the special cases in which these estimators reduce to the Fama-MacBeth estimator.

3.1 Three Stage Estimation

We first provide a feasible estimation procedure for the parameters of interest in the model. In particular, we provide consistent and asymptotically normal estimators of the state variable dynamics Ψ, prices of risk Λ, and risk factor exposures B.[2]

Theorem 1 *Suppose Assumptions 1 and 2 hold and we observe R_e generated by equation (6). Denote $M_Z = I_T - Z_-' \left(Z_- Z_-'\right)^{-1} Z_-$. Then the following estimators,*

$$\hat{\Psi}_{\text{ols}} = XZ_-' \left(Z_- Z_-'\right)^{-1}, \qquad \hat{B}_{\text{ols}} = \left(XM_Z X'\right)^{-1} XM_Z R_e',$$

$$\hat{\Lambda}_{\text{ols}} = \left(\hat{B}_{\text{ols}} \hat{B}_{\text{ols}}'\right)^{-1} \hat{B}_{\text{ols}} R_e Z_-' \left(Z_- Z_-'\right)^{-1}, \qquad \hat{\Lambda}_{\text{gls}} = \left(\hat{B}_{\text{ols}} \Sigma_e^{-1} \hat{B}_{\text{ols}}'\right)^{-1} \hat{B}_{\text{ols}} \Sigma_e^{-1} R_e Z_-' \left(Z_- Z_-'\right)^{-1},$$

satisfy

$$\sqrt{T}(\hat{\psi}_{\text{ols}} - \psi) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\hat{\psi},\text{ols}}), \qquad \sqrt{T}(\hat{b}_{\text{ols}} - b) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\hat{b},\text{ols}}),$$

$$\sqrt{T}(\hat{\lambda}_{\text{ols}} - \lambda) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\hat{\lambda},\text{ols}}), \qquad \sqrt{T}(\hat{\lambda}_{\text{gls}} - \lambda) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\hat{\lambda},\text{gls}}),$$

where

$$\begin{aligned}
\mathcal{V}_{\hat{\psi},\text{ols}} &= \left(\Upsilon^{-1} \otimes \Sigma_v\right), \quad \mathcal{V}_{\hat{b},\text{ols}} = \left(\Sigma_e \otimes \Sigma_v^{-1}\right), \\
\mathcal{V}_{\hat{\lambda},\text{ols}} &= \mathcal{V}_{\hat{\psi},\text{ols}} + \left(\left(\Lambda' \Sigma_v^{-1} \Lambda + \Upsilon^{-1}\right) \otimes \left(BB'\right)^{-1} B \Sigma_e B' \left(BB'\right)^{-1}\right), \\
\mathcal{V}_{\hat{\lambda},\text{gls}} &= \mathcal{V}_{\hat{\psi},\text{ols}} + \left(\left(\Lambda' \Sigma_v^{-1} \Lambda + \Upsilon^{-1}\right) \otimes \left(B \Sigma_e^{-1} B'\right)^{-1}\right),
\end{aligned}$$

and $\Upsilon = \Upsilon\left(\Psi, \Sigma_v\right) = \text{plim}_{T\to\infty} \left(Z_- Z_-' \big/ T\right)$.

Theorem 1 can be intuitively summarized as a three stage estimator:[3]

1. The estimator $\hat{\Psi}_{\text{ols}} = XZ_-' \left(Z_- Z_-'\right)^{-1}$ is the OLS estimator of the vector autoregression (VAR) that governs the dynamics of the state variables X_t. The estimated state variable residuals are the orthogonal complement of the projection of X on Z_-, so $\hat{V}_{\text{ols}} = X - \hat{\Psi}_{\text{ols}} Z_- = XM_Z$. Finally Σ_v may be estimated via $\hat{\Sigma}_{v,\text{ols}} = \hat{V}_{\text{ols}} \hat{V}_{\text{ols}}' \big/ T$.

2. The estimator $\hat{B}_{\text{ols}} = \left(XM_Z X'\right)^{-1} XM_Z R_e' = \left(\hat{V}_{\text{ols}} \hat{V}_{\text{ols}}'\right)^{-1} \hat{V}_{\text{ols}} R_e'$ is the (transpose of the) OLS seemingly unrelated regression (OLS-SUR) estimator of the excess returns across assets

[2] In the discussion of the affine model we treat Σ_e as known as a matter of expository and notational convenience. For instructions on how to implement a feasible GLS estimator see Section A.1 (in the Appendix). Furthermore, in the Appendix we show that this feasible GLS estimator has the same limiting distribution as the infeasible version discussed in the main text.

[3] It is with some abuse of notation that we label our estimators of Λ as "OLS" and "GLS"; however, we do so to remain consistent with standard practices.

R_e onto the estimated state variable innovations $\hat{V}_{\text{ols}}$. Since V is unobserved we cannot regress R_e directly onto V and instead replace it by the estimated residuals based on $\hat{\Psi}_{\text{ols}}$.

3. Finally, with estimators of Ψ and B, we may estimate the prices of risk Λ by a regression involving two components: first, the OLS-SUR estimator of the quantity $B'\Lambda$ in equation (6), $R_e Z_-' \left(Z_- Z_-'\right)^{-1}$; second, the estimated risk factor exposures, $\hat{B}_{\text{ols}}$. Regressing the former onto the latter yields, $\hat{\Lambda}_{\text{ols}} = \left(\hat{B}_{\text{ols}} \hat{B}_{\text{ols}}'\right)^{-1} \hat{B}_{\text{ols}} R_e Z_-' \left(Z_- Z_-'\right)^{-1}$. The GLS estimator of the prices of risk simply weights the estimated risk exposures by the (inverse of the) variances of the pricing errors so that $\hat{\Lambda}_{\text{gls}} = \left(\hat{B}_{\text{ols}} \Sigma_e^{-1} \hat{B}_{\text{ols}}'\right)^{-1} \hat{B}_{\text{ols}} \Sigma_e^{-1} R_e Z_-' \left(Z_- Z_-'\right)^{-1}$.

This three stage OLS estimator was previously studied by Adrian and Moench (2008) in an application to affine term structure models. Let us make a few further observations about the results presented in Theorem 1. First, $\mathcal{V}_{\hat{\psi},\text{ols}}$ and $\mathcal{V}_{\hat{b},\text{ols}}$ are the familiar variance formulas for the OLS-SUR estimator applied to equations (1) and (6), respectively. Notice in particular that the form of $\mathcal{V}_{\hat{b},\text{ols}}$ shows that the impact on the asymptotic variance of replacing V by an estimate is negligible. Meanwhile, the asymptotic variance formulas for the estimators of Λ require further discussion. For concreteness, let us discuss the asymptotic variance of $\hat{\Lambda}_{\text{gls}}$ (a similar intuition holds for that of $\hat{\Lambda}_{\text{ols}}$). The asymptotic variance formula for $\hat{\Lambda}_{\text{gls}}$ is comprised of three terms. The first term in the asymptotic variance formula $\mathcal{V}_{\hat{\psi},\text{ols}}$ arises because we must replace V by an estimate based on $\hat{\Psi}_{\text{ols}}$. Unlike in the case of $\hat{B}_{\text{ols}}$ the impact of replacing V by an estimate affects the form of the asymptotic variance. The second term $(\Lambda' \Sigma_v^{-1} \Lambda \otimes (B \Sigma_e^{-1} B')^{-1})$ arises because we do not observe B and so we must replace it by an estimate, namely, $\hat{B}_{\text{ols}}$. If B and V were both known only the third term, $(\Upsilon^{-1} \otimes (B \Sigma_e^{-1} B')^{-1})$, would remain in the formula for the asymptotic variance. Furthermore, if we examine the form of the asymptotic variances of $\hat{\Lambda}_{\text{ols}}$ and $\hat{\Lambda}_{\text{gls}}$ we can see that they are equivalent when Σ_e is a scalar variance matrix. Moreover, in general, $\hat{\Lambda}_{\text{gls}}$ is asymptotically efficient relative to $\hat{\Lambda}_{\text{ols}}$.[4]

3.2 Four Stage Estimation

We next extend the three stage estimator of Theorem 1 to a four stage estimator that involves the re-estimation of B. In order to motivate this four stage regression estimator of B consider the situation where Λ and V are known. Then, if we rewrite equation (6) as

$$R_e = B' \left(\Lambda Z_- + V\right) + E, \tag{8}$$

it is clear that we could estimate B by the regression of R_e on $(\Lambda Z_- + V)$. Of course, in practice, we do not observe Λ and V and so instead we replace $(\Lambda Z_- + V)$ by an estimate based on the estimators introduced in Theorem 1. These alternative estimators of B are detailed in the following:

[4] Throughout the paper comparisons between matrices will be understood to be in a positive-definite sense.

Theorem 2 *Suppose the assumptions of Theorem 1 hold. Put $\hat{V}_{\text{ols}} = X - \hat{\Psi}_{\text{ols}} Z_-$. The following estimators,*

$$\begin{aligned}
\hat{B}_{4\text{ols}} &= [(\hat{\Lambda}_{\text{ols}} Z_- + \hat{V}_{\text{ols}})(\hat{\Lambda}_{\text{ols}} Z_- + \hat{V}_{\text{ols}})']^{-1}(\hat{\Lambda}_{\text{ols}} Z_- + \hat{V}_{\text{ols}}) R_e', \\
\hat{B}_{4\text{gls}} &= [(\hat{\Lambda}_{\text{gls}} Z_- + \hat{V}_{\text{ols}})(\hat{\Lambda}_{\text{gls}} Z_- + \hat{V}_{\text{ols}})']^{-1}(\hat{\Lambda}_{\text{gls}} Z_- + \hat{V}_{\text{ols}}) R_e',
\end{aligned}$$

satisfy

$$\sqrt{T}\begin{pmatrix} \hat{b}_{4\text{ols}} - b \\ \hat{\lambda}_{\text{ols}} - \lambda \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} \mathcal{V}_{\hat{b},4\text{ols}} & \mathcal{C}_{\hat{b},\hat{\lambda},4\text{ols}} \\ \mathcal{C}'_{\hat{b},\hat{\lambda},4\text{ols}} & \mathcal{V}_{\hat{\lambda},\text{ols}} \end{bmatrix}\right), \quad \sqrt{T}\begin{pmatrix} \hat{b}_{4\text{gls}} - b \\ \hat{\lambda}_{\text{gls}} - \lambda \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} \mathcal{V}_{\hat{b},4\text{gls}} & \mathcal{C}_{\hat{b},\hat{\lambda},4\text{gls}} \\ \mathcal{C}'_{\hat{b},\hat{\lambda},4\text{gls}} & \mathcal{V}_{\hat{\lambda},\text{gls}} \end{bmatrix}\right),$$

where

$$\begin{aligned}
\mathcal{V}_{\hat{b},4\text{ols}} &= \left(\Sigma_e \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\right) + \left(B'\left(BB'\right)^{-1} B\Sigma_e B'\left(BB'\right)^{-1} B \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\Lambda\Upsilon\Lambda'\Sigma_v^{-1}\right), \\
\mathcal{V}_{\hat{b},4\text{gls}} &= \left(\Sigma_e \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\right) + \left(B'\left(B\Sigma_e^{-1}B'\right)^{-1} B \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\Lambda\Upsilon\Lambda'\Sigma_v^{-1}\right), \\
\mathcal{C}_{\hat{b},\hat{\lambda},4\text{ols}} &= -\left(B'\left(BB'\right)^{-1} B\Sigma_e B'\left(BB'\right)^{-1} \otimes \Sigma_v^{-1}\Lambda\right)\kappa_{K(K+1)}, \\
\mathcal{C}_{\hat{b},\hat{\lambda},4\text{gls}} &= -\left(B'\left(B\Sigma_e^{-1}B'\right)^{-1} \otimes \Sigma_v^{-1}\Lambda\right)\kappa_{K(K+1)}.
\end{aligned}$$

We make a few further observations about the results presented in Theorem 2. First, we discuss the intuition behind the form of the asymptotic variance of the four stage estimators of B (again, we will focus our attention on the GLS asymptotic variance formula). Suppose that Ψ is known (so that V is observed) but Λ remains unknown. Then it turns out that the asymptotic variance formula would be exactly the same. Thus, just as in the case of the OLS estimator of B, replacing V by an estimate does not affect the expression for the asymptotic variance. If we could additionally observe Λ then only the first term of the asymptotic variance formula, $(\Sigma_e \otimes [\Lambda\Upsilon\Lambda' + \Sigma_v]^{-1})$, would remain. The second term, $(B'\left(B\Sigma_e^{-1}B'\right)^{-1} B \otimes [\Lambda\Upsilon\Lambda' + \Sigma_v]^{-1}\Lambda\Upsilon\Lambda'\Sigma_v^{-1})$, arises because we must replace Λ by the estimate $\hat{\Lambda}_{\text{gls}}$.

Second, it is straightforward to show that $\mathcal{V}_{\hat{b},\text{ols}}$ is generally larger than the asymptotic variance of $\hat{B}_{4\text{gls}}$. In particular, after some algebra we see that

$$\mathcal{V}_{\hat{b},\text{ols}} - \mathcal{V}_{\hat{b},4\text{gls}} = \left(\left(\Sigma_e - B'\left(B\Sigma_e^{-1}B'\right)^{-1} B\right) \otimes \Sigma_v^{-1}\Lambda\left[\Upsilon^{-1} + \Lambda'\Sigma_v^{-1}\Lambda\right]^{-1}\Lambda'\Sigma_v^{-1}\right).$$

The right hand side matrix is positive semi-definite.[5] There are two noteworthy special cases to consider. First, when $K = N$ then Assumption 1 (b) implies that the two asymptotic variance formulas are identical (as the first term in the Kronecker product is a zero matrix). In the sequel, we will ignore this special case as it has limited empirical relevance and instead proceed under the assumption that $N > K$. Second, when $\Lambda = 0$ the two asymptotic variance formulas are again

[5] When $\mathcal{V}_{\hat{b},4\text{gls}} \neq \mathcal{V}_{\hat{b},4\text{ols}}$ the difference $\mathcal{V}_{\hat{b},\text{ols}} - \mathcal{V}_{\hat{b},4\text{ols}}$ is non-definite.

identical. To provide some intuition for this case consider that when $\Lambda = 0$ we are no longer ignoring information about the parameter B found in $B'\Lambda Z_-$ since this expression is identically zero. An alternative way to say this is that when $\Lambda = 0$ equation (8) becomes $R_e = B'V + E$ and so the analogous four stage estimator would be exactly the same formula as $\hat{B}_{\text{ols}}$. When $\Lambda \neq 0$ we are able to exploit the additional information about the parameter B contained in the term $B'\Lambda Z_-$. Consequently, in general, the four stage estimator $\hat{B}_{4\text{gls}}$ will be asymptotically efficient relative to $\hat{B}_{\text{ols}}$.

