

Adrian, Tobias; Crump, Richard K.; Moench, Emanuel

**Working Paper**

## Efficient, regression-based estimation of dynamic asset pricing models

Staff Report, No. 493

**Provided in Cooperation with:**  
Federal Reserve Bank of New York

*Suggested Citation:* Adrian, Tobias; Crump, Richard K.; Moench, Emanuel (2011) : Efficient, regression-based estimation of dynamic asset pricing models, Staff Report, No. 493, Federal Reserve Bank of New York, New York, NY

This Version is available at:

<https://hdl.handle.net/10419/60952>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

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

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

## **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

JEL classification: E58, G28

### **Abstract**

We study regression-based estimators for beta representations of dynamic asset pricing models with affine and exponentially affine pricing kernel specifications. 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](mailto:tobias.adrian@ny.frb.org), [richard.crump@ny.frb.org](mailto:richard.crump@ny.frb.org), [emanuel.moench@ny.frb.org](mailto: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. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

## 1 Introduction

This paper proposes efficient, regression based estimators for dynamic asset pricing models with time varying prices of risk. Estimators are developed for both affine and exponentially affine pricing kernel specifications commonly used in equity and fixed income applications. 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 typical asset pricing applications that allow the comparison of various estimators and provide 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. We first study regression based estimators for the affine pricing kernel specification. 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 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.

We next study estimation of dynamic asset pricing models when the pricing kernel has an exponentially affine formulation. In this setting, it is common to assume conditionally normal distributions for innovations. In the exponentially affine case, the return generating process implied by no arbitrage has a convexity term which involves a quadratic function of betas. Any asymptotically efficient estimator has to take the quadratic dependence of the convexity adjustment on the risk factor exposures into account. As a result, the recomputation of efficient betas in the fourth step cannot be done and hence the OLS and GLS estimators do not achieve efficiency. We therefore also provide an LML estimator which is asymptotically efficient for the exponentially affine case. As in the affine pricing kernel specification, this estimator is easily programmed based on outputs from the regression based approaches and does not involve numerical optimization.

**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 provide regression based estimators for dynamic asset pricing models with exponentially affine pricing kernels and conditionally Gaussian shocks. The assumption of an exponentially affine pricing kernel and conditionally Gaussian shocks is a special case of the general equilibrium framework of Merton (1973) and Cox, Ingersoll, and Ross (1985a). It is also commonly used in affine dynamic term structure models (e.g., Vasicek (1977), Cox, Ingersoll, and Ross (1985b)). In practice, the term structure models usually impose additional restrictions deriving from boundary conditions and are estimated using numerical methods for maximum likelihood estimation (see e.g., Dai and Singleton (2000), Duffee (2002) or Piazzesi (2003)). When no additional restrictions are imposed, we show how to solve the maximum likelihood problem using an LML estimator, thus providing closed form solutions to highly dimensional optimization problems. In addition, we contrast the optimal LML estimation approach to the regression based approach, allowing us to exactly quantify the asymptotic efficiency loss by using the direct regression estimator relative to the LML estimator.

We illustrate the usefulness of our approach and associated estimators by applying them to two well known asset pricing models, the conditional CAPM suggested by Lettau and Ludvigson (2001) and the ICAPM considered by Campbell (1996). Both models implicitly assume time variation in risk premia, but have 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, and that the unrestricted beta representation of the Campbell (1996) model is equivalent to that implied by our exponentially affine pricing kernel when prices of risk are constant. Following the original papers, we estimate the Lettau and Ludvigson (2001) model for a cross section of size and book-to-market sorted equity portfolios, and apply the Campbell (1996) model to a cross section of equity and fixed income portfolios. For both models, we compare OLS, feasible GLS, and the LML estimators. In both empirical applications, and 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 models with time varying prices of risk give rise to economically important reductions in conditional pricing errors relative to models with constant prices of risk. Finally, even though the sample sizes are moderate in both models, 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. Section 5 provides the analysis of the exponentially affine model. We provide empirical illustrations of our estimators in Section 6, and Section 7 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  $\text{vec}(\cdot)$  operator applied to a matrix (e.g.,  $\gamma = \text{vec}(\Gamma)$ ). For an  $m \times n$  matrix  $A$ , define the  $mn \times mn$  commutation matrix  $\kappa_{mn}$  which satisfies  $\text{vec}(A') = \kappa_{mn} \text{vec}(A)$ .  $\text{bdiag}_{[pq],[mn]}(A_i)$  will denote a  $p \times q$  block-diagonal matrix (not necessarily square) with  $i$ th diagonal element equal to the  $m \times n$  matrix  $A_i$ . Finally, let  $I_m$  be the  $m \times m$  identity matrix and let  $\iota_m$  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:

$$X_{t+1} = \zeta + \Phi X_t + v_{t+1}, \quad t = 1, \dots, T, \quad (1)$$

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 (1985a)'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:

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

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:

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

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

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

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

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

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

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

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

$$\beta_{i,t} = \Sigma_{v,t}^{-1} \mathbb{C} [X_{t+1}, R_{i,t+1}^e | \mathcal{F}_t], \quad (4)$$

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:

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

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

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

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  $e_{i,t+1}$ :

$$R_{i,t+1}^e = \beta_{i,t}' (\lambda_0 + \Lambda_1 X_t) + \beta_{i,t}' v_{t+1} + e_{i,t+1}, \quad t = 1, \dots, T. \quad (5)$$

The excess return thus depends on the expected return,  $\beta_{i,t}' (\lambda_0 + \Lambda_1 X_t)$ , 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  $i = 1, \dots, N$  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  $R_{i,t+1}^e = a_{i,t} + c_{i,t}X_t + \beta'_{i,t}v_{t+1} + e_{i,t+1}$ , 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  $\mathbb{E} \left[ M_{t+1} R_{i,t+1}^e \middle| \mathcal{F}_t \right] = 0$  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  $\Phi$  have modulus less than one.

**Assumption 2** Put  $\varepsilon_t = (v'_t, e'_t)'$  and define  $\mathcal{G}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \dots)$ . (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  $\Sigma_v$  and (2,2)-element  $\Sigma_e$ ; (b)  $\Sigma_v$  is a positive definite matrix and  $\Sigma_e$  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  $\Lambda$ . 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),

$$R_e = B'\Lambda Z_- + B'V + E \quad (6)$$

$$X = \Psi Z_- + V, \quad (7)$$

where  $\Psi = [\zeta \ \Phi]$ ,  $\Lambda = [\lambda_0 \ \Lambda_1]$ ,  $X = [X_1 \ X_2 \ \dots \ X_T]$ ,  $X_- = [X_0 \ X_1 \ \dots \ X_{T-1}]$  and  $Z_- = [\iota_T \ X_-']'$ .  $R_e$ ,  $E$  and  $V$  are matrices which are formed by stacking  $R_{i,t}^e$ ,  $e_{i,t}$  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  $\Psi$ , prices of risk  $\Lambda$ , and risk factor exposures  $B$ .<sup>1</sup>

---

<sup>1</sup>In the discussion of the affine model we treat  $\Sigma_e$  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.

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

$$\begin{aligned} \hat{\Psi}_{\text{ols}} &= X Z'_- (Z_- Z'_-)^{-1}, & \hat{B}_{\text{ols}} &= (X M_Z X')^{-1} X M_Z R'_e, \\ \hat{\Lambda}_{\text{ols}} &= \left( \hat{B}_{\text{ols}} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} R_e Z'_- (Z_- Z'_-)^{-1}, & \hat{\Lambda}_{\text{gls}} &= \left( \hat{B}_{\text{ols}} \Sigma_e^{-1} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} \Sigma_e^{-1} R_e Z'_- (Z_- Z'_-)^{-1}, \end{aligned}$$

satisfy

$$\begin{aligned} \sqrt{T}(\hat{\psi}_{\text{ols}} - \psi) &\xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\hat{\psi}, \text{ols}}), & \sqrt{T}(\hat{b}_{\text{ols}} - b) &\xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\hat{b}, \text{ols}}), \\ \sqrt{T}(\hat{\lambda}_{\text{ols}} - \lambda) &\xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\hat{\lambda}, \text{ols}}), & \sqrt{T}(\hat{\lambda}_{\text{gls}} - \lambda) &\xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\hat{\lambda}, \text{gls}}), \end{aligned}$$

where

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

and  $\Upsilon = \Upsilon(\Psi, \Sigma_v) = \text{plim}_{T \rightarrow \infty} (Z_- Z'_- / T)$ .

Theorem 1 can be intuitively summarized as a three stage estimator:<sup>2</sup>

1. The estimator  $\hat{\Psi}_{\text{ols}} = X Z'_- (Z_- Z'_-)^{-1}$  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  $\hat{V}_{\text{ols}} = X - \hat{\Psi}_{\text{ols}} Z_- = X M_Z$ . Finally  $\Sigma_v$  may be estimated via  $\hat{\Sigma}_{v, \text{ols}} = \hat{V}_{\text{ols}} \hat{V}'_{\text{ols}} / T$ .
2. The estimator  $\hat{B}_{\text{ols}} = (X M_Z X')^{-1} X M_Z R'_e = \left( \hat{V}_{\text{ols}} \hat{V}'_{\text{ols}} \right)^{-1} \hat{V}_{\text{ols}} R'_e$  is the (transpose of the) OLS seemingly unrelated regression (OLS-SUR) estimator of the excess returns across assets  $R_e$  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  $\Psi$  and  $B$ , we may estimate the prices of risk  $\Lambda$  by a regression involving two components: first, the OLS-SUR estimator of the quantity  $B' \Lambda$  in equation (6),  $R_e Z'_- (Z_- Z'_-)^{-1}$ ; second, the estimated risk factor exposures,  $\hat{B}_{\text{ols}}$ . Regressing the former onto the latter yields,  $\hat{\Lambda}_{\text{ols}} = \left( \hat{B}_{\text{ols}} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} R_e Z'_- (Z_- Z'_-)^{-1}$ . 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  $\hat{\Lambda}_{\text{gls}} = \left( \hat{B}_{\text{ols}} \Sigma_e^{-1} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} \Sigma_e^{-1} R_e Z'_- (Z_- Z'_-)^{-1}$ .

<sup>2</sup>It is with some abuse of notation that we label our estimators of  $\Lambda$  as "OLS" and "GLS"; however, we do so to remain consistent with standard practices.

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  $\Lambda$  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  $\Sigma_e$  is a scalar variance matrix. Moreover, in general,  $\hat{\Lambda}_{\text{gls}}$  is asymptotically efficient relative to  $\hat{\Lambda}_{\text{ols}}$ .<sup>3</sup>

### 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  $\Lambda$  and  $V$  are known. Then, if we rewrite equation (6) as

$$R_e = B' (\Lambda Z_- + V) + E, \quad (8)$$

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  $\Lambda$  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:

**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*

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

<sup>3</sup>Throughout the paper comparisons between matrices will be understood to be in a positive-definite sense.

where

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

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  $\Psi$  is known (so that  $V$  is observed) but  $\Lambda$  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  $\Lambda$  then only the first term of the asymptotic variance formula,  $(\Sigma_e \otimes [\Lambda\Upsilon\Lambda' + \Sigma_v]^{-1})$ , would remain. The second term,  $(B' (B\Sigma_e^{-1}B')^{-1} B \otimes [\Lambda\Upsilon\Lambda' + \Sigma_v]^{-1} \Lambda\Upsilon\Lambda'\Sigma_v^{-1})$ , arises because we must replace  $\Lambda$  by the estimate  $\hat{\Lambda}_{gls}$ .

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

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

The right hand side matrix is positive semi-definite.<sup>4</sup> 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 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  $\hat{B}_{ols}$ . 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  $\hat{B}_{4gls}$  will be asymptotically efficient relative to  $\hat{B}_{ols}$ .

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

<sup>4</sup>When  $\mathcal{V}_{\hat{b},4gls} \neq \mathcal{V}_{\hat{b},4ols}$  the difference  $\mathcal{V}_{\hat{b},ols} - \mathcal{V}_{\hat{b},4ols}$  is non-definite.