Theorem 2 implies that the re-estimation of B sharpens the estimation and inference about the quantities that are of primary interest from an economic point of view, namely, the estimation of risk premia $B'\Lambda Z_-$ and of conditional pricing errors $B'V + E$. Furthermore, we will see in the next section that this fourth step in the estimation procedure detailed in Theorem 1 leads to asymptotically efficient estimates of risk factor exposures when the error terms are normally distributed.

The re-estimation of B in the fourth step differs from the estimation in the second step only slightly. While $\hat{B}_{\text{ols}} = \left(\hat{V}_{\text{ols}}\hat{V}'_{\text{ols}}\right)^{-1}\hat{V}_{\text{ols}}R'_e$, $\hat{B}_{4\text{ols}} = [(\hat{\Lambda}_{\text{ols}}Z_- + \hat{V}_{\text{ols}})(\hat{\Lambda}_{\text{ols}}Z_- + \hat{V}_{\text{ols}})']^{-1}(\hat{\Lambda}_{\text{ols}}Z_- + \hat{V}_{\text{ols}})R'_e$. The only difference between the two estimators is the addition of the conditional risk premium $\hat{\Lambda}_{\text{ols}}Z_-$ to $\hat{V}_{\text{ols}}$ in the time series regression. This addition of the mean reduces sampling errors of B. Note that we can write

$$\hat{\Lambda}_{\text{ols}}Z_- + \hat{V}_{\text{ols}} = X - \left(\hat{\Psi}_{\text{ols}} - \hat{\Lambda}_{\text{ols}}\right)Z_-,$$

so that $\hat{\Lambda}_{\text{ols}}Z_- + \hat{V}_{\text{ols}}$ are the estimated innovations to the state variables under the risk neutral measure.

Remark 1 (Multivariate Predictive Regressions) *We provide joint convergence results for the estimators of B and Λ in Theorem 2 to facilitate inference on the quantity $B'\Lambda$.*[6] *In particular, it can be shown that $\hat{B}'_{4\text{gls}}\hat{\Lambda}_{\text{gls}}$ is, in general, a (more) efficient estimator than any other combination from the two sets of estimators: $\left(\hat{B}_{\text{ols}}, \hat{B}_{4\text{ols}}, \hat{B}_{4\text{gls}}\right)$ and $\left(\hat{\Lambda}_{\text{ols}}, \hat{\Lambda}_{\text{gls}}\right)$. We may also estimate the risk premium, $B'\Lambda$ directly. The OLS-SUR estimator of equation (6), $R_e Z'_- \left(Z_- Z'_-\right)^{-1}$, is consistent and asymptotically normal. This estimator is equivalent to equation-by-equation OLS regressions of individual asset returns on the lagged forecasting variables. Specifically, each equation is a predictive regression (in general, a multivariate predictive regression.*[7]*) The asymptotic variance of this OLS-SUR estimator is, in general, larger than the asymptotic variance of $\hat{B}'_{4\text{gls}}\hat{\Lambda}_{\text{gls}}$. Thus, there is a clear sense in which the cross sectional constraints implied by our model may be exploited to improve inference in (systems of) predictive regressions.*

[6] Section A.3 (in the Appendix) provides a generic result on the limiting variance of estimators of $B'\Lambda$ based on the (joint) limiting distribution of individual estimators of B and Λ. Consequently, asymptotic standard errors for elements of $B'\Lambda$ may be constructed using the results from Theorem 2.

[7] For a discussion of the multivariate predictive regression model and its properties see Amihud, Hurvich, and Wang (2009).

3.3 Comparison to Static Models

It is natural to compare our results to the classical Fama-MacBeth approach. The Fama-MacBeth model can be nested into our framework when the prices of risk are constant (i.e., $\Lambda_1 = 0$) and the factors are uncorrelated across time (i.e., $\Phi = 0$). In this case, the usual Fama-MacBeth estimator and the "GLS" version (see, for example, Shanken (1985), Shanken (1992)) are equal to

$$\hat{\lambda}_{0,FM,\text{ols}} = \left(\hat{B}_{\text{ols}}\hat{B}'_{\text{ols}}\right)^{-1} \hat{B}_{\text{ols}}\bar{R}_e,$$

$$\hat{\lambda}_{0,FM,\text{gls}} = \left(\hat{B}_{\text{ols}}\Sigma_e^{-1}\hat{B}'_{\text{ols}}\right)^{-1} \hat{B}_{\text{ols}}\Sigma_e^{-1}\bar{R}_e,$$

where $\bar{R}_e = R_e\iota_T/T$ is the average across rows of the matrix R_e. The form of these estimators is the same as the estimators of Λ presented in Theorem 1 with Z_- replaced by ι_T. Thus, our model may be interpreted as a dynamic version of the classical Fama-MacBeth approach. In direct agreement with Theorem 2 we may re-estimate B to construct a (more) efficient estimator than $\hat{B}_{\text{ols}}$. If we let $\tilde{V}_{\text{ols}} = X - \bar{X}\iota'_T$, $\bar{X} = X\iota_T/T$ and define

$$\hat{B}_{FM,4\text{ols}} = [(\hat{\lambda}_{0,FM,\text{ols}}\iota'_T + \tilde{V}_{\text{ols}})(\hat{\lambda}_{0,FM,\text{ols}}\iota'_T + \tilde{V}_{\text{ols}})']^{-1}(\hat{\lambda}_{0,FM,\text{ols}}\iota'_T + \tilde{V}_{\text{ols}})R'_e,$$

$$\hat{B}_{FM,4\text{gls}} = [(\hat{\lambda}_{0,FM,\text{gls}}\iota'_T + \tilde{V}_{\text{ols}})(\hat{\lambda}_{0,FM,\text{gls}}\iota'_T + \tilde{V}_{\text{ols}})']^{-1}(\hat{\lambda}_{0,FM,\text{gls}}\iota'_T + \tilde{V}_{\text{ols}})R'_e.$$

We will see in the next section that this re-estimation step, in the form of $\hat{B}_{FM,4\text{gls}}$, produces asymptotically efficient estimates of the betas.

The results of Shanken (1992), in the Fama-MacBeth setting, were extended in Jagannathan and Wang (1998). In particular Jagannathan and Wang (1998) relaxed the conditional homoskedasticity assumption made in Shanken (1992). Although we have not made any explicit distributional assumptions thus far, our conditional homoskedasticity assumption (Assumption 2 (c)) has simplified the form of our asymptotic variance formulas. If we were instead to relax this assumption (analogous to Assumption 1 of Jagannathan and Wang (1998)) and define the following asymptotic covariance matrices,

$$\begin{aligned}
\Pi_1 &= \lim_{T\to\infty} \mathbb{E}\left[\left(\text{vec}\left(T^{-1/2}EZ'_-\right)\text{vec}\left(T^{-1/2}EV'\right)\right)'\right], \\
\Pi_2 &= \lim_{T\to\infty} \mathbb{E}\left[\left(\text{vec}\left(T^{-1/2}VZ'_-\right)\text{vec}\left(T^{-1/2}EV'\right)\right)'\right],
\end{aligned}$$

then the asymptotic variance (and covariance) formulas for the estimators $\left(\hat{\Lambda}_{\text{ols}}, \hat{\Lambda}_{\text{gls}}, \hat{B}_{4\text{ols}}, \hat{B}_{4\text{gls}}\right)$ of Theorems 1 and 2 change. In particular, it can be shown using results in the Appendix that $\mathcal{V}_{\hat{\lambda},\text{gls}}$ becomes $\mathcal{V}^*_{\hat{\lambda},\text{gls}} = \mathcal{V}_{\hat{\lambda},\text{gls}} + \mathcal{C}_{\hat{\lambda},\text{gls}} + \mathcal{C}'_{\hat{\lambda},\text{gls}}$, where

$$\mathcal{C}_{\hat{\lambda},\text{gls}} = -\left[\left(\Upsilon^{-1}\otimes\left(B\Sigma_e^{-1}B'\right)^{-1}B\Sigma_e^{-1}\right)\Pi_1 + \left(\Upsilon^{-1}\otimes I_K\right)\Pi_2\right]\left(\Sigma_v^{-1}\Lambda\otimes\Sigma_e^{-1}B'\left(B\Sigma_e^{-1}B'\right)^{-1}\right),$$

and $\mathcal{V}_{\hat{\lambda},\text{ols}}$ becomes $\mathcal{V}^*_{\hat{\lambda},\text{ols}} = \mathcal{V}_{\hat{\lambda},\text{ols}} + \mathcal{C}_{\hat{\lambda},\text{ols}} + \mathcal{C}'_{\hat{\lambda},\text{ols}}$,where

$$\mathcal{C}_{\hat{\lambda},\text{ols}} = -\left[\left(\Upsilon^{-1} \otimes (BB')^{-1} B\right) \Pi_1 + \left(\Upsilon^{-1} \otimes I_K\right) \Pi_2\right] \left(\Sigma_v^{-1}\Lambda \otimes B' (BB')^{-1}\right).$$

By similar steps $\mathcal{V}_{\hat{b},4\text{gls}}$ becomes $\mathcal{V}^*_{\hat{b},4\text{gls}} = \mathcal{V}_{\hat{b},4\text{gls}} + \mathcal{C}_{\hat{b},4\text{gls}} + \mathcal{C}'_{\hat{b},4\text{gls}}$, where

$$\begin{aligned}\mathcal{C}_{\hat{b},4\text{gls}} &= \left[\left(I_N - B'\left(B\Sigma_e^{-1}B'\right)^{-1} B\Sigma_e^{-1}\right) \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1} \Lambda\right] \kappa_{N(K+1)}\Pi_1\kappa_{KN} \times \\ &\quad \left[\left(I_N \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\right) + \left(B'\left(B\Sigma_e^{-1}B'\right)^{-1} B\Sigma_e^{-1} \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1} \Lambda\Upsilon\Lambda'\Sigma_v^{-1}\right)\right].\end{aligned}$$

The result for $\hat{B}_{4\text{ols}}$ follows similarly. Thus, $\mathcal{V}^*_{\hat{\lambda},\text{ols}}$ and $\mathcal{V}^*_{\hat{\lambda},\text{gls}}$ are an extension of the standard errors by Jagannathan and Wang (1998), which are valid in the static setting, to our dynamic setting (see Remark 3 in the Appendix).[8]

3.4 Constant Prices of Risk

In some applications, prices of risk are assumed to be constant, while risk factors are obtained as residuals from a vector autoregression (see, for example, Chen, Roll, and Ross (1986), Campbell (1996) and Petkova (2006)). This case corresponds to $\Phi \neq 0$ and $\Lambda_1 = 0$.[9] Equation (6) then becomes

$$R_e^{\mu} = B'\lambda_0\iota_T' + B'V + E, \tag{9}$$

while equation (7) is unchanged. We use a superscript μ to differentiate this case. Under these assumptions, the counterpart to Theorems 1 and 2 is,

Theorem 3 *Suppose Assumptions 1 and 2 hold and we observe R_e^{μ} generated by equation (9). Put $\bar{R}_e^{\mu} = R_e^{\mu}\iota_T/T$ and $\hat{V}_{\text{ols}} = X - \hat{\Psi}_{\text{ols}}Z_-$. Then the following estimators,*

$$\hat{\lambda}^{\mu}_{0,\text{ols}} = \left(\hat{B}_{\text{ols}}\hat{B}'_{\text{ols}}\right)^{-1} \hat{B}_{\text{ols}}\bar{R}_e^{\mu}, \quad \hat{\lambda}^{\mu}_{0,\text{gls}} = \left(\hat{B}_{\text{ols}}\Sigma_e^{-1}\hat{B}'_{\text{ols}}\right)^{-1} \hat{B}_{\text{ols}}\Sigma_e^{-1}\bar{R}_e^{\mu},$$

and

$$\begin{aligned}\hat{B}^{\mu}_{4\text{ols}} &= [(\hat{\lambda}^{\mu}_{0,\text{ols}}\iota_T' + \hat{V}_{\text{ols}})(\hat{\lambda}^{\mu}_{0,\text{ols}}\iota_T' + \hat{V}_{\text{ols}})']^{-1}(\hat{\lambda}^{\mu}_{0,\text{ols}}\iota_T' + \hat{V}_{\text{ols}})R_e^{\mu\prime}, \\ \hat{B}^{\mu}_{4\text{gls}} &= [(\hat{\lambda}^{\mu}_{0,\text{gls}}\iota_T' + \hat{V}_{\text{ols}})(\hat{\lambda}^{\mu}_{0,\text{gls}}\iota_T' + \hat{V}_{\text{ols}})']^{-1}(\hat{\lambda}^{\mu}_{0,\text{gls}}\iota_T' + \hat{V}_{\text{ols}})R_e^{\mu\prime},\end{aligned}$$

[8] Note, there is no counterpart to $\mathcal{V}^*_{\hat{b},4\text{gls}}$ in Jagannathan and Wang (1998).