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  $\Lambda$  in Theorem 2 to facilitate inference on the quantity  $B'\Lambda$ .<sup>5</sup> 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:  $(\hat{B}_{\text{ols}}, \hat{B}_{4\text{ols}}, \hat{B}_{4\text{gls}})$  and  $(\hat{\Lambda}_{\text{ols}}, \hat{\Lambda}_{\text{gls}})$ . We may also estimate the risk premium,  $B'\Lambda$  directly. The OLS-SUR estimator of equation (6),  $R_eZ'_-(Z_-Z'_-)^{-1}$ , 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.<sup>6</sup>) 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.*

### 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

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

<sup>5</sup>Section A.5 (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  $\Lambda$ . Consequently, asymptotic standard errors for elements of  $B'\Lambda$  may be constructed using the results from Theorem 2.

<sup>6</sup>For a discussion of the multivariate predictive regression model and its properties see Amihud, Hurvich, and Wang (2009).

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

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  $\Lambda$  presented in Theorem 1 with  $Z_-$  replaced by  $\iota_T$ . 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  $\hat{B}_{ols}$ . If we let  $\tilde{V}_{ols} = X - \bar{X} \iota'_T$ ,  $\bar{X} = X \iota_T / T$  and define

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

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

We will see in the next section that this re-estimation step, in the form of  $\hat{B}_{FM,4glS}$ , 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,

$$\begin{aligned} \Pi_1 &= \lim_{T \rightarrow \infty} \mathbb{E} \left[ \left( \text{vec} \left( T^{-1/2} E Z_- \right) \text{vec} \left( T^{-1/2} E V' \right) \right)' \right], \\ \Pi_2 &= \lim_{T \rightarrow \infty} \mathbb{E} \left[ \left( \text{vec} \left( T^{-1/2} V Z_- \right) \text{vec} \left( T^{-1/2} E V' \right) \right)' \right], \end{aligned}$$

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

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

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

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

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

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

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 4 in the Appendix).<sup>7</sup>

### 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  $\Lambda_1 = 0$ .<sup>8</sup> Equation (6) then becomes

$$R_e^\mu = B'\lambda_0\iota'_T + B'V + E, \quad (9)$$

while equation (7) is unchanged. We use a superscript  $\mu$  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  $R_e^\mu$  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,*

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

and

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

satisfy

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

<sup>7</sup>Note, there is no counterpart to  $\mathcal{V}_{\hat{b},4\text{gls}}^*$  in Jagannathan and Wang (1998).

<sup>8</sup>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.



where,

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

Again, we may re-estimate the parameter  $B$  to obtain  $\hat{B}_{4gls}^\mu$  which is generally more efficient than  $\hat{B}_{ols}$ . The interpretation of each term of  $\mathcal{V}_{\hat{b},4gls}^\mu$  and  $\mathcal{V}_{\hat{b},4ols}^\mu$  is in perfect analogy with the general case. The second term reflects the need to provide an estimate of  $\lambda_0$  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  $\hat{\lambda}_{0,FM,ols}$  and  $\hat{\lambda}_{0,FM,gls}$ , respectively. When prices of risk are constant, the asymptotic variability of our proposed estimators of  $B$  and  $\Lambda$  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  $\Upsilon$ , 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

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

where  $E = E(\Psi, B, \Lambda)$  and  $V = V(\Psi)$  are treated as functions of the parameters.<sup>9</sup> 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  $(\psi, b, \lambda)$ . We may focus exclusively on these parameters because  $(\psi, b, \lambda)$  and  $(\sigma_e, \sigma_v)$  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  $\Sigma_e$  and  $\Sigma_v$  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*

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

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  $\Psi$ .<sup>10</sup> The second and third diagonal elements confirm that  $\hat{\Lambda}_{\text{gls}}$  and  $\hat{B}_{4\text{gls}}$  are efficient estimators of the parameters  $\Lambda$  and  $B$ . Shanken (1992) shows that the two pass GLS estimator is efficient when prices of risk are constant and factors are uncorrelated 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

<sup>9</sup>The ordering of parameters  $(\psi, b, \lambda)$  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  $\lim_{T \rightarrow \infty} -T^{-1} \mathbb{E}[\partial \ell / \partial \psi \partial b']$ .

<sup>10</sup>However, we will show in the next section that this result no longer holds when prices of risk are constant.

$\hat{B}_{\text{ols}}$  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.<sup>11</sup> We appeal to this approach because it does not appear that the maximum likelihood estimators of the parameters  $B$  and  $\Lambda$  are available without using numerical maximization.<sup>12</sup> 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  $\theta$ ,  $\bar{\theta}$  which satisfies  $\sqrt{T}(\bar{\theta} - \theta) = O_p(1)$  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  $\hat{\Upsilon} = Z_- Z_-'/T$ . Then the estimators formed by*

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

satisfy

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

**Remark 2** *When evaluating the score vector  $\dot{\ell}(\theta; \Sigma_e, \Sigma_v)$  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  $\Sigma_v$  and  $\Sigma_e$  (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  $\Psi$ . Recall that  $\hat{\Psi}_{\text{ols}}$  is the MLE when prices of risk are time varying. Using the results from Theorem 4 we have,

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

However, by construction  $X - \hat{\Psi}_{\text{ols}} Z_- = 0$  and so  $\hat{\psi}_{\text{lml}} = \hat{\psi}_{\text{ols}}$ . 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.

<sup>11</sup>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).

<sup>12</sup>Note, when we only assume that  $\Sigma_e$  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).

## 4.2 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  $\lambda_0$  (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  $\theta^\mu = (\psi', b', \lambda_0')'$  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  $\dot{\ell}^\mu = \dot{\ell}^\mu(\theta^\mu; \Sigma_v, \Sigma_e)$ , the  $K(K+N+2) \times 1$ , partitioned vector, with elements*

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

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

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

where

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

and  $\Upsilon_1$  is the first column of the matrix  $\Upsilon$ .

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,  $((\Upsilon - \Upsilon_1 \Upsilon_1') \otimes B \Sigma_e^{-1} B')$ , a function of parameters of the return equation,  $B$  and  $\Sigma_e$ , which governs the degree of efficiency loss. Fortunately, this is only a concern if inference on  $\Psi$  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  $\hat{B}_{4\text{gls}}^\mu$  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  $\hat{\lambda}_{0,FM,glS}$  is asymptotically equivalent to the MLE of  $\lambda_0$  under Assumption 3. Define  $\theta^{FM} = (\zeta', b', \lambda_0)'$ . It can be shown that in the Fama-MacBeth framework (under Assumption 3) the information matrix,  $\mathcal{I}^{FM}(\theta^{FM}; \Sigma_e, \Sigma_v)$ , has an inverse,  $\mathcal{H}^{FM} = \mathcal{H}^{FM}(\theta^{FM}; \Sigma_e, \Sigma_v)$ , which is a partitioned matrix comprised of the following elements,

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

This confirms the Shanken (1992) result and adds the additional result that the four stage estimator of  $B$ ,  $\hat{B}_{FM,4glS}$ , 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 The Exponentially Affine Model

### 5.1 Pricing Kernel Assumptions and Return Generation

In this section, we maintain Assumption 3 throughout. In addition, we assume the same dynamics of a  $K \times 1$  vector of state variables  $X_t$  as in the affine model (i.e., from equation (1)). In the exponentially affine setup, we study the dynamic properties of log excess returns  $r_{i,t+1}^e$ , which we define as:

$$r_{i,t+1}^e = \ln(P_{t+1}^i + D_{t+1}^i) - \ln P_t^i - r_t^F, \quad (10)$$

where  $P^i$  and  $D^i$  denote the market price and dividend of asset  $i$ , respectively, and where the continuously compounded risk free rate is denoted  $r_t^F$ . We can again decompose unexpected returns  $r_{i,t+1}^e - E[r_{i,t+1}^e | \mathcal{F}_t]$  into a systematic component  $\beta'_{i,t} v_{t+1}$  and an idiosyncratic component  $e_{i,t+1}$ :

$$r_{i,t+1}^e - E[r_{i,t+1}^e | \mathcal{F}_t] = \beta'_{i,t} v_{t+1} + e_{i,t+1} \quad (11)$$

where we now denote  $\beta_{i,t} = \Sigma_{v,t}^{-1} \mathbb{C} \left[ X_{t+1}, r_{i,t+1}^e \mid \mathcal{F}_t \right]$ . In a term structure model fitted to zero coupon bonds, the idiosyncratic component would usually equal zero, while the idiosyncratic shock would correspond to idiosyncratic cash flow news in an equity model.

The pricing kernel is now assumed to be exponentially affine:

$$M_{t+1}^\epsilon = \exp \left( -r_t^F - \frac{1}{2} \lambda_t' \Sigma_{v,t}^{-1} \lambda_t - \lambda_t' \Sigma_{v,t}^{-1/2} v_{t+1} \right) \quad (12)$$

where as before  $\lambda_t = \Sigma_{v,t}^{-1/2} (\lambda_0 + \Lambda_1 X_t)$ . We use a superscript of  $\epsilon$  to differentiate the exponentially

affine model.<sup>13</sup> With the exponentially affine pricing kernel, we can write the no arbitrage condition as

$$1 = E \left[ M_{t+1}^\epsilon \left( \frac{P_{t+1}^i + D_{t+1}^i}{P_t^i} \right) \middle| \mathcal{F}_t \right] = E \left[ \exp \left( -\frac{1}{2} \lambda_t' \Sigma_{v,t}^{-1} \lambda_t - \lambda_t' \Sigma_{v,t}^{-1/2} v_{t+1} + r_{i,t+1}^\epsilon \right) \middle| \mathcal{F}_t \right] \quad (13)$$

Using Assumption 3 and plugging equation (11) into equation (13) gives an expression for expected returns:

$$E [r_{i,t+1}^\epsilon | \mathcal{F}_t] = \beta_{i,t}' \lambda_0 + \beta_{i,t}' \lambda_1 X_t - \frac{1}{2} \text{vec} (\beta_{i,t} \beta_{i,t}')' \text{vec} (\Sigma_{v,t}) - \frac{1}{2} \sigma_i^2. \quad (14)$$

If we compare the return generating equation of the affine model, equation (3), to equation (14), we see that the exponentially affine model requires an additional ‘‘Jensen’’ term to adjust for convexity. This term is proportional to  $\beta_{i,t}' \Sigma_{v,t} \beta_{i,t} + \sigma_i^2$ , which makes clear that the betas now enter into the return generating equation in both a linear and quadratic fashion. Excess returns can be written as:

$$r_{i,t+1}^\epsilon = \beta_{i,t}' \lambda_0 + \beta_{i,t}' \lambda_1 X_t - \frac{1}{2} \text{vec} (\beta_{i,t} \beta_{i,t}')' \text{vec} (\Sigma_{v,t}) - \frac{1}{2} \sigma_i^2 + \beta_{i,t}' v_{t+1} + e_{i,t+1}, \quad (15)$$

If we impose Assumption 1, we can again stack this equation to obtain:

$$r_e^\epsilon = B' \lambda_0 \iota_T' - \frac{1}{2} J_B \text{vec} (\Sigma_v) \iota_T' - \frac{1}{2} d_e \iota_T' + B' \Lambda_1 X_- + B' V + E, \quad (16)$$

where  $J_B = [\text{vec} (\beta_1 \beta_1') \cdots \text{vec} (\beta_N \beta_N')]'$  is a  $N \times K^2$  matrix and  $d_e = (\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)'$  is the diagonal of  $\Sigma_e$ . We use  $r_e^\epsilon$  to emphasize that this is a  $N \times T$  matrix of log excess returns.

The assumptions of the exponentially affine setting are often used in the term structure literature (see Piazzesi (2003) and Singleton (2006) for overviews). As will be discussed in Section 6, this return generating process is equivalent to that implied by the unconstrained pricing model of Campbell (1996) when prices of risk are constant, i.e. when  $\Lambda_1 = 0$ .

## 5.2 Three Stage Estimators

In direct analogy with the affine model we may estimate the parameters in equation (16) using a three stage regression approach. The asymptotic properties of these estimators are detailed in the following theorem. Note that all of the quantities in the theorem will be fully described in the discussion after the theorem.<sup>14</sup>

**Theorem 6** *Suppose Assumptions 1 and 3 hold and we observe  $r_e^\epsilon$  generated by equation (16).*

<sup>13</sup>The superscript ‘‘ $\epsilon$ ’’ on  $\beta_{i,t}$  and  $\lambda_t$  is suppressed for notational convenience.