[9] We could also consider the special case of $\Lambda_1 \neq 0$ and $\Phi = 0$. The results would follow by similar steps and so we omit this case for brevity.

satisfy

$$\sqrt{T}\begin{pmatrix}\hat{b}^{\mu}_{4\mathrm{ols}}-b\\ \hat{\lambda}^{\mu}_{0,\mathrm{ols}}-\lambda\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(0,\begin{bmatrix}\mathcal{V}^{\mu}_{\hat{b},4\mathrm{ols}} & \mathcal{C}^{\mu}_{\hat{b},\hat{\lambda},4\mathrm{ols}}\\ \mathcal{C}^{\mu\prime}_{\hat{b},\hat{\lambda},4\mathrm{ols}} & \mathcal{V}^{\mu}_{\hat{\lambda},\mathrm{ols}}\end{bmatrix}\right),\quad \sqrt{T}\begin{pmatrix}\hat{b}_{4\mathrm{gls}}-b\\ \hat{\lambda}^{\mu}_{0,\mathrm{gls}}-\lambda\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(0,\begin{bmatrix}\mathcal{V}^{\mu}_{\hat{b},4\mathrm{gls}} & \mathcal{C}^{\mu}_{\hat{b},\hat{\lambda},4\mathrm{gls}}\\ \mathcal{C}^{\mu\prime}_{\hat{b},\hat{\lambda},4\mathrm{gls}} & \mathcal{V}^{\mu}_{\hat{\lambda},\mathrm{gls}}\end{bmatrix}\right),$$

where,

$$\begin{aligned}
\mathcal{V}^{\mu}_{\hat{b},4\mathrm{ols}} &= (\Sigma_e\otimes\left[\Sigma_v+\lambda_0\lambda_0'\right]^{-1})+(B'\left(BB'\right)^{-1}B\Sigma_eB'\left(BB'\right)^{-1}B\otimes\left[\Sigma_v+\lambda_0\lambda_0'\right]^{-1}\lambda_0\lambda_0'\Sigma_v^{-1}),\\
\mathcal{V}^{\mu}_{\hat{b},4\mathrm{gls}} &= (\Sigma_e\otimes\left[\Sigma_v+\lambda_0\lambda_0'\right]^{-1})+(B'\left(B\Sigma_e^{-1}B'\right)^{-1}B\otimes\left[\Sigma_v+\lambda_0\lambda_0'\right]^{-1}\lambda_0\lambda_0'\Sigma_v^{-1}),\\
\mathcal{C}^{\mu}_{\hat{b},\hat{\lambda},4\mathrm{ols}} &= -(B'\left(BB'\right)^{-1}B\Sigma_eB'\left(BB'\right)^{-1}\otimes\Sigma_v^{-1}\lambda_0),\\
\mathcal{C}^{\mu}_{\hat{b},\hat{\lambda},4\mathrm{gls}} &= -(B'\left(B\Sigma_e^{-1}B'\right)^{-1}\otimes\Sigma_v^{-1}\lambda_0),\\
\mathcal{V}^{\mu}_{\hat{\lambda},\mathrm{ols}} &= \Sigma_v+\left(\lambda_0'\Sigma_v^{-1}\lambda_0+1\right)\left(BB'\right)^{-1}B\Sigma_eB'\left(BB'\right)^{-1},\\
\mathcal{V}^{\mu}_{\hat{\lambda},\mathrm{gls}} &= \Sigma_v+\left(\lambda_0'\Sigma_v^{-1}\lambda_0+1\right)\left(B\Sigma_e^{-1}B'\right)^{-1}.
\end{aligned}$$

Again, we may re-estimate the parameter B to obtain $\hat{B}^{\mu}_{4\mathrm{gls}}$ which is generally more efficient than $\hat{B}_{\mathrm{ols}}$. The interpretation of each term of $\mathcal{V}^{\mu}_{\hat{b},4\mathrm{gls}}$ and $\mathcal{V}^{\mu}_{\hat{b},4\mathrm{ols}}$ is in perfect analogy with the general case. The second term reflects the need to provide an estimate of λ_0 and neither term is affected by replacing V with an estimate. Also, when $\lambda_0=0$, then all three estimators of B have the exact same limiting distribution. The last two formulas in Theorem 3 may be recognized as the so-called "Shanken correction" of Shanken (1992) for the estimators $\hat{\lambda}_{0,FM,\mathrm{ols}}$ and $\hat{\lambda}_{0,FM,\mathrm{gls}}$, respectively. When prices of risk are constant, the asymptotic variability of our proposed estimators of B and Λ are unaffected by the transition from static to dynamic state variables. This stands in stark contrast to the case when prices of risk vary over time: If we compare Theorem 3 to Theorem 2, we can see that the parameter Υ, the limiting second moment matrix of the state variables, is incorporated in all of the asymptotic variance and covariance formulas.

4 Efficiency in the Affine Model

4.1 Time Varying Prices of Risk

In order to make precise statements regarding efficiency we make the following distributional assumption.

Assumption 3 *We have that* $\varepsilon_t=(v_t',e_t')'$ $(t=1,\dots,T)$ *are i.i.d. copies of a vector* $\varepsilon\sim\mathcal{N}(0,\Sigma)$.

Under Assumptions (2) and (3) we now have that e_t and v_t are independent. Define $\theta = (\psi', b', \lambda')'$ and so we may write the log-likelihood (up to a constant) as

$$\begin{aligned}\ell(\theta; \Sigma_e, \Sigma_v) &= \frac{T}{2}\left(\log\left(\left|\Sigma_e^{-1}\right|\right) + \log\left(\left|\Sigma_v^{-1}\right|\right)\right) - \frac{1}{2}\operatorname{vec}(E)'\left(I_T \otimes \Sigma_e^{-1}\right)\operatorname{vec}(E) \\ &\quad -\frac{1}{2}\operatorname{vec}(V)'\left(I_T \otimes \Sigma_v^{-1}\right)\operatorname{vec}(V),\end{aligned}$$

where $E = E(\Psi, B, \Lambda)$ and $V = V(\Psi)$ are treated as functions of the parameters.[10] We suppress the dependence of the likelihood on the data for notational simplicity. In the following theorem we provide expressions for the score vector and the inverse of the information matrix for the affine model with respect to the parameters (ψ, b, λ). We may focus exclusively on these parameters because (ψ, b, λ) and (σ_e, σ_v) are orthogonal in the sense that the (full) information matrix is block-diagonal. As a consequence, we may work under the "as if" assumption that Σ_e and Σ_v are known without affecting our conclusions about asymptotic efficiency results.

Theorem 4 *Suppose Assumptions 1 and 3 hold and we observe R_e generated by equation (6). Then the (scaled) score vector is $\dot{\ell} = \dot{\ell}(\theta; \Sigma_e, \Sigma_v)$, the $(2K(K+1) + KN) \times 1$, partitioned vector with elements*

$$\begin{aligned}\left[\dot{\ell}\right]_1 &= \frac{\partial \ell(\theta; \Sigma_e, \Sigma_v)}{\partial \psi} = T^{-1} \cdot \left[\operatorname{vec}\left(\Sigma_v^{-1} V Z_-'\right) - \operatorname{vec}\left(B \Sigma_e^{-1} E Z_-'\right)\right] \\ \left[\dot{\ell}\right]_2 &= \frac{\partial \ell(\theta; \Sigma_e, \Sigma_v)}{\partial b} = T^{-1} \cdot \operatorname{vec}\left(\left(\Lambda Z_- + V\right) E' \Sigma_e^{-1}\right) \\ \left[\dot{\ell}\right]_3 &= \frac{\partial \ell(\theta; \Sigma_e, \Sigma_v)}{\partial \lambda} = T^{-1} \cdot \operatorname{vec}\left(B \Sigma_e^{-1} E Z_-'\right)\end{aligned}$$

Moreover, the information matrix $\mathcal{I}(\theta; \Sigma_e, \Sigma_v, \Upsilon)$, has an inverse, $\mathcal{H} = \mathcal{H}(\theta; \Sigma_e, \Sigma_v, \Upsilon)$, which is a partitioned matrix comprised of the following elements,

$$\begin{aligned}&[\mathcal{H}]_{11} = \mathcal{V}_{\hat{\psi},\text{ols}}, && [\mathcal{H}]_{22} = \mathcal{V}_{\hat{b},4\text{gls}}, && [\mathcal{H}]_{33} = \mathcal{V}_{\hat{\lambda},\text{gls}}, \\ &[\mathcal{H}]_{12} = 0_{K(K+1) \times NK}, && [\mathcal{H}]_{13} = \mathcal{V}_{\hat{\psi},\text{ols}}, && [\mathcal{H}]_{23} = \mathcal{C}_{\hat{b},\hat{\lambda},4\text{gls}}.\end{aligned}$$

The inverse of the information matrix yields the lowest attainable bound for regular estimators under Assumption 3. As a consequence, we may draw explicit conclusions regarding the efficiency properties of the estimators we have thus far proposed. The first diagonal element (the $(1,1)$-element) is exactly equal to the asymptotic variance of the OLS estimator, $\hat{\Psi}_{\text{ols}}$, given in Theorem 1. This is because $\hat{\Psi}_{\text{ols}}$ is in fact the MLE of Ψ.[11] The second and third diagonal elements confirm that $\hat{\Lambda}_{\text{gls}}$ and $\hat{B}_{4\text{gls}}$ are efficient estimators of the parameters Λ and B. Shanken (1992) shows that the two pass GLS estimator is efficient when prices of risk are constant and factors are uncorrelated

[10] The ordering of parameters (ψ, b, λ) will be followed for all derivatives of the log-likelihood for the affine model. For example, the first element of the score is $\partial\ell/\partial\psi$ and the $(1,2)$ element of the information matrix is $\lim_{T\to\infty} -T^{-1}\mathbb{E}\left[\partial\ell/\partial\psi\partial b'\right]$.

[11] However, we will show in the next section that this result no longer holds when prices of risk are constant.

across time. Thus we extend Shanken (1992)'s result to the case where prices of risk are time varying and factors follow a first order VAR. Furthermore, Theorem 4 shows that the simple, four stage regression estimator of B proposed in Theorem 2 is also asymptotically equivalent to the maximum likelihood estimator.

The asymptotic efficiency of the estimators $\hat{\Lambda}_{\text{gls}}$ and $\hat{B}_{4\text{gls}}$ is both a surprising and an appealing result. It is surprising because $\hat{\Lambda}_{\text{gls}}$ is based on a generally inefficient estimator of B, namely $\hat{B}_{\text{ols}}$ and $\hat{B}_{4\text{gls}}$ is, in turn, based on $\hat{\Lambda}_{\text{gls}}$. It is appealing, because it provides, in a clearly defined sense, optimality properties for our four stage regression procedure. These estimators can all be implemented just as simply as commonly used cross sectional asset pricing regressions.

We may also construct asymptotically efficient estimators via a linearized maximum likelihood (LML) approach.[12] We appeal to this approach because it does not appear that the maximum likelihood estimators of the parameters B and Λ are available without using numerical maximization.[13] Moreover, this approach is as simple to implement as the four stage regression estimators and at negligible computational cost since numerical maximization is rendered unnecessary.

Corollary 1 *Suppose the assumptions of Theorem 4 hold. In addition assume there exists an estimator of θ, $\bar{\theta}$ which satisfies $\sqrt{T}\left(\bar{\theta}-\theta\right) = O_p(1)$ and estimators $\bar{\Sigma}_v$ and $\bar{\Sigma}_e$ which satisfy $\bar{\Sigma}_v = \Sigma_v + o_p(1)$ and $\bar{\Sigma}_e = \Sigma_e + o_p(1)$. Put $\hat{\Upsilon} = Z_- Z_-' \big/ T$. Then the estimators formed by*

$$\hat{\theta}_{\text{lmle}} = \bar{\theta} + \mathcal{H}\left(\bar{\theta}; \bar{\Sigma}_e, \bar{\Sigma}_v, \hat{\Upsilon}\right) \times \dot{\ell}\left(\bar{\theta}; \bar{\Sigma}_e, \bar{\Sigma}_v\right)$$

satisfy

$$\sqrt{T}\left(\hat{\theta}_{\text{lmle}} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{H}\left(\theta; \Sigma_e, \Sigma_v, \Upsilon\right)\right).$$

Remark 2 *When evaluating the score vector $\dot{\ell}\left(\theta; \Sigma_e, \Sigma_v\right)$ with respect to estimators $\bar{\theta}$ one would replace V and E in the expressions given in Theorem 4 by $\bar{V} = X - \bar{\Psi} Z_-$ and $\bar{E} = R_e - \bar{B}'\bar{\Lambda} Z_- - \bar{B}'\bar{V}$, respectively.*

The LML estimators use pilot estimators of the parameters as inputs and produce asymptotically efficient estimators as outputs. An appealing choice for these pilot estimators are the OLS estimators of Theorem 1 along with the usual residual based estimators of Σ_v and Σ_e (see Step 1 in the discussion after Theorem 1 and Section A.1). They are simple to compute as they are in closed form. To provide some intuition for the procedure let us consider the LML estimator of Ψ. Recall that $\hat{\Psi}_{\text{ols}}$ is the MLE when prices of risk are time varying. Using the results from Theorem

[12] For a detailed discussion of this approach for the Fama-MacBeth setting see Shanken (1992). The method was introduced in the Econometrics literature in Rothenberg and Leenders (1964).

[13] Note, when we only assume that Σ_e is positive definite then a maximum likelihood solution is available without appealing to numerical methods (see Section A.2 in Appendix A). Also, in the special case when $K = N$ then $\hat{B}_{\text{ols}}$ and $\hat{\Lambda}_{\text{gls}}$ are the maximum likelihood estimators. Maximum likelihood estimation has been considered in the Fama-MacBeth case by Gibbons (1982), Kandel (1984), Roll (1985), Shanken (1985), Shanken (1986), Chen and Kan (2004), Shanken and Zhou (2007), and Kleibergen (2009).

4 we have,

$$\hat{\psi}_{\text{lmle}} = \hat{\psi}_{\text{ols}} + T^{-1} \cdot \left(\hat{\Upsilon}^{-1} \otimes \bar{\Sigma}_v\right) \text{vec}\left(\bar{\Sigma}_v^{-1}\left(X - \hat{\Psi}_{\text{ols}} Z_-\right) Z_-'\right).$$

However, by construction $X - \hat{\Psi}_{\text{ols}} Z_- = 0$ and so $\hat{\Psi}_{\text{lmle}} = \hat{\Psi}_{\text{ols}}$. The LML procedure returns the pilot estimator when that estimator is already the MLE. When the pilot estimator is not the MLE it returns an estimator that is asymptotically equivalent to the MLE.

4.2 Constant Prices of Risk

In this section, we discuss efficiency when prices of risk are constant. The most noteworthy result is that the standard OLS estimator of the VAR parameters is no longer the MLE. Despite this, the GLS estimators of B and λ_0 (based on $\hat{\Psi}_{\text{ols}}$) given in Theorem 3 are still asymptotically efficient. To begin we provide expressions for the score vector and the inverse of the information matrix with respect to the parameters $\theta^\mu = \left(\psi', b', \lambda_0'\right)'$ when prices of risk are constant.

Theorem 5 *Suppose Assumptions 1 and 3 hold and we observe R_e generated by equation (9). Then the (scaled) score vector is $\dot{\ell}^\mu = \dot{\ell}^\mu\left(\theta^\mu; \Sigma_v, \Sigma_e\right)$, the $K(K+N+2) \times 1$, partitioned vector, with elements*

$$\begin{aligned}
\left[\dot{\ell}^\mu\right]_1 &= \frac{\partial \ell^\mu\left(\theta^\mu; \Sigma_v, \Sigma_e\right)}{\partial \psi} = T^{-1} \cdot \left[\text{vec}\left(\Sigma_v^{-1} V Z_-'\right) - \text{vec}\left(B \Sigma_e^{-1} E^\mu Z_-'\right)\right] \\
\left[\dot{\ell}^\mu\right]_2 &= \frac{\partial \ell^\mu\left(\theta^\mu; \Sigma_v, \Sigma_e\right)}{\partial b} = T^{-1} \cdot \text{vec}\left(\left(\lambda_0 \iota_T' + V\right) E^{\mu\prime} \Sigma_e^{-1}\right) \\
\left[\dot{\ell}^\mu\right]_3 &= \frac{\partial \ell^\mu\left(\theta^\mu; \Sigma_v, \Sigma_e\right)}{\partial \lambda_0} = T^{-1} \cdot \text{vec}\left(B \Sigma_e^{-1} E^\mu \iota_T\right)
\end{aligned}$$

where $E^\mu = E^\mu\left(\Psi, B, \lambda_0\right)$ and $V = V(\Psi)$. Moreover, the information matrix $\mathcal{I}^\mu\left(\theta^\mu; \Sigma_e, \Sigma_v, \Upsilon\right)$, has an inverse, $\mathcal{H}^\mu = \mathcal{H}^\mu\left(\theta^\mu; \Sigma_e, \Sigma_v, \Upsilon\right)$, which is a partitioned matrix comprised of the following elements,

$$\begin{array}{lll}
\left[\mathcal{H}^\mu\right]_{11} = \mathcal{V}^\mu_{\hat{\psi},\text{mle}}, & \left[\mathcal{H}^\mu\right]_{22} = \mathcal{V}^\mu_{\hat{b},\text{4gls}}, & \left[\mathcal{H}^\mu\right]_{33} = \mathcal{V}^\mu_{\hat{\lambda},\text{gls}}, \\
\left[\mathcal{H}^\mu\right]_{12} = 0_{K(K+1) \times NK}, & \left[\mathcal{H}^\mu\right]_{13} = \mathcal{V}^\mu_{\hat{\psi},\text{mle}}\left(\Upsilon_1 \otimes I_K\right), & \left[\mathcal{H}^\mu\right]_{23} = \mathcal{C}^\mu_{\hat{b},\hat{\lambda},\text{4gls}},
\end{array}$$

where

$$\mathcal{V}^\mu_{\hat{\psi},\text{mle}} = \left[\left(\left(\Upsilon - \Upsilon_1 \Upsilon_1'\right) \otimes B \Sigma_e^{-1} B'\right) + \left(\Upsilon \otimes \Sigma_v^{-1}\right)\right]^{-1}$$

and Υ_1 is the first column of the matrix Υ.

As we foreshadowed earlier, the most striking result of Theorem 5 is that the first diagonal element is not the same as when prices of risk are time varying, or equivalently, that $\hat{\Psi}_{\text{ols}}$ is not the MLE, and is instead an asymptotically inefficient estimator. In fact, using parameter values that would be commonly encountered in empirical finance applications the loss in efficiency can be substantial. It is the first term, $\left(\left(\Upsilon - \Upsilon_1 \Upsilon_1'\right) \otimes B \Sigma_e^{-1} B'\right)$, a function of parameters of the return equation, B and Σ_e, which governs the degree of efficiency loss. Fortunately, this is only a concern if inference on Ψ is of interest. Otherwise we again have the appealing result that our four stage estimation procedure may be used to obtain asymptotically efficient estimators. This is again a surprising result, perhaps more so in this case, because now both $\hat{\lambda}_{\text{gls}}$ and $\hat{B}^{\mu}_{4\text{gls}}$ are directly based on inefficient estimators ($\hat{B}_{\text{ols}}$ and $\hat{\Psi}_{\text{ols}}$, respectively).

4.3 Comparison to Static Models

Shanken (1992) shows that $\hat{\lambda}_{0,FM,\text{gls}}$ is asymptotically equivalent to the MLE of λ_0 under Assumption 3. Define $\theta^{FM} = \left(\zeta', b', \lambda_0'\right)'$. It can be shown that in the Fama-MacBeth framework (under Assumption 3) the information matrix, $\mathcal{I}^{FM}\left(\theta^{FM}; \Sigma_e, \Sigma_v\right)$, has an inverse, $\mathcal{H}^{FM} = \mathcal{H}^{FM}\left(\theta^{FM}; \Sigma_e, \Sigma_v\right)$, which is a partitioned matrix comprised of the following elements,

$$\begin{aligned} \left[\mathcal{H}^{FM}\right]_{11} &= \Sigma_v, & \left[\mathcal{H}^{FM}\right]_{22} &= \mathcal{V}^{\mu}_{\hat{b},4\text{gls}}, & \left[\mathcal{H}^{FM}\right]_{33} &= \mathcal{V}^{\mu}_{\hat{\lambda},\text{gls}}, \\ \left[\mathcal{H}^{FM}\right]_{12} &= 0_{K\times NK}, & \left[\mathcal{H}^{FM}\right]_{13} &= \Sigma_v, & \left[\mathcal{H}^{FM}\right]_{23} &= \mathcal{C}^{\mu}_{\hat{b},\hat{\lambda},4\text{gls}}. \end{aligned}$$

This confirms the Shanken (1992) result and adds the additional result that the four stage estimator of B, $\hat{B}_{FM,4\text{gls}}$, and the MLE of B are asymptotically equivalent. Thus, as a special case of our results, we extend the results of Shanken (1992) to show that an asymptotically efficient, multistage regression estimator for B is also available in the Fama-MacBeth setup.

5 Empirical Application

We illustrate the estimators in an empirical application drawn from a prominent asset pricing model and document the statistical and economic significance of modeling prices of risk as time varying. We start by showing that the conditional CAPM of Lettau and Ludvigson (2001) is a special case of our affine asset pricing specification. Although the model implicitly assumes time variation in risk premia, it has previously been estimated using methods designed for constant price of risk specifications. We explicitly allow for time varying prices of risk by applying the estimators suggested above. We find strong empirical support in favor of dynamic price of risk specifications, as Λ_1 is highly significant, and as pricing errors are reduced in economically meaningful magnitudes relative to a constant price of risk specification.

Lettau and Ludvigson (2001) present evidence that conditional versions of the CAPM reduce pricing errors in the cross section of average size and book-to-market sorted stock returns relative to unconditional specifications. Lettau and Ludvigson (2001) follow Cochrane (1996) in specifying the parameters of an affine pricing kernel:

$$M_{t+1} = a_t + b_t f_{t+1},$$

$$a_t = a_0 + a_1 z_t, \qquad b_t = b_0 + b_1 z_t,$$

where M_{t+1} denotes the pricing kernel at date $t+1$, f_{t+1} is a scalar pricing factor, and z_t is a scalar conditioning variable. Lettau and Ludvigson (2001) present a conditional CAPM specification where the pricing factor is the excess return of the market portfolio, $f_{t+1} = R^M_{t+1}$ and a consumption CAPM specification with consumption growth, $f_{t+1} = \Delta c_{t+1}$, as pricing factor. We focus on the CAPM version of Lettau and Ludvigson (2001), which gives rise to the following expected return-beta representation

$$\mathbb{E}\left[R_{t+1} \middle| \mathcal{F}_t \right] - R^f_t = \beta_t \lambda_t$$

where $\beta_t = \mathbb{V}\left(R^M_{t+1} \middle| \mathcal{F}_t \right)^{-1} \mathbb{C}\left[R^M_{t+1}, R_{t+1} \middle| \mathcal{F}_t \right]$ is the factor risk exposure and

$$\lambda_t = -R^f_t \mathbb{V}\left(R^M_{t+1} \middle| \mathcal{F}_t \right) b_t = -R^f_t \mathbb{V}\left(R^M_{t+1} \middle| \mathcal{F}_t \right) (b_0 + b_1 z_t)$$

is the market price of risk. Based on previous work documenting that the log consumption-wealth ratio, measured as the cointegrating residual between consumption, total asset wealth and labor income and labeled *cay*, has predictive power for equity premia, Lettau and Ludvigson (2001) propose to use this indicator as the conditioning variable z_t. Consistent with Assumptions 1 and 2, and Lettau and Ludvigson (2001), we assume that there is no time variation in conditional second moments, i.e. $\beta_t = \beta \; \forall t$ and $\mathbb{V}\left[R^M_{t+1} \middle| \mathcal{F}_t \right] = \Sigma_M \; \forall t$. While Lettau and Ludvigson (2001) assume a constant risk free rate, we allow it to be time varying. We then obtain the expected return-beta representation

$$\mathbb{E}\left[R_{t+1} \middle| \mathcal{F}_t \right] - R^f_t = \beta_t \lambda_t,$$

where

$$\lambda_t = -\Sigma_M R^f_t \left(b_0 + b_1 z_t \right),$$

i.e., time variation in market prices of risk is due to time varying risk-free rates and time variation in the log consumption-wealth ratio scaled by the risk-free rate. In our general modeling framework, the vector of state variables thus becomes

$$X_t = \left(R^M_t, R^f_t, R^f_t \cdot cay_t \right)'.$$

In what follows, we present results based on the estimation of the pricing model (6)-(7) with X as defined above. Note that our model is more general than the specification estimated by Lettau and Ludvigson (2001), since we allow time variation in risk premia to emanate from all elements of X rather than *cay* alone. We follow Lettau and Ludvigson (2001) and assess the model's performance in pricing the cross section of 25 size and book-to-market sorted equity portfolios of Fama and French (1993).[14]

We can assess whether the prices of risk associated with the risk factors in the model feature

[14] We use the equity return data from the website of Kenneth French and *cay* from the website of Sydney Ludvigson.

time variation by testing if the rows of Λ_1 are statistically different from zero. Given the asymptotic distributions of the estimators for the affine model derived in Sections 3 and 4, this can be done using the Wald test for the null hypothesis that a given row of Λ_1 is equal to zero. In particular, let $\lambda_1^{i\prime}$ be the i-th row of Λ_1. Then, under the null that $\lambda_1^i = 0_{k\times 1}$, the Wald statistic

$$W_{\lambda_1^i} = \hat{\lambda}_1^{i\prime} \hat{\mathcal{V}}_{\lambda_1^i}^{-1} \hat{\lambda}_1^i \overset{a}{\sim} \chi^2(k) \tag{10}$$

has a chi-square distribution with k degrees of freedom. We compute these Wald statistics for the different estimators proposed in Sections 3 and 4 using as inputs point estimates $\hat{\lambda}_1^{i\prime}$ with corresponding asymptotic variance-covariance matrix $\hat{\mathcal{V}}_{\lambda_1^i}$, respectively.

Table 1 reports estimates and corresponding standard errors for the market price of risk parameters λ_0 and Λ_1 in our version of the Lettau-Ludvigson model based on three different estimators: OLS, FGLS, and LML. The last column provides the Wald statistics and corresponding p-values for tests of whether rows of Λ_1 are significantly different from zero.

There are two takeaways from Table 1. First, the upper-right element of Λ_1 corresponding to the impact of $R_f \cdot cay$ on the price of market risk is strongly significant independently of the particular estimator considered. This corroborates Lettau-Ludvigson's findings that *cay* captures time variation in the market risk premium. Second, individual elements of the remaining two rows of Λ_1 are also found to be significant across the three estimators considered. Moreover, for both the FGLS and the LML estimator, the Wald tests strongly reject the null hypothesis that rows of Λ_1 are equal to zero. This suggests that not only the market risk premium features time variation but also the prices of risk associated with the additional risk factors R_f and $R_f \cdot cay$.

When we compare the three different estimators in Table 1, we note that the point estimates for the prices of risk are very similar for the LML and FGLS estimators, while the point estimates for the OLS estimator differ marginally. This difference between the OLS and the FGLS and LML estimators is due to the weighting in the efficient estimation approach. When we compare the standard errors, we can see that the OLS standard errors are generally larger. The difference in the size of the standard errors particularly has an impact on the Wald test. We conclude that the efficient estimators allow for sharper inference.