<sup>14</sup>In the Appendix we also provide all necessary results for the constant price of risk case ( $\Lambda_1 = 0$ ).

Then the following estimators,

$$\begin{aligned}\hat{\Lambda}_{\text{ols}}^\epsilon &= \left(\hat{B}_{\text{ols}}\hat{B}'_{\text{ols}}\right)^{-1}\hat{B}_{\text{ols}}\left[r_e^\epsilon + \frac{1}{2}\hat{J}_{B,\text{ols}}\text{vec}(\hat{\Sigma}_{v,\text{ols}})\iota'_T + \frac{1}{2}\hat{d}_{e,\text{sur}}\iota'_T\right]Z'_-(Z_-Z'_-)^{-1}, \\ \hat{\Lambda}_{\text{fgls}}^\epsilon &= \left(\hat{B}_{\text{ols}}\hat{\Sigma}_{e,\text{ols}}^{-1}\hat{B}'_{\text{ols}}\right)^{-1}\hat{B}_{\text{ols}}\hat{\Sigma}_{e,\text{ols}}^{-1}\left[r_e^\epsilon + \frac{1}{2}\hat{J}_{B,\text{ols}}\text{vec}(\hat{\Sigma}_{v,\text{ols}})\iota'_T + \frac{1}{2}\hat{d}_{e,\text{sur}}\iota'_T\right]Z'_-(Z_-Z'_-)^{-1},\end{aligned}$$

satisfy

$$\sqrt{T}(\hat{\lambda}_{\text{ols}}^\epsilon - \lambda) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\lambda,\text{ols}}^\epsilon), \quad \sqrt{T}(\hat{\lambda}_{\text{fgls}}^\epsilon - \lambda) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\lambda,\text{fgls}}^\epsilon),$$

where  $\mathcal{V}_{\lambda,\text{ols}}^\epsilon$  and  $\mathcal{V}_{\lambda,\text{fgls}}^\epsilon$  are given in equations (21) and (22) in Appendix A.3.1.

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

1. First, exactly as in the affine case, we estimate the VAR by OLS to obtain  $\hat{\Psi}_{\text{ols}}$  and  $\hat{V}_{\text{ols}}$ . We may then construct  $\hat{\Sigma}_{v,\text{ols}} = \hat{V}_{\text{ols}}\hat{V}'_{\text{ols}}/T$ .
2. Estimate a OLS-SUR regression of  $r_e^\epsilon$  on  $Z_-$  and  $\hat{V}_{\text{ols}}$ . We can then estimate  $E$  via,

$$\hat{E}_{\text{sur}}^\epsilon = r_e^\epsilon - \left[r_e^\epsilon Z'_-(Z_-Z'_-)^{-1}\right]Z_- - \hat{B}'_{\text{ols}}\hat{V}_{\text{ols}}, \quad \hat{B}_{\text{ols}} = \left(\hat{V}_{\text{ols}}\hat{V}'_{\text{ols}}\right)^{-1}\hat{V}_{\text{ols}}r_e^{\epsilon'}$$

This yields  $\hat{B}_{\text{ols}}$ . If we let  $\hat{\beta}_{i,\text{ols}}$  be the  $i$ th column of  $\hat{B}_{\text{ols}}$  then we can estimate  $J_B$  by  $\hat{J}_{B,\text{ols}} = \left[\text{vec}(\hat{\beta}_{1,\text{ols}}\hat{\beta}'_{1,\text{ols}}) \cdots \text{vec}(\hat{\beta}_{N,\text{ols}}\hat{\beta}'_{N,\text{ols}})\right]'$ . Moreover, the elements of  $\hat{d}_{e,\text{sur}}$  are equal to the diagonal elements of the  $N \times N$  matrix  $\hat{E}_{\text{sur}}^\epsilon\hat{E}_{\text{sur}}^{\epsilon'}/T$ .

- 3a. Now that we have estimates of  $B$ ,  $J_B$ ,  $\Sigma_v$  and  $d_e$  we need only estimate the quantity  $B'\Lambda$ . This can be achieved by a OLS-SUR regression of the convexity adjusted log excess returns on the lagged forecasting variables,  $\left[r_e^\epsilon + \frac{1}{2}\hat{J}_{B,\text{ols}}\text{vec}(\hat{\Sigma}_{v,\text{ols}})\iota'_T + \frac{1}{2}\hat{d}_{e,\text{sur}}\iota'_T\right]Z'_-(Z_-Z'_-)^{-1}$ . Regressing this quantity onto  $\hat{B}_{\text{ols}}$  yields  $\hat{\Lambda}_{\text{ols}}^\epsilon$ .
- 3b. With the estimates obtained from steps 1, 2, and 3a we can now construct

$$\hat{E}_{\text{ols}}^\epsilon = r_e^\epsilon + \frac{1}{2}\hat{J}_{B,\text{ols}}\text{vec}(\hat{\Sigma}_{v,\text{ols}})\iota'_T + \frac{1}{2}\hat{d}_{e,\text{sur}}\iota'_T - \hat{B}'_{\text{ols}}\left(\hat{\Lambda}_{\text{ols}}^\epsilon Z_- + \hat{V}_{\text{ols}}\right).$$

We can then construct the estimator  $\hat{\Sigma}_{e,\text{ols}}$  as a diagonal matrix with diagonal elements equal to the diagonal elements of the matrix  $T^{-1} \cdot \hat{E}_{\text{ols}}^\epsilon M_\iota \hat{E}_{\text{ols}}^{\epsilon'}$  where  $M_\iota = I_T - T^{-1} \cdot \iota_T \iota'_T$ . The matrix  $M_\iota$  de-means the rows of  $\hat{E}_{\text{ols}}^\epsilon$ . This is a necessary step because  $\hat{E}_{\text{ols}}^\epsilon \iota_T$  is not equal to zero in general. We now have all the ingredients necessary to construct  $\hat{\Lambda}_{\text{fgls}}^\epsilon$ .