In sum, the data clearly favor a specification with time varying prices of risk. This can also be seen in Table 2 which reports the average absolute conditional pricing errors implied by the Lettau-Ludvigson model based on the different estimators proposed in Section 3. In particular, each row of Table 2 reports the quantity $\frac{1}{T-1}\sum_{t=1}^{T-1}\left(|\hat{\alpha}_{i,t+1}| - |\hat{\alpha}_{i,t+1}^{FM}|\right)$ where

$$\hat{\alpha}_{i,t+1} = R_{i,t+1}^e - \hat{\beta}_i' \hat{\Lambda} Z_t \tag{11}$$

and $\Lambda = \lambda_0$ and $Z_t = 1$ in the constant price of risk case and $\Lambda = [\lambda_0 \;\; \Lambda_1]$ and $Z_t = [1 \;\; X_t']'$ in the time varying prices of risk specification. Meanwhile, each column reports the conditional pricing errors for the estimators OLS, FGLS, 4FGLS and LML along with their constant price of risk coun-

Table 1: Lettau-Ludvigson - Market Price of Risk Parameter Estimates

This table reports coefficient estimates and the corresponding standard errors for the market price of risk parameters λ_0 and Λ_1 in the conditional CAPM with MKT, R_t^f, and $R_t^f \cdot cay_t$ as pricing factors. Three different estimators are shown: OLS, FGLS, and LML. The last column reports Wald statistics and the corresponding p-values for the null of a respective row of Λ_1 being equal to zero. The sample period is 1952Q1-2010Q2.

	λ_0	$\Lambda_1^{R^M}$	$\Lambda_1^{R_f}$	$\Lambda_1^{R_f \cdot cay}$	W_{Λ_1}
			OLS		
R^M	**2.274**	0.055	-0.801	**0.211**	**11.176**
	(1.088)	(0.066)	(0.769)	(0.070)	(0.011)
R_f	-0.078	**-0.012**	-0.051	**0.017**	**9.310**
	(0.094)	(0.006)	(0.058)	(0.006)	(0.025)
$R_f \cdot cay$	**3.457**	**0.265**	0.689	-0.057	5.736
	(1.662)	(0.111)	(1.037)	(0.111)	(0.125)
			FGLS		
R^M	**2.476**	0.063	-0.845	**0.230**	**13.452**
	(1.079)	(0.066)	(0.765)	(0.070)	(0.004)
R_f	-0.026	-0.005	-0.070	**0.020**	**13.169**
	(0.077)	(0.005)	(0.050)	(0.006)	(0.004)
$R_f \cdot cay$	**2.992**	**0.332**	0.587	-0.142	**12.132**
	(1.400)	(0.098)	(0.914)	(0.103)	(0.007)
			LML		
R^M	**2.476**	0.063	-0.845	**0.230**	**13.466**
	(1.078)	(0.065)	(0.765)	(0.070)	(0.004)
R_f	-0.026	-0.005	-0.070	**0.020**	**13.902**
	(0.075)	(0.005)	(0.049)	(0.006)	(0.003)
$R_f \cdot cay$	**2.992**	**0.332**	0.587	-0.142	**14.195**
	(1.298)	(0.091)	(0.849)	(0.096)	(0.003)

terparts, OLS$^\mu$, FGLS$^\mu$, 4FGLS$^\mu$ and LML$^\mu$. Here, FGLS denotes the pricing errors corresponding to $\hat{B} = \hat{B}_{\text{ols}}$ and $\hat{\Lambda} = \hat{\Lambda}_{\text{fgls}}$ while 4FGLS denotes the pricing errors corresponding to $\hat{B} = \hat{B}_{4\text{fgls}}$ and $\hat{\Lambda} = \hat{\Lambda}_{\text{fgls}}$ (and similarly for FGLS$^\mu$ and 4FGLS$^\mu$). We directly compare the conditional pricing errors implied by these estimators to those of Lettau-Ludvigson's benchmark model, denoted $\hat{\alpha}_{i,t+1}^{FM}$, and estimated using the Fama-MacBeth method with factors $\tilde{X}_t = \left(R_t^M, cay_{t-1}, R_t^M \cdot cay_{t-1}\right)'$.

Based on the results in Table 2, we make the following observations. First, for most portfolios, the constant price of risk specifications with the three risk factors imply average absolute conditional pricing errors that are of similar magnitude as those given by the Fama-MacBeth estimator of Lettau-Ludvigson's CAPM with the scaled market factor. Still, on average across all 25 portfolios, the efficient $4FGLS^\mu$ and LML^μ estimators with constant prices of risk provide a 2 basis point reduction in average absolute conditional pricing errors relative to the scaled CAPM estimated via Fama-MacBeth.

Second, independently of the estimator used, the affine model with time varying prices of risk consistently yields lower conditional average absolute pricing errors than the affine model with constant prices of risk. While the difference between the specifications with constant and time varying prices of risk is relatively small for small growth stocks, it reaches more than 30 basis points per quarter for large value stocks as captured in the FF51 portfolio. On average across the 25 portfolios, the improvement due to time varying prices of risk is about 13 basis points per quarter for all four estimators considered here. Hence, there is an economically meaningful advantage of using the time varying price of risk specification of our model. This is also true when the estimators

Table 2: Lettau-Ludvigson - Mean Absolute Conditional Pricing Errors

This table reports time series averages of differences of absolute conditional pricing errors in the Lettau-Ludvigson model implied by our estimators with respect to the Fama-MacBeth estimator. Specifically, we report the quantity $\frac{1}{T-1}\sum_{t=1}^{T-1}|\hat{\alpha}_{i,t+1}| - |\hat{\alpha}^{FM}_{i,t+1}|$ where $\hat{\alpha}_{i,t+1} = R^e_{i,t+1} - \hat{\beta}_i'\hat{\Lambda}Z_t$ is the conditional pricing error implied by the respective estimator and $\hat{\alpha}^{FM}_{i,t+1}$ is the conditional pricing error implied by the Fama-MacBeth estimator applied to the scaled CAPM. The test assets are the returns of 25 size and book-to-market sorted stock portfolios provided on Ken French's website in excess of the risk-free rate also from Ken French's website. Six different estimators are considered. OLS^μ denotes the OLS estimator of the affine pricing model with constant prices of risk; OLS denotes the OLS estimator of the affine pricing model with time varying prices of risk; $FGLS^\mu$ and $FGLS$ denote the FGLS estimators of the affine pricing model with constant and time varying prices of risk, respectively; $4FGLS^\mu$ and $4FGLS$ denote the four stage FGLS estimators of the affine pricing model with constant and time varying prices of risk, respectively; LML^μ and LML denote the linearized maximum likelihood estimators of the affine pricing model with constant and time varying prices of risk, respectively. The sample period is 1952Q1-2010Q2.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS^μ	OLS	$FGLS^\mu$	$FGLS$	$4FGLS^\mu$	$4FGLS$	LML^μ	LML
FF11	-0.169	-0.248	-0.144	-0.187	-0.207	-0.215	-0.204	-0.239
FF12	0.020	-0.031	0.024	-0.027	0.010	-0.033	0.010	-0.038
FF13	0.018	-0.092	0.017	-0.092	0.016	-0.079	0.016	-0.082
FF14	-0.037	-0.188	-0.035	-0.186	-0.043	-0.178	-0.043	-0.177
FF15	-0.026	-0.146	-0.025	-0.143	-0.022	-0.130	-0.023	-0.134
FF21	-0.098	-0.201	-0.093	-0.177	-0.103	-0.210	-0.102	-0.206
FF22	0.003	0.015	0.004	0.013	-0.001	-0.051	-0.001	-0.042
FF23	-0.032	-0.170	-0.033	-0.175	-0.055	-0.172	-0.054	-0.172
FF24	-0.004	-0.152	-0.004	-0.151	-0.005	-0.158	-0.005	-0.159
FF25	0.017	-0.078	0.012	-0.083	0.003	-0.089	0.004	-0.089
FF31	-0.014	-0.170	-0.013	-0.194	-0.013	-0.192	-0.013	-0.193
FF32	0.006	-0.151	0.003	-0.149	-0.003	-0.152	-0.003	-0.152
FF33	-0.036	-0.104	-0.038	-0.105	-0.047	-0.105	-0.047	-0.105
FF34	-0.007	-0.167	-0.007	-0.167	-0.019	-0.183	-0.019	-0.181
FF35	-0.012	-0.043	-0.016	-0.047	-0.018	-0.059	-0.018	-0.056
FF41	0.050	-0.121	0.029	-0.162	0.009	-0.174	0.010	-0.176
FF42	-0.069	-0.161	-0.080	-0.163	-0.056	-0.159	-0.057	-0.159
FF43	0.003	-0.146	-0.001	-0.149	-0.004	-0.150	-0.004	-0.150
FF44	-0.004	-0.129	-0.004	-0.130	-0.006	-0.133	-0.006	-0.133
FF45	0.013	-0.134	0.014	-0.120	0.001	-0.134	0.001	-0.134
FF51	0.037	-0.287	0.022	-0.314	0.024	-0.313	0.024	-0.316
FF52	0.059	-0.212	0.038	-0.247	0.021	-0.261	0.022	-0.259
FF53	0.008	-0.228	-0.004	-0.223	-0.013	-0.218	-0.012	-0.218
FF54	0.001	-0.141	0.000	-0.140	0.002	-0.139	0.002	-0.139
FF55	0.008	-0.156	0.000	-0.169	-0.007	-0.193	-0.007	-0.187
Avg	**-0.011**	**-0.146**	**-0.013**	**-0.148**	**-0.021**	**-0.155**	**-0.021**	**-0.156**

are compared to the traditional Fama-MacBeth estimator of the scaled CAPM. Indeed, as columns 2, 4, 6 and 8 of Table 2 reveal, with each of the four considered estimators, the time varying price of risk specification of the affine model consistently outperforms the Fama-MacBeth estimator of the scaled CAPM suggested by Lettau-Ludvigson. As an example, the difference in average absolute pricing errors between the Fama-MacBeth estimator and the LML estimator ranges from 4 basis points for the FF12 portfolio to 32 basis points per quarter for the FF51 portfolio, with the average across the 25 portfolios being about 15 basis points. In sum, while we confirm that *cay* captures time variation in excess returns, we also find that our more general price of risk specification implies considerably lower conditional pricing errors than the empirical specification used by Lettau and Ludvigson (2001).

6 Conclusion

Dynamic asset pricing models are at the heart of modern finance theory. Moreover, virtually all of the macro-finance literature of recent decades is cast in dynamic terms. The goal of this paper is to provide a unifying framework for estimating generic dynamic asset pricing models which impose cross sectional no arbitrage restrictions and allow for prices of risk to vary with observable state variables. Our approach nests the popular Fama-MacBeth two pass estimator for static pricing models and dynamic pricing models that do not impose cross sectional restrictions. Alternatively, one could also view our framework as a system of (multivariate) predictive regressions with cross equation constraints implied by no arbitrage.

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 fast and robust, as well as efficient under conditions that we discuss. In our empirical illustration, we revisit an influential asset pricing model that may be cast as special case of our generic approach. We find support that time variation of risk premia is present in the model, and that explicitly accounting for this time variation can substantially reduce conditional pricing errors.

Appendix A: Additional Results

Here we collect results that are too cumbersome to be placed in the main text.

A.1 Feasible GLS

In this section we provide instructions on how to implement a feasible version of the GLS estimators, $(\hat{\Lambda}_{\text{gls}}, \hat{B}_{4\text{gls}})$ (the corresponding steps for $(\hat{\Lambda}^{\mu}_{\text{gls}}, \hat{B}^{\mu}_{4\text{gls}})$ follow analogously).

(i) Follow steps 1-3 as discussed in the main text after the presentation of Theorem 1. With the output from these steps we may construct,

$$\hat{E}_{\text{ols}} = R_e - \hat{B}'_{\text{ols}} \left(\hat{\Lambda}_{\text{ols}} Z_- + \hat{V}_{\text{ols}} \right).$$

We may then estimate $\hat{\Sigma}_{e,\text{ols}}$ by the diagonal elements of the matrix $T^{-1} \cdot \hat{E}_{\text{ols}} M_{\iota} \hat{E}'_{\text{ols}}$ where $M_{\iota} = I_T - T^{-1} \cdot \iota_T \iota'_T$.

(ii) With $\hat{\Sigma}_{e,\text{ols}}$ we may then construct the feasible GLS estimator of Λ via,

$$\hat{\Lambda}_{\text{fgls}} = \left(\hat{B}_{\text{ols}} \hat{\Sigma}^{-1}_{e,\text{ols}} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} \hat{\Sigma}^{-1}_{e,\text{ols}} R_e Z'_- \left(Z_- Z'_- \right)^{-1}.$$

A.2 Maximum Likelihood Estimation when Σ_e is Unrestricted

When we relax the diagonality assumption of Assumption 2 (b) and instead only assume that Σ_e is positive definite, then we can easily characterize the maximum likelihood solution. It follows from Theorem 4 that $\hat{\Psi}_{\text{mle}} = \hat{\Psi}_{\text{ols}}$. This implies that we may replace V by $\hat{V}_{\text{ols}} = X - \hat{\Psi}_{\text{ols}} Z_-$. Next note that given (Ψ, B, Λ), $\hat{\Sigma}_{e,\text{mle}} = \frac{EE'}{T}$ and given Λ and Ψ,

$$\hat{B}_{\text{mle}\,|\,\Psi,\Lambda} = \left[\left(\Lambda Z_- + V \right) \left(\Lambda Z_- + V \right)' \right]^{-1} \left(\Lambda Z_- + V \right) R'_e.$$

Thus, if we concentrate the likelihood with respect to Σ_e, B, it is clear that maximizing the likelihood is equivalent to minimizing the expression $|EE'|$, where

$$E = R_e - R_e \left(\Lambda Z_- + \hat{V}_{\text{ols}} \right)' \left[\left(\Lambda Z_- + \hat{V}_{\text{ols}} \right) \left(\Lambda Z_- + \hat{V}_{\text{ols}} \right)' \right]^{-1} \left(\Lambda Z_- + \hat{V}_{\text{ols}} \right).$$

By standard properties of determinants,

$$\begin{aligned} \min_{\Lambda} |EE'| &= \min_{\Lambda} \left| R_e R'_e - R_e \left(\Lambda Z_- + \hat{V}_{\text{ols}} \right)' \left[\left(\Lambda Z_- + \hat{V}_{\text{ols}} \right) \left(\Lambda Z_- + \hat{V}_{\text{ols}} \right)' \right]^{-1} \left(\Lambda Z_- + \hat{V}_{\text{ols}} \right) R'_e \right| \\ &= |R_e R'_e| \min_{\Lambda} \frac{|D'YM_R Y'D|}{|D'YY'D|}, \end{aligned}$$

where $M_R = I_T - R_e' \left(R_e R_e'\right)^{-1} R_e$, $D = [I_K \ \Lambda]'$ and Y is the $(2K+1) \times T$ matrix $Y = \left[\hat{V}_{\text{ols}}' \ Z_-'\right]'$. Next, if we define $G_1 = Y M_R Y'$, $G_2 = YY'$ then the solution to this minimization problem is,

$$\min_{\Lambda} \frac{|D' G_1 D|}{|D' G_2 D|} = \prod_{i=1}^{K} \mu_i,$$

where $\{\mu_i : i = 1, \ldots, K\}$ are the K smallest roots of the equation $|G_1 - \mu G_2| = 0$. Moreover, the maximum likelihood estimator of D is the eigenvectors associated with these K smallest roots. Thus, $\hat{D} = (a_1, \ldots, a_K)$,where a_i is the eigenvector associated with the eigenvalue μ_i. Finally, we need to normalize the first K rows of $\hat{D}$. First, partition $\hat{D}$ as

$$\hat{D} = \begin{bmatrix} \hat{D}_1 \\ \hat{D}_2 \end{bmatrix},$$

where $\hat{D}_1$ and $\hat{D}_2$ are $K \times K$ and $(K+1) \times K$, respectively. Then the normalized version is $\hat{D}\left(\hat{D}_1\right)^{-1}$ which implies that

$$\hat{\Lambda}_{\text{mle}} = \left(\hat{D}_1'\right)^{-1} \hat{D}_2'.$$

Given $\left(\hat{\Psi}_{\text{mle}}, \hat{\Lambda}_{\text{mle}}\right)$,

$$\hat{B}_{\text{mle}} = \left[\left(\hat{\Lambda}_{\text{mle}} Z_- + \hat{V}_{\text{ols}}\right)\left(\hat{\Lambda}_{\text{mle}} Z_- + \hat{V}_{\text{ols}}\right)'\right]^{-1} \left(\hat{\Lambda}_{\text{mle}} Z_- + \hat{V}_{\text{ols}}\right) R_e'.$$

A.3 Inference on $B'\Lambda$

Suppose that we have two generic estimators of B and Λ, say $\bar{B}$ and $\bar{\Lambda}$ which satisfy,

$$\sqrt{T} \begin{pmatrix} \bar{b} - b \\ \bar{\lambda} - \lambda \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} \bar{\mathcal{V}}_b & \bar{\mathcal{C}}_{b,\lambda} \\ \bar{\mathcal{C}}_{b,\lambda}' & \bar{\mathcal{V}}_\lambda \end{bmatrix}\right).$$

Then it may be shown that

$$\sqrt{T} \operatorname{vec}\left(\bar{B}'\bar{\Lambda} - B'\Lambda\right) \xrightarrow{d} \mathcal{N}\left(0, \bar{\mathcal{V}}_{b\lambda}\right),$$

where

$$\begin{aligned} \bar{\mathcal{V}}_{b\lambda} \ = \ & \left(\Lambda' \otimes I_N\right) \kappa_{KN} \bar{\mathcal{V}}_b \kappa_{KN}' \left(\Lambda \otimes I_N\right) + \left(I_{(K+1)} \otimes B'\right) \bar{\mathcal{V}}_\lambda \left(I_{(K+1)} \otimes B\right) \\ & + \left(\Lambda' \otimes I_N\right) \kappa_{KN} \bar{\mathcal{C}}_{b,\lambda} \left(I_{(K+1)} \otimes B\right) + \left(I_{(K+1)} \otimes B'\right) \bar{\mathcal{C}}_{b,\lambda}' \kappa_{KN}' \left(\Lambda \otimes I_N\right). \end{aligned}$$

When prices of risk are constant,

$$\sqrt{T} \operatorname{vec}\left(\bar{B}'\bar{\lambda}_0 - B'\lambda_0\right) \xrightarrow{d} \mathcal{N}\left(0, \bar{\mathcal{V}}_{b\lambda}^{\mu}\right),$$

where

$$\begin{aligned} \bar{\mathcal{V}}_{b\lambda}^{\mu} \ = \ & \left(\lambda_0' \otimes I_N\right) \kappa_{KN} \bar{\mathcal{V}}_b^{\mu} \kappa_{KN}' \left(\lambda_0 \otimes I_N\right) + B' \bar{\mathcal{V}}_\lambda^{\mu} B \\ & + \left(\lambda_0' \otimes I_N\right) \kappa_{KN} \bar{\mathcal{C}}_{b,\lambda}^{\mu} B + B' \bar{\mathcal{C}}_{b,\lambda}^{\mu\prime} \kappa_{KN}' \left(\lambda_0 \otimes I_N\right). \end{aligned}$$

B Appendix B: Preliminary Lemmas

We first present some preliminary lemmas that will be useful for the proofs of theorems. Define $\Omega_{vz} = VZ_-'/\sqrt{T}$ and similarly for Ω_{ez} and Ω_{ev}; define $\Omega_{v\iota} = V\iota_T/\sqrt{T}$ and similarly for $\Omega_{e\iota}$; define $\Omega_{vv} = VV'/T$ and similarly for Ω_{ee}. Recall that $\hat{\Upsilon} = Z_-Z_-'/T$. Throughout the Appendix let $M_\iota = I_T - T^{-1}\cdot\iota_T\iota_T'$, $M_Z = I_T = Z_-'\left(Z_-Z_-'\right)^{-1}Z_-$ and $\hat{V}_{\text{ols}} = XM_Z$.

Lemma 1 *Suppose Assumptions 1 and 2 hold. Then,*

(i) Ω_{vv}, Ω_{ee}, Ω_{vz}, Ω_{ez}, Ω_{ev} *are all* $O_p(1)$*;*

(ii) $\mathbb{E}[\hat{\Upsilon}] = \Upsilon + o(1)$*;*

(iii) Let $\sqrt{T}(\hat{\Psi}-\Psi) = O_p(1)$. *Put* $\hat{V} \equiv X - \hat{\Psi}Z_-$. *Then,* $T^{-1}\cdot\hat{V}\hat{V}' = \Omega_{vv} + O_p(T^{-1})$*;*

(iv) Let $\sqrt{T}(\hat{B}-B)$, $\sqrt{T}(\hat{\Lambda}-\Lambda)$ *and* $\sqrt{T}(\hat{\Psi}-\Psi)$ *all be* $O_p(1)$. *Put* $\hat{V} = X-\hat{\Psi}Z_-$ *and* $\hat{E} = R_e-\hat{B}'(\hat{\Lambda}Z_-+\hat{V})$. *Then,* $T^{-1}\cdot\hat{E}M_\iota\hat{E}' = \Omega_{ee} + O_p(T^{-1})$, *where* $M_\iota = I_T - T^{-1}\cdot\iota_T\iota_T'$*;*

(v) $\sqrt{T}(\hat{\Psi}_{\text{ols}}-\Psi) = \Delta_{\Psi,\text{ols}} + o_p(1)$ *where* $\Delta_{\Psi,\text{ols}} = \Omega_{vz}\Upsilon^{-1}$*;*

(vi) $\sqrt{T}(\hat{B}_{\text{ols}}-B) = \Delta_{B,\text{ols}} + o_p(1)$ *where* $\Delta_{B,\text{ols}} = \Sigma_v^{-1}\Omega_{ev}'$.

Proof of Lemma 1. (i) and (ii) follow by Assumption 2 and standard calculations. (iii) follows since,

$$T^{-1}\cdot\hat{V}\hat{V}' = \Omega_{vv} - T^{-1/2}\cdot\left(\hat{\Psi}-\Psi\right)\Omega_{vz}' - T^{-1/2}\cdot\Omega_{vz}\left(\hat{\Psi}-\Psi\right)' + \left(\hat{\Psi}-\Psi\right)\hat{\Upsilon}\left(\hat{\Psi}-\Psi\right)',$$

and using the results from (i) and (ii). For (iv) note that

$$T^{-1}\cdot\hat{E}M_i\hat{E}' = T^{-1}\cdot EM_\iota E' + T^{-1}\cdot\left[\left(\hat{E}-E\right)M_\iota E' + EM_\iota\left(\hat{E}-E\right)' + \left(\hat{E}-E\right)M_\iota\left(\hat{E}-E\right)'\right].$$

Then the result follows by (i) and since $\left(\hat{E}-E\right)E'$ and $T^{-1/2}\cdot\left(\hat{E}-E\right)\iota_T$ are $O_p(1)$. (v) follows since,

$$\sqrt{T}\left(\hat{\Psi}_{\text{ols}}-\Psi\right) = \Omega_{vz}\hat{\Upsilon}^{-1} = \Delta_{\Psi,\text{ols}} + o_p(1),$$

by (i) and (ii). For (vi)

$$\hat{B}_{\text{ols}} = \left(\hat{V}_{\text{ols}}\hat{V}_{\text{ols}}'\right)^{-1}\hat{V}_{\text{ols}}R_e' = B + \left(\hat{V}_{\text{ols}}\hat{V}_{\text{ols}}'\right)^{-1}\hat{V}_{\text{ols}}\left(\left(V-\hat{V}_{\text{ols}}\right)'B + E'\right).$$

Thus,

$$\sqrt{T}\left(\hat{B}_{\text{ols}}-B\right) = \sqrt{T}\left(\hat{V}_{\text{ols}}\hat{V}_{\text{ols}}'\right)^{-1}\hat{V}_{\text{ols}}\left(V-\hat{V}_{\text{ols}}\right)'B + \sqrt{T}\left(\hat{V}_{\text{ols}}\hat{V}_{\text{ols}}'\right)^{-1}\hat{V}_{\text{ols}}E'.$$

The first term is $o_p(1)$ by (i), (ii), and (iii). The second term is

$$\sqrt{T}\left(\hat{V}_{\text{ols}}\hat{V}_{\text{ols}}'\right)^{-1}\hat{V}_{\text{ols}}E' = \left(T^{-1}\cdot\hat{V}_{\text{ols}}\hat{V}_{\text{ols}}'\right)^{-1}T^{-1/2}\hat{V}_{\text{ols}}E' = \Sigma_v^{-1}\Omega_{ev}' + o_p(1),$$

by (i), (iii), and the continuous mapping theorem.

Lemma 2 *Suppose Assumptions 1 and 2 hold. Then,*

$$\begin{pmatrix}\omega_{vz}\\ \omega_{ez}\\ \omega_{ev}\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix}(\Upsilon\otimes\Sigma_v) & 0 & 0\\ 0 & (\Upsilon\otimes\Sigma_e) & 0\\ 0 & 0 & (\Sigma_v\otimes\Sigma_e)\end{bmatrix}\right),$$

and

$$\Omega_{vv} \xrightarrow{p} \Sigma_v, \qquad \Omega_{ee} \xrightarrow{p} \Sigma_e.$$

Proof of Lemma 2. These results follow by Assumption 2 and standard properties of martingale difference sequences, see, for example, White (2001).

For simplicity of notation define $\hat{\Lambda}(\Gamma)$ and $\hat{B}(\Gamma)$ as

$$\begin{aligned}\hat{\Lambda}(\Gamma) &= \left(\hat{B}_{\text{ols}}\Gamma\hat{B}'_{\text{ols}}\right)^{-1}\hat{B}_{\text{ols}}\Gamma R_e Z'_-\left(Z_-Z'_-\right)^{-1}\\ \hat{B}(\Gamma) &= \left[\left(\hat{\Lambda}(\Gamma)Z_- + \hat{V}_{\text{ols}}\right)\left(\hat{\Lambda}(\Gamma)Z_- + \hat{V}_{\text{ols}}\right)'\right]^{-1}\left(\hat{\Lambda}(\Gamma)Z_- + \hat{V}_{\text{ols}}\right)R'_e,\end{aligned}$$

and similarly for $\hat{\lambda}^{\mu}(\Gamma)$ and $\hat{B}^{\mu}(\Gamma)$. For example, $\hat{\Lambda}(I_N) = \hat{\Lambda}_{\text{ols}}$ and $\hat{\Lambda}\left(\hat{\Sigma}^{-1}_{e,\text{ols}}\right) = \hat{\Lambda}_{\text{fgls}}$.

Lemma 3 *Suppose Assumptions 1 and 2 hold and $\hat{\Gamma}$ satisfies $\hat{\Gamma} - \Gamma = o_p(1)$. Then, if the data are generated by equation (6),*

(*i*) $\sqrt{T}(\hat{\Lambda}(\hat{\Gamma}) - \Lambda) = \Delta_{\Lambda,\Gamma} + o_p(1)$ *where* $\Delta_{\Lambda,\Gamma} = \Delta_{\Psi,\text{ols}} + (B\Gamma B')^{-1}B\Gamma\left[\Omega_{ez}\Upsilon^{-1} - \Omega_{ev}\Sigma_v^{-1}\Lambda\right]$,

(*ii*) $\sqrt{T}(\hat{B}(\hat{\Gamma}) - B) = \Delta_{B,\Gamma} + o_p(1)$ *where* $\Delta_{B,\Gamma} = [\Lambda\Upsilon\Lambda' + \Sigma_v]^{-1}\left(\Lambda\Omega'_{vz}B + (\Lambda\Omega'_{ez} + \Omega'_{ev}) - \Lambda\Upsilon\Delta'_{\Lambda,\Gamma}B\right)$.