### 5.3 Efficiency in the Exponentially Affine Model

Define  $\theta^\epsilon = (\theta', \sigma'_v, d'_e)'$  and so we may write the log-likelihood (up to a constant) as

$$\begin{aligned} \ell^\epsilon(\theta^\epsilon) &\approx \frac{T}{2} \log(|\Sigma_e^{-1}|) + \frac{T}{2} \log(|\Sigma_v^{-1}|) - \frac{1}{2} \text{vec}(E^\epsilon)' (I_T \otimes \Sigma_e^{-1}) \text{vec}(E^\epsilon) \\ &\quad - \frac{1}{2} \text{vec}(V)' (I_T \otimes \Sigma_v^{-1}) \text{vec}(V), \end{aligned}$$

where  $E^\epsilon = E^\epsilon(\Psi, B, \Lambda, \Sigma_v, d_e)$  and  $V = V(\Psi)$  are treated as functions of the parameters. We again 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 exponentially affine model with respect to all of the parameters. The exact expressions are relegated to Section A.4.1 (in the Appendix) to conserve space.

**Theorem 7** *Suppose Assumptions 1 and 3 hold and we observe  $r_e^\epsilon$  generated by equation (16). Then the (scaled) score vector is  $\dot{\ell}^\epsilon(\theta^\epsilon)$  and the information matrix is  $\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)$ , with inverse  $\mathcal{H}^\epsilon(\theta^\epsilon; \Upsilon)$ , each with partitioned elements as presented in Appendix A.4.1.*

Although  $\hat{\Psi}_{\text{ols}}$  is still the MLE of  $\Psi$ , it does not appear that closed form expressions exists for the maximum likelihood estimators of  $\Lambda$ ,  $B$ ,  $\Sigma_v$ , or  $d_e$ . Moreover, unlike in the affine model,  $\hat{\Lambda}_{\text{gls}}^\epsilon$  is not asymptotically efficient.<sup>15</sup> To obtain asymptotically efficient estimators of all parameters we again utilize a feasible LML procedure.

**Corollary 2** *Suppose the assumptions of Theorem 7 hold. In addition assume there exists an estimator of  $\theta^\epsilon$ ,  $\bar{\theta}^\epsilon$  which satisfies  $\sqrt{T}(\bar{\theta}^\epsilon - \theta) = O_p(1)$ . Put  $\hat{\Upsilon} = Z'Z'/T$ . Then the estimators formed by*

$$\hat{\theta}_{\text{lml}}^\epsilon = \bar{\theta}^\epsilon + \mathcal{H}^\epsilon(\bar{\theta}^\epsilon; \hat{\Upsilon}) \times \dot{\ell}^\epsilon(\bar{\theta}^\epsilon)$$

satisfy

$$\sqrt{T}(\hat{\theta}_{\text{lml}}^\epsilon - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{H}^\epsilon(\theta^\epsilon; \Upsilon)).$$

**Remark 3** *In Appendix A we provide expressions for the partitioned elements of the information matrix in the exponentially affine model. Unlike in the affine model, not all of the elements of the inverse of the information matrix are available in a tidy form. However, we provide expressions for each partitioned element of the inverse information matrix so that it is easy to compute the LML estimator.*

The obvious choice of pilot estimators are  $\hat{\Psi}_{\text{ols}}$ ,  $\hat{B}_{\text{ols}}$ ,  $\hat{\Sigma}_{v,\text{ols}}$ ,  $\hat{d}_{e,\text{ols}}$  (the diagonal elements of  $\hat{\Sigma}_{e,\text{ols}}$ ) and either  $\hat{\Lambda}_{\text{ols}}^\epsilon$  or  $\hat{\Lambda}_{\text{fgls}}^\epsilon$  from Theorem 6. With these choices the LML estimator is still regression based as all the quantities used to construct the LML estimator come directly from

<sup>15</sup>However, for parameter values generally encountered in financial applications, we find that the loss of efficiency is concentrated in the estimation of  $\lambda_0$  whereas for  $\Lambda_1$  the estimator  $\hat{\Lambda}_{\text{fgls}}^\epsilon$  is very nearly asymptotically efficient.



standard regression output. Moreover, the appeal of this estimation approach is clear. We may easily construct asymptotically efficient estimators of  $(\Psi, B, \Lambda, \Sigma_v, \Sigma_e)$  through the LML procedure. For example, the empirically relevant choice of  $(N, K) = (20, 3)$  presents 110 parameters that may be estimated efficiently based on the output of familiar regression based methods.

## 6 Applications

We illustrate the estimators in two empirical applications drawn from prominent asset pricing models. Both applications illustrate 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. In the second application, we show that the return generating process of the unrestricted version of the ICAPM of Campbell (1996) is nested in our exponentially affine model when prices of risk are constant. While both models implicitly assume time variation in risk premia, they have 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. In both applications, we find strong empirical support in favor of dynamic price of risk specifications, as  $\Lambda_1$  is highly significant, and as pricing errors are reduced in economically meaningful magnitudes relative to constant price of risk specifications.

### 6.1 Lettau and Ludvigson (2001)

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:

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

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

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,  $f_{t+1} = R_{t+1}^M$  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

$$\mathbb{E}[R_{t+1} | \mathcal{F}_t] - R_t^f = \beta_t \lambda_t$$

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

$$\lambda_t = -R_t^f \mathbb{V}(R_{t+1}^M | \mathcal{F}_t) b_t = -R_t^f \mathbb{V}(R_{t+1}^M | \mathcal{F}_t) (b_0 + b_1 z_t)$$

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  $\mathbb{V}[R_{t+1}^M | \mathcal{F}_t] = \Sigma_M \forall t$ . 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

$$\mathbb{E}[R_{t+1} | \mathcal{F}_t] - R_t^f = \beta_t \lambda_t,$$

where

$$\lambda_t = -\Sigma_M R_t^f (b_0 + b_1 z_t),$$

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

$$X_t = \left( R_t^M, R_t^f, R_t^f \cdot cay_t \right)'$$

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).<sup>16</sup>

We can assess whether the prices of risk associated with the risk factors in the model feature time variation by testing if the rows of  $\Lambda_1$  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  $\Lambda_1$  is equal to zero. In particular, let  $\lambda_1^{i'}$  be the  $i$ -th row of  $\Lambda_1$ . Then, under the null that  $\lambda_1^i = 0_{k \times 1}$ , the Wald statistic

$$W_{\lambda_1^i} = \hat{\lambda}_1^{i'} \hat{\mathcal{V}}_{\lambda_1^i}^{-1} \hat{\lambda}_1^i \stackrel{a}{\sim} \chi^2(k) \quad (17)$$

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  $\hat{\lambda}_1^{i'}$  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  $\lambda_0$  and  $\Lambda_1$  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

<sup>16</sup>We use the equity return data from the website of Kenneth French and *cay* from the website of Sydney Ludvigson.

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  $\lambda_0$  and  $\Lambda_1$  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  $\Lambda_1$  being equal to zero. The sample period is 1952Q1-2010Q2.

	$\lambda_0$	$\Lambda_1^{R^M}$	$\Lambda_1^{R^f}$	$\Lambda_1^{R^f \cdot cay}$	$W_{\Lambda_1}$
<b>OLS</b>					
$R^M$	<b>2.274</b> (1.088)	0.055 (0.066)	-0.801 (0.769)	<b>0.211</b> (0.070)	<b>11.176</b> (0.011)
$R_f$	-0.078 (0.094)	<b>-0.012</b> (0.006)	-0.051 (0.058)	<b>0.017</b> (0.006)	<b>9.310</b> (0.025)
$R_f \cdot cay$	<b>3.457</b> (1.662)	<b>0.265</b> (0.111)	0.689 (1.037)	-0.057 (0.111)	5.736 (0.125)
<b>FGLS</b>					
$R^M$	<b>2.476</b> (1.079)	0.063 (0.066)	-0.845 (0.765)	<b>0.230</b> (0.070)	<b>13.452</b> (0.004)
$R_f$	-0.026 (0.077)	-0.005 (0.005)	-0.070 (0.050)	<b>0.020</b> (0.006)	<b>13.169</b> (0.004)
$R_f \cdot cay$	<b>2.992</b> (1.400)	<b>0.332</b> (0.098)	0.587 (0.914)	-0.142 (0.103)	<b>12.132</b> (0.007)
<b>LML</b>					
$R^M$	<b>2.476</b> (1.078)	0.063 (0.065)	-0.845 (0.765)	<b>0.230</b> (0.070)	<b>13.466</b> (0.004)
$R_f$	-0.026 (0.075)	-0.005 (0.005)	-0.070 (0.049)	<b>0.020</b> (0.006)	<b>13.902</b> (0.003)
$R_f \cdot cay$	<b>2.992</b> (1.298)	<b>0.332</b> (0.091)	0.587 (0.849)	-0.142 (0.096)	<b>14.195</b> (0.003)

for tests of whether rows of  $\Lambda_1$  are significantly different from zero.

There are two takeaways from Table 1. First, the upper-right element of  $\Lambda_1$  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  $\Lambda_1$  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  $\Lambda_1$  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  $\frac{1}{T-1} \sum_{t=1}^{T-1} (|\hat{\alpha}_{i,t+1}| - |\hat{\alpha}_{i,t+1}^{FM}|)$  where

$$\hat{\alpha}_{i,t+1} = R_{i,t+1}^e - \hat{\beta}_i' \hat{\Lambda} Z_t \quad (18)$$

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 counterparts,  $OLS^\mu$ ,  $FGLS^\mu$ ,  $4FGLS^\mu$  and  $LML^\mu$ . Here, FGLS denotes the pricing errors corresponding to  $\hat{B} = \hat{B}_{ols}$  and  $\hat{\Lambda} = \hat{\Lambda}_{fgls}$  while 4FGLS denotes the pricing errors corresponding to  $\hat{B} = \hat{B}_{4fgls}$  and  $\hat{\Lambda} = \hat{\Lambda}_{4fgls}$  (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  $\tilde{X}_t = (R_t^M, cay_{t-1}, R_t^M \cdot cay_{t-1})'$ .

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  $LML^\mu$  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 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).

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  $\frac{1}{T-1} \sum_{t=1}^{T-1} |\hat{\alpha}_{i,t+1}| - |\hat{\alpha}_{i,t+1}^{FM}|$  where  $\hat{\alpha}_{i,t+1} = R_{i,t+1}^e - \beta_i' \hat{\Lambda} Z_t$  is the conditional pricing error implied by the respective estimator and  $\hat{\alpha}_{i,t+1}^{FM}$  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.  $OLS^\mu$  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;  $LML^\mu$  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)
	$OLS^\mu$	$OLS$	$FGLS^\mu$	$FGLS$	$4FGLS^\mu$	$4FGLS$	$LML^\mu$	$LML$
<b>FF11</b>	-0.169	-0.248	-0.144	-0.187	-0.207	-0.215	-0.204	-0.239
<b>FF12</b>	0.020	-0.031	0.024	-0.027	0.010	-0.033	0.010	-0.038
<b>FF13</b>	0.018	-0.092	0.017	-0.092	0.016	-0.079	0.016	-0.082
<b>FF14</b>	-0.037	-0.188	-0.035	-0.186	-0.043	-0.178	-0.043	-0.177
<b>FF15</b>	-0.026	-0.146	-0.025	-0.143	-0.022	-0.130	-0.023	-0.134
<b>FF21</b>	-0.098	-0.201	-0.093	-0.177	-0.103	-0.210	-0.102	-0.206
<b>FF22</b>	0.003	0.015	0.004	0.013	-0.001	-0.051	-0.001	-0.042
<b>FF23</b>	-0.032	-0.170	-0.033	-0.175	-0.055	-0.172	-0.054	-0.172
<b>FF24</b>	-0.004	-0.152	-0.004	-0.151	-0.005	-0.158	-0.005	-0.159
<b>FF25</b>	0.017	-0.078	0.012	-0.083	0.003	-0.089	0.004	-0.089
<b>FF31</b>	-0.014	-0.170	-0.013	-0.194	-0.013	-0.192	-0.013	-0.193
<b>FF32</b>	0.006	-0.151	0.003	-0.149	-0.003	-0.152	-0.003	-0.152
<b>FF33</b>	-0.036	-0.104	-0.038	-0.105	-0.047	-0.105	-0.047	-0.105
<b>FF34</b>	-0.007	-0.167	-0.007	-0.167	-0.019	-0.183	-0.019	-0.181
<b>FF35</b>	-0.012	-0.043	-0.016	-0.047	-0.018	-0.059	-0.018	-0.056
<b>FF41</b>	0.050	-0.121	0.029	-0.162	0.009	-0.174	0.010	-0.176
<b>FF42</b>	-0.069	-0.161	-0.080	-0.163	-0.056	-0.159	-0.057	-0.159
<b>FF43</b>	0.003	-0.146	-0.001	-0.149	-0.004	-0.150	-0.004	-0.150
<b>FF44</b>	-0.004	-0.129	-0.004	-0.130	-0.006	-0.133	-0.006	-0.133
<b>FF45</b>	0.013	-0.134	0.014	-0.120	0.001	-0.134	0.001	-0.134
<b>FF51</b>	0.037	-0.287	0.022	-0.314	0.024	-0.313	0.024	-0.316
<b>FF52</b>	0.059	-0.212	0.038	-0.247	0.021	-0.261	0.022	-0.259
<b>FF53</b>	0.008	-0.228	-0.004	-0.223	-0.013	-0.218	-0.012	-0.218
<b>FF54</b>	0.001	-0.141	0.000	-0.140	0.002	-0.139	0.002	-0.139
<b>FF55</b>	0.008	-0.156	0.000	-0.169	-0.007	-0.193	-0.007	-0.187
<b>Avg</b>	<b>-0.011</b>	<b>-0.146</b>	<b>-0.013</b>	<b>-0.148</b>	<b>-0.021</b>	<b>-0.155</b>	<b>-0.021</b>	<b>-0.156</b>