Proof of Lemma 3. For (i) note first that

$$\begin{aligned}R_e Z'_-\left(Z_-Z'_-\right)^{-1} &= \hat{B}'_{\text{ols}}\Lambda + \hat{B}'_{\text{ols}}VZ'_-\left(Z_-Z'_-\right)^{-1} - \left(\hat{B}_{\text{ols}} - B\right)'\Lambda\\ &\quad - \left(\hat{B}_{\text{ols}} - B\right)'VZ'_-\left(Z_-Z'_-\right)^{-1} + EZ'_-\left(Z_-Z'_-\right)^{-1}.\end{aligned}$$

Thus,

$$\hat{\Lambda}(\Gamma) = \mathcal{T}_{\Lambda,\Gamma,1} + \mathcal{T}_{\Lambda,\Gamma,2} + \mathcal{T}_{\Lambda,\Gamma,3} + \mathcal{T}_{\Lambda,\Gamma,4} + \mathcal{T}_{\Lambda,\Gamma,5},$$

where

$$\mathcal{T}_{\Lambda,\Gamma,1} = \Lambda,$$

$$\mathcal{T}_{\Lambda,\Gamma,2} = VZ'_-\left(Z_-Z'_-\right)^{-1},$$

$$\mathcal{T}_{\Lambda,\Gamma,3} = -\left(\hat{B}_{\text{ols}}\hat{\Gamma}\hat{B}'_{\text{ols}}\right)^{-1}\hat{B}_{\text{ols}}\hat{\Gamma}\left(\hat{B}_{\text{ols}} - B\right)'\Lambda,$$

$$\mathcal{T}_{\Lambda,\Gamma,4} = -\left(\hat{B}_{\text{ols}}\hat{\Gamma}\hat{B}'_{\text{ols}}\right)^{-1}\hat{B}_{\text{ols}}\hat{\Gamma}\left(\hat{B}_{\text{ols}} - B\right)'VZ'_-\left(Z_-Z'_-\right)^{-1},$$

$$\mathcal{T}_{\Lambda,\Gamma,5} = \left(\hat{B}_{\text{ols}}\hat{\Gamma}\hat{B}'_{\text{ols}}\right)^{-1}\hat{B}_{\text{ols}}\hat{\Gamma}EZ'_-\left(Z_-Z'_-\right)^{-1}.$$

Note that $\mathcal{T}_{\Lambda,\Gamma,2} = \hat{\Psi}_{\text{ols}} - \Psi$ and $\mathcal{T}_{\Lambda,\Gamma,4} = o_p\left(T^{-1/2}\right)$ by Lemma 1. By Lemma 1 and the assumptions on $\hat{\Gamma}$,

$$\sqrt{T}\mathcal{T}_{\Lambda,\Gamma,3} = -\sqrt{T}\left(\hat{B}_{\text{ols}}\hat{\Gamma}\hat{B}'_{\text{ols}}\right)^{-1}\hat{B}_{\text{ols}}\hat{\Gamma}\left(\hat{B}_{\text{ols}} - B\right)'\Lambda = -(B\Gamma B')^{-1}B\Gamma\Delta'_{B,\text{ols}}\Lambda + o_p(1),$$

and

$$\sqrt{T}\mathcal{T}_{\Lambda,\Gamma,5} = \left(\hat{B}_{\text{ols}}\hat{\Gamma}\hat{B}'_{\text{ols}}\right)^{-1}\hat{B}_{\text{ols}}\hat{\Gamma}EZ_-\left(Z_-Z'_-\right)^{-1} = (B\Gamma B')^{-1}B\Gamma\Omega_{ez}\Upsilon^{-1} + o_p(1).$$

Using Lemma 1 again we have,

$$\sqrt{T}\left(\hat{\Lambda}(\Gamma) - \Lambda\right) = \Delta_{\Lambda,\Gamma} + o_p(1),$$

where

$$\Delta_{\Lambda,\Gamma} = \Omega_{vz}\Upsilon^{-1} + \left(B\Gamma B'\right)^{-1} B\Gamma\Omega_{ez}\Upsilon^{-1} - \left(B\Gamma B'\right)^{-1} B\Gamma\Omega_{ev}\Sigma_v^{-1}\Lambda.$$

For (ii) we will suppress the dependence on Γ of $\hat{\Lambda}(\Gamma)$. Note first that

$$\begin{aligned}
\sqrt{T}\hat{B}\left(\hat{\Gamma}\right) &= \left[T^{-1}\left(\hat{\Lambda}Z_- + \hat{V}\right)\left(\hat{\Lambda}Z_- + \hat{V}\right)'\right]^{-1} T^{-1/2}\left(\hat{\Lambda}Z_- + \hat{V}\right)R_e' \\
&= \left[\hat{\Lambda}\hat{\Upsilon}\hat{\Lambda}' + \hat{\Sigma}_v\right]^{-1} T^{-1/2}\left(\hat{\Lambda}Z_- + \hat{V}\right)\left(\left(Z_-'\Lambda' + V'\right)B + E'\right) \\
&= \sqrt{T}B + \mathcal{T}_{B,\Gamma,1} + \mathcal{T}_{B,\Gamma,2},
\end{aligned}$$

where

$$\mathcal{T}_{B,\Gamma,1} = \left[\hat{\Lambda}\hat{\Upsilon}\hat{\Lambda}' + \hat{\Sigma}_v\right]^{-1} T^{-1/2}\left(\Lambda Z_- + V\right)\left(\left(Z_-'\left(\Lambda - \hat{\Lambda}\right)' + \left(V - \hat{V}\right)'\right)B + E'\right),$$

$$\mathcal{T}_{B,\Gamma,2} = \left[\hat{\Lambda}\hat{\Upsilon}\hat{\Lambda}' + \hat{\Sigma}_v\right]^{-1} T^{-1/2}\left(\left(\hat{\Lambda} - \Lambda\right)Z_- + \left(\hat{V} - V\right)\right)\left(\left(Z_-'\left(\Lambda - \hat{\Lambda}\right)' + \left(V - \hat{V}\right)'\right)B + E'\right).$$

By Lemma 1 all of the terms of $\mathcal{T}_{B,\Gamma,2}$ are $o_p(1)$ so we need only consider $\mathcal{T}_{B,\Gamma,1}$.

$$\mathcal{T}_{B,\Gamma,1} = \mathcal{T}_{B,\Gamma,1,1} + \mathcal{T}_{B,\Gamma,1,2} + \mathcal{T}_{B,\Gamma,1,3} + \mathcal{T}_{B,\Gamma,1,4} + \mathcal{T}_{B,\Gamma,1,5} + \mathcal{T}_{B,\Gamma,1,6},$$

where

$$\mathcal{T}_{B,\Gamma,1,1} = \left[\hat{\Lambda}\hat{\Upsilon}\hat{\Lambda}' + \hat{\Sigma}_v\right]^{-1} T^{-1/2}\Lambda Z_- Z_-'\left(\Lambda - \hat{\Lambda}\right)' B,$$

$$\mathcal{T}_{B,\Gamma,1,2} = \left[\hat{\Lambda}\hat{\Upsilon}\hat{\Lambda}' + \hat{\Sigma}_v\right]^{-1} T^{-1/2}\Lambda Z_-\left(V - \hat{V}\right)' B,$$

$$\mathcal{T}_{B,\Gamma,1,3} = \left[\hat{\Lambda}\hat{\Upsilon}\hat{\Lambda}' + \hat{\Sigma}_v\right]^{-1} T^{-1/2}\Lambda Z_- E',$$

$$\mathcal{T}_{B,\Gamma,1,4} = \left[\hat{\Lambda}\hat{\Upsilon}\hat{\Lambda}' + \hat{\Sigma}_v\right]^{-1} T^{-1/2} V Z_-'\left(\Lambda - \hat{\Lambda}\right)' B,$$

$$\mathcal{T}_{B,\Gamma,1,5} = \left[\hat{\Lambda}\hat{\Upsilon}\hat{\Lambda}' + \hat{\Sigma}_v\right]^{-1} T^{-1/2} V\left(V - \hat{V}\right)' B,$$

$$\mathcal{T}_{B,\Gamma,1,6} = \left[\hat{\Lambda}\hat{\Upsilon}\hat{\Lambda}' + \hat{\Sigma}_v\right]^{-1} T^{-1/2} V E'.$$

$\mathcal{T}_{B,\Gamma,1,4}$ and $\mathcal{T}_{B,\Gamma,1,5}$ are $o_p(1)$ by Lemma 1. By Lemma 1 and (i),

$$\mathcal{T}_{B,\Gamma,1,1} = -\left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\Lambda\Upsilon\Delta_{\Lambda,\Gamma}' B + o_p(1), \qquad \mathcal{T}_{B,\Gamma,1,2} = \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\Lambda\Omega_{vz}' B + o_p(1),$$

$$\mathcal{T}_{B,\Gamma,1,3} = \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\Lambda\Omega_{ez}' + o_p(1), \qquad \mathcal{T}_{B,\Gamma,1,6} = \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\Omega_{ev}' + o_p(1).$$

Thus,

$$\sqrt{T}\left(\hat{B}(\Gamma) - B\right) = \mathcal{T}_{B,\Gamma,1} + \mathcal{T}_{B,\Gamma,2} = \Delta_{B,\Gamma} + o_p(1),$$

where

$$\Delta_{B,\Gamma} = \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\left(\Lambda\Omega_{vz}' B + \Lambda\Omega_{ez}' + \Omega_{ev}' - \Lambda\Upsilon\Delta_{\Lambda,\Gamma}' B\right). \qquad \blacksquare$$

Lemma 4 *Suppose Assumptions 1 and 2 hold and that Γ satisfies the assumptions of Lemma 3. Then,if the data are generated by equation (9),*

1. $\sqrt{T}(\hat{\lambda}_0^{\mu}(\hat{\Gamma}) - \lambda_0) = \Delta_{\lambda,\Gamma}^{\mu} + o_p(1)$ *where* $\Delta_{\lambda,\Gamma}^{\mu} = \Omega_{v\iota} + \left(B\Gamma B'\right)^{-1} B\Gamma\Omega_{e\iota} - \left(B\Gamma B'\right)^{-1} B\Gamma\Omega_{ev}\Sigma_v^{-1}\lambda_0$,
2. $\sqrt{T}(\hat{B}^{\mu}(\hat{\Gamma}) - B) = \Delta_{B,\Gamma}^{\mu} + o_p(1)$ *where* $\Delta_{B,\Gamma}^{\mu} = \left[\Sigma_v + \lambda_0\lambda_0'\right]^{-1}\left(\lambda_0\Omega_{v\iota}' B + \lambda_0\Omega_{e\iota}' + \Omega_{ev}' - \lambda_0\Delta_{\lambda,\Gamma}^{\mu\prime} B\right)$.

Proof of Lemma 4. This proof follows by similar steps as in the proof of Lemma 3 and so is omitted to conserve space.

Appendix C: Proofs of Theorems

Proof of Theorem 1. Let us first consider $\hat{\psi}_{\text{ols}}$ and $\hat{b}_{\text{ols}}$. For $\hat{\psi}_{\text{ols}}$ by Lemma 1 we have that $\delta_{\Psi,\text{ols}} = \left(\Upsilon^{-1} \otimes I_K\right) \omega_{vz}$ and so by Lemma 2 and Slutsky's Lemma $\delta_{\Psi,\text{ols}} \to_d \mathcal{N}\left(0, \left(\Upsilon^{-1} \otimes \Sigma_v\right)\right)$. Similarly, $\delta_{B,\text{ols}} = \left(I_N \otimes \Sigma_v^{-1}\right) \kappa_{NK} \omega_{ev}$ and so by Lemma 2, Slutsky's Lemma and standard properties of the commutation matrix $\delta_{B,\text{ols}} \to_d \mathcal{N}\left(0, \Sigma_e \otimes \Sigma_v^{-1}\right)$. Next, by Lemma 3 we have,

$$\delta_{\Lambda,\Gamma} = \left(\Upsilon^{-1} \otimes I_K\right) \kappa_{K(K+1)} \omega_{zv} + \left(\Upsilon^{-1} \otimes \left(B\Gamma B'\right)^{-1} B\Gamma\right) \omega_{ez} - \left(\Lambda' \Sigma_v^{-1} \otimes \left(B\Gamma B'\right)^{-1} B\Gamma\right) \omega_{ev}.$$

By Assumption 2 the covariance terms are zero and so by Lemma 2 and Slutsky's Lemma, $\delta_{\Lambda,\Gamma} \to_d \mathcal{N}\left(0, \mathcal{V}_{\hat{\lambda},\Gamma}\right)$, where

$$\begin{aligned}
\mathcal{V}_{\hat{\lambda},\Gamma} &= \left(\Upsilon^{-1} \otimes \Sigma_v\right) + \left(\Upsilon^{-1} \otimes \left(B\Gamma B'\right)^{-1} B\Gamma\Sigma_e\Gamma B' \left(B\Gamma B'\right)^{-1}\right) \\
&\quad + \left(\Lambda' \Sigma_v^{-1} \Lambda \otimes \left(B\Gamma B'\right)^{-1} B\Gamma\Sigma_e\Gamma B' \left(B\Gamma B'\right)^{-1}\right),
\end{aligned}$$

and so the result follows by setting Γ to I_N or Σ_e^{-1}.

Proof of Theorem 2. First we will calculate the asymptotic variance of $\hat{B}\left(\Gamma\right)$. By Lemma 3 we have,

$$\begin{aligned}
\Delta_{B,\Gamma} &= \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1} \left(\Lambda\Omega_{vz}' B + \left(\Lambda\Omega_{ez}' + \Omega_{ev}'\right) - \Lambda\Upsilon\Delta_{\Lambda,\Gamma}' B\right) \\
&= \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1} \left(\Lambda\Omega_{ez}' \left(I_N - \Gamma B' \left(B\Gamma B'\right)^{-1} B\right) + \Omega_{ev}' + \Lambda\Upsilon\Lambda'\Sigma_v^{-1}\Omega_{ev}'\Gamma B' \left(B\Gamma B'\right)^{-1} B\right),
\end{aligned}$$

or equivalently,

$$\begin{aligned}
\delta_{B,\Gamma} &= \left(\left(I_N - B' \left(B\Gamma B'\right)^{-1} B\Gamma\right) \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1} \Lambda\right) \kappa_{N(K+1)} \omega_{ez} \\
&\quad + \left[\left(I_N \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\right) + \left(B' \left(B\Gamma B'\right)^{-1} B\Gamma \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1} \Lambda\Upsilon\Lambda'\Sigma_v^{-1}\right)\right] \kappa_{NK} \omega_{ev}.
\end{aligned}$$

By Assumption 2 the covariance terms are zero and so by Lemma 2, Slutsky's Lemma, and standard properties of the commutation matrix $\delta_{\Lambda,\Gamma} \to_d \mathcal{N}(0, \mathcal{V}_{\hat{b},\Gamma})$, where

$$\begin{aligned}
\mathcal{V}_{\hat{b},\Gamma} &= \left(\left(I_N - B' \left(B\Gamma B'\right)^{-1} B\Gamma\right) \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1} \Lambda\right) \left(\Sigma_e \otimes \Upsilon\right) \left(\left(I_N - B' \left(B\Gamma B'\right)^{-1} B\Gamma\right) \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1} \Lambda\right)' \\
&\quad + \left(I_N \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\right) \left(\Sigma_e \otimes \Sigma_v\right) \left(I_N \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\right)' \\
&\quad + \left(B' \left(B\Gamma B'\right)^{-1} B\Gamma \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1} \Lambda\Upsilon\Lambda'\Sigma_v^{-1}\right) \left(\Sigma_e \otimes \Sigma_v\right) \left(B' \left(B\Gamma B'\right)^{-1} B\Gamma \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1} \Lambda\Upsilon\Lambda'\Sigma_v^{-1}\right)' \\
&= \left(\Sigma_e \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\right) \\
&\quad + \left(B' \left(B\Gamma B'\right)^{-1} B\Gamma\Sigma_e\Gamma B' \left(B\Gamma B'\right)^{-1} B \otimes \left(\Sigma_v^{-1} - \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\right)\right)
\end{aligned}$$

and so the result follows by setting Γ to I_N or Σ_e. For the asymptotic covariance between $\hat{B}\left(\Gamma\right)$ and $\hat{\Lambda}\left(\Gamma\right)$ we need only to calculate

$$\lim_{T\to\infty} \mathbb{E}\left[\delta_{B,\Gamma}\delta_{\Lambda,\Gamma}'\right] = \mathcal{T}_{B,\Lambda,\Gamma,1} + \mathcal{T}_{B,\Lambda,\Gamma,2},$$

where

$$\mathcal{T}_{B,\Lambda,\Gamma,1} = \mathbb{E}\left[\left(\left(I_N - \Gamma B'\left(B\Gamma B'\right)^{-1}B\right)' \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\Lambda\right)\operatorname{vec}\left(\Omega_{ez}'\right)\omega_{ez}'\left(\Upsilon^{-1}\otimes\left(B\Gamma B'\right)^{-1}B\Gamma\right)'\right],$$

$$\begin{aligned}\mathcal{T}_{B,\Lambda,\Gamma,2} \quad = \quad & \mathbb{E}\left[\left[\left(I_N \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\right) + \left(B'\left(B\Gamma B'\right)^{-1}B\Gamma \otimes \left[\Lambda\Upsilon\Lambda' + \Sigma_v\right]^{-1}\Lambda\Upsilon\Lambda'\Sigma_v^{-1}\right)\right]\operatorname{vec}\left(\Omega_{ev}'\right) \times \right.\\ & \left. \omega_{ev}'\left(-\left(\Lambda'\Sigma_v^{-1}\otimes\left(B\Gamma B'\right)^{-1}B\Gamma\right)\right)'\right].\end{aligned}$$

By similar calculations as above we have that

$$\begin{aligned}\mathcal{T}_{B,\Lambda,\Gamma,1} \quad = \quad & \left(\Sigma_e\Gamma B'\left(B\Gamma B'\right)^{-1}\otimes\left[\Lambda\Upsilon\Lambda'+\Sigma_v\right]^{-1}\Lambda\right)\kappa_{K(K+1)} \\ & -\left(B'\left(B\Gamma B'\right)^{-1}B\Gamma\Sigma_e\Gamma B'\left(B\Gamma B'\right)^{-1}\otimes\left[\Lambda\Upsilon\Lambda'+\Sigma_v\right]^{-1}\Lambda\right)\kappa_{K(K+1)},\end{aligned}$$

and

$$\begin{aligned}\mathcal{T}_{B,\Lambda,\Gamma,2} \quad = \quad & -\left(\Sigma_e\Gamma B'\left(B\Gamma B'\right)^{-1}\otimes\left[\Lambda\Upsilon\Lambda'+\Sigma_v\right]^{-1}\Lambda\right)\kappa_{K(K+1)} \\ & -\left(B'\left(B\Gamma B'\right)^{-1}B\Gamma\Sigma_e\Gamma B'\left(B\Gamma B'\right)^{-1}\otimes\left[\Lambda\Upsilon\Lambda'+\Sigma_v\right]^{-1}\Lambda\Upsilon\Lambda'\Sigma_v^{-1}\Lambda\right)\kappa_{K(K+1)}.\end{aligned}$$

Cancelling the common term, simplifying the resulting expression, and setting Γ to I_N or Σ_e^{-1} yields the result.

Remark 3 *If we relax the conditional homoskedasticity assumption (Assumption 2 (c)), then the asymptotic covariance between some terms in the asymptotically linear representation of $\hat{\Lambda}(\Gamma)$ and $\hat{B}(\Gamma)$ are nonzero. Including these terms yields the variance formulas discussed in the main text. For concreteness, consider*

$$\delta_{\Lambda,\Gamma} = \left(\Upsilon^{-1}\otimes I_K\right)\kappa_{K(K+1)}\omega_{zv} + \left(\Upsilon^{-1}\otimes\left(B\Gamma B'\right)^{-1}B\Gamma\right)\omega_{ez} - \left(\Lambda'\Sigma_v^{-1}\otimes\left(B\Gamma B'\right)^{-1}B\Gamma\right)\omega_{ev}.$$

We still have that $\mathbb{E}\left[\omega_{zv}\omega_{ez}'\right]=0$ but $\mathbb{E}\left[\omega_{zv}\omega_{ev}'\right]$ and $\mathbb{E}\left[\omega_{ez}\omega_{ev}'\right]$ are now nonzero and so we must incorporate terms involving these expressions in the asymptotic variance formula.

Proof of Theorem 3. This proof follows by similar steps as in the proofs of Theorems 1 and 2 and so is omitted to conserve space.

Proof of Theorem 4. The results of Theorem 4 follow by standard matrix calculus. The (transpose of the) first element of the score vector is,

$$\begin{aligned}\frac{\partial\ell\left(\theta;\sigma_e,\sigma_v\right)}{\partial\psi'} \quad & = \quad -\operatorname{vec}(E)'\left(I_T\otimes\Sigma_e^{-1}\right)\frac{\partial\operatorname{vec}(E)}{\partial\psi'} - \operatorname{vec}(V)'\left(I_T\otimes\Sigma_v^{-1}\right)\frac{\partial\operatorname{vec}(V)}{\partial\psi'} \\ & = \quad -\operatorname{vec}\left(B\Sigma_e^{-1}EZ_-'\right)' + \operatorname{vec}\left(\Sigma_v^{-1}VZ_-'\right)'.\end{aligned}$$

By similar steps we may obtain the second and third elements of the score vector. Similarly, the (1,1) element of the information matrix is,

$$\begin{aligned}\frac{\partial^2\ell\left(\theta;\sigma_e,\sigma_v\right)}{\partial\psi\partial\psi'} \quad & = \quad \frac{\partial}{\partial\psi}\left[-\operatorname{vec}\left(B\Sigma_e^{-1}EZ_-'\right)+\operatorname{vec}\left(\Sigma_v^{-1}VZ_-'\right)\right] \\ & = \quad -\left(Z_-Z_-'\otimes B\Sigma_e^{-1}B'\right) - \left(Z_-Z_-'\otimes\Sigma_v^{-1}\right) \\ & = \quad -T\cdot\left(\hat{\Upsilon}\otimes\left(B\Sigma_e^{-1}B'+\Sigma_v^{-1}\right)\right),\end{aligned}$$

and so by Lemma 1,

$$\lim_{T\to\infty}\left(-\frac{1}{T}\right)\mathbb{E}\left[\frac{\partial^2\ell\left(\theta;\sigma_e,\sigma_v\right)}{\partial\psi\partial\psi'}\right]=\left(\Upsilon\otimes\left(B\Sigma_e^{-1}B'+\Sigma_v^{-1}\right)\right).$$

The rest of the elements may be obtained in a similar fashion. Finally, we may utilize standard results on inverses of partitioned matrices to obtain the inverse information matrix.

Proof of Corollary 1. This follows by Theorem 4 and standard properties of one step estimators. See, for example, van der Vaart (1998).

Proof of Theorem 5. This proof follows by similar steps as in the proof of Theorems 4 and so is omitted to conserve space.

References

Adrian, T., R. K. Crump, and E. Moench (2012): "Pricing the Term Structure with Linear Regressions," *Federal Reserve Bank of New York Staff Reports*, 340.

Adrian, T., and E. Moench (2008): "Pricing the Term Structure with Linear Regressions," *Federal Reserve Bank of New York Staff Reports*, 340.

Amihud, Y., C. M. Hurvich, and Y. Wang (2009): "Multiple-Predictor Regressions: Hypothesis Testing," *Review of Financial Studies*, 22(1), 413–434.

Ang, A., and M. Piazzesi (2003): "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50(4), 745–787.

Campbell, J. Y. (1987): "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18(2), 373–399.

——— (1996): "Understanding Risk and Return," *Journal of Political Economy*, 104(2), 572–621.

Campbell, J. Y., A. W. Lo, and A. C. MacKinley (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton.

Chen, N.-F., R. Roll, and S. A. Ross (1986): "Economic Forces and the Stock Market," *Journal of Business*, 59(3), 383–403.

Chen, R., and R. Kan (2004): "Finite Sample Analysis of Two-Pass Cross-sectional Regressions," Working paper, University of Toronto.

Cochrane, J. (1996): "A Cross-sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104(3), 572–621.

——— (2005): *Asset Pricing*. Princeton University Press, Princeton, Revised edition.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985): "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53(2), 363–384.

Dai, Q., and K. Singleton (2002): "Expecations Puzzle, Time-Varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics*, 63(3), 415–441.

Duffee, G. R. (2002): "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57(1), 405–443.

Fama, E. F., and K. R. French (1992): "The Cross-section of Expected Stock Returns," *Journal of Finance*, 47(2), 427–65.

Fama, E. F., and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33(1), 3 – 56.

Fama, E. F., and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 113(3), 607–636.

Ferson, W. E., and C. R. Harvey (1991): "The Variation of Economic Risk Premiums," *Journal of Political Economy*, 99(2), 385–415.

Gibbons, M. R. (1982): "Multivariate Tests of Financial Models: A New Approach," *Journal of Financial Economics*, 10(1), 3–27.

Gibbons, M. R., and W. Ferson (1985): "Testing Asset Pricing Models with Changing Expectations and an Unobservable Market Portfolio," *Journal of Financial Economics*, 14(2), 217–236.

Hansen, L. P., and R. J. Hodrick (1983): "Risk Averse Speculation in the Forward Foreign Exchange Market: An Econometric Analysis of Linear Models," in *Exchange Rates and International Macroeconomics*, NBER Chapters, pp. 113–152. National Bureau of Economic Research, Inc.

Harvey, C. R. (1989): "Time-varying Conditional Covariances in Tests of Asset Pricing Models," *Journal of Financial Economics*, 24(2), 289–317.

——— (1991): "The World Price of Covariance Risk," *Journal of Finance*, 46(1), 111–157.

Jagannathan, R., and Z. Wang (1996): "The Conditional CAPM and the Cross-section of Expected Returns," *Journal of Finance*, 51(1), 3–53.

——— (1998): "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-sectional Regression," *Journal of Finance*, 53(4), 1285–1309.

Kandel, S. (1984): "The Likelihood Ratio Test Statistic of Mean-variance Efficiency without a Riskless Asset," *Journal of Financial Economics*, 13(4), 575–592.

Kleibergen, F. (2009): "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149(2), 149–173.

Lettau, M., and S. Ludvigson (2001): "Resurrecting the (C)CAPM: A Cross-sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109(6), 1238–1287.

Merton, R. C. (1973): "An Intertemporal Asset Pricing Model," *Econometrica*, 41(5), 867–887.

Nagel, S., and K. J. Singleton (2010): "Estimation and Evaluation of Conditional Asset Pricing Models," *Journal of Finance*, Forthcoming.

Petkova, R. (2006): "Do the Fama-French Factors Proxy for Innovations in Predictive Variables?," *Journal of Finance*, 61(2), 581–612.

Roll, R. (1985): "A Note on the Geometry of Shanken's CSR T^2 Test for Mean/Variance Efficiency," *Journal of Financial Economics*, 14(3), 349–357.

Rothenberg, T. J., and C. Leenders (1964): "Efficient Estimation of Simultaneous Equation Systems," *Econometrica*, 32(1), 57–76.

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

——— (1986): "Testing Portfolio Efficiency when the Zero-beta Rate is Unknown: A Note," *Journal of Finance*, 41(1), 269–276.

——— (1992): "On the Estimation of Beta-Pricing Models," *Review of Financial Studies*, 5(1), 1–33.

Shanken, J., and G. Zhou (2007): "Estimating and Testing Beta Pricing Models: Alternative Methods and their Performance in Simulations," *Journal of Financial Economics*, 84(1), 40–86.

Singleton, K. (2006): *Empirical Dynamic Asset Pricing*. Princeton University Press.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, New York.

White, H. (2001): *Asymptotic Theory for Econometricians*. Academic Press, San Diego, Revised edition.