## 6.2 Campbell (1996)

In this section, we illustrate the usefulness of the proposed estimators for the exponentially affine pricing model by applying it to the version of the ICAPM suggested by Campbell (1996). Campbell combines the log-linear approximation of a representative agents' budget constraint with a log-linear solution to the consumption Euler equation implied by an Epstein-Zin-Weil utility specification in order to derive a reduced form expected log excess return equation of the form

$$\mathbb{E} [r_{i,t+1}^e | \mathcal{F}_t] + \frac{1}{2} V_i = \beta_i' \lambda_0 \quad (19)$$

where  $V_i = \mathbb{V} \left( r_{i,t+1}^e - \mathbb{E} \left[ r_{i,t+1}^e | \mathcal{F}_t \right] \right)$  is the unconditional variance of returns,  $\beta_i$  denotes a vector of risk factor exposures for asset  $i$  and  $\lambda_0$  a vector of factor risk premia. In Campbell's approxi-

mate solution, equilibrium expected returns have a multi factor structure. Prices of risk, however, are constant functions of preference parameters and the parameters governing the state variable dynamics. In the baseline specification, Campbell (1996) selects five factors that forecast future returns, or future labor income: the real return on the stock market index ( $RVW$ ), real labor income growth ( $LBR$ ), the dividend yield on the aggregate stock market index ( $DIV$ ), the yield spread between long- and short-term government bonds ( $TRM$ ), and the “relative bill rate” ( $RTB$ ) which is given by the difference between the 1-month Treasury bill rate and its 1-year backward moving average. Campbell’s unrestricted expected return specification from equation (19) is a special case of our exponentially affine model (15) when  $\Lambda_1 = 0$ .<sup>17</sup> Using the estimators derived in Section 5, we can thus estimate a version of Campbell’s model which explicitly allows for time varying market prices of risk.<sup>18</sup> Moreover, given the limiting variances of the estimators, we can test whether the restriction that market prices of risk are constant is borne out by the data.

Before discussing the results, it is important to note that our estimation approach requires the matrix  $B$  to have full row rank (see Assumption 1 (b)). This implies that none of the pricing factors is redundant, uninformative or unspanned. A factor is redundant if its risk factor exposures are linear combinations of the risk factor exposures of the remaining factors. Uninformative factors feature zero risk exposures for all assets and do not predict returns. Unspanned factors also carry zero factor risk exposures but do drive time variation in risk premia. If the rank condition on  $B$  is not satisfied, we cannot identify  $\Lambda$ .<sup>19</sup> Therefore, before interpreting market price of risk estimates, we test whether the estimated risk factor exposures  $\beta_i$  are jointly statistically different from zero. This can be done using a Wald-type test for rows of  $B$  analogous to equation (17). When we compute such test statistics for the rows of the estimated matrix  $\hat{B}$  in the Campbell model, we find that all factors except the real labor income growth  $LBR$  carry strongly significant factor risk exposures across assets. For labor income growth the  $OLS$  and  $FGLS$  estimators do not provide statistical support for the hypothesis that its betas are significantly different from zero, while the  $LML$  estimator rejects the null of all betas associated with  $LBR$  being jointly zero at the 10% confidence level. Given these findings, we therefore exclude labor income growth from the model and report results based on a model specification featuring the four factors,  $RVW$ ,  $DIV$ ,  $TRM$  and  $RTB$ .

We estimate the model using a cross section of size sorted equity, corporate bond, and Trea-

---

<sup>17</sup>Note that in our model  $V_i = \mathbb{V}(r_{i,t+1}^e - E_t r_{i,t+1}^e) = \mathbb{V}(\beta'_{i,t} v_{t+1} + e_{i,t+1}) = \text{vec}(\beta_i \beta'_i)' \text{vec}(\Sigma_v) + \sigma_i^2$ , i.e., the Jensen adjustment term features squared risk factor exposures. This arises due to the fact that we decompose return innovations into a spanned part, i.e., a component that is proportional to pricing factor innovations, and an idiosyncratic return component.

<sup>18</sup>We construct  $RVW$  as the value weighted excess market return from CRSP in excess of monthly CPI inflation. We follow Jagannathan and Wang (1996) in constructing  $LBR$  using labor income per capita from the Bureau of Economic Analysis’ NIPA Table 2.6. Furthermore,  $DIV$  is the dividend yield of the CRSP market portfolio,  $TRM$  is the difference between the 10-year and the 3-month constant maturity Treasury yields as reported in the Federal Reserve Board’s H.15 Release. Finally, the relative bill rate  $RTB$  is constructed using the 1-month Treasury bill rate obtained from Kenneth French’s website.

<sup>19</sup>A number of recent papers discuss the identifiability of factor risk premia in static Fama-MacBeth regressions, see e.g., Burnside (2009), Burnside (2010), Kleibergen (2009), and Kleibergen (2010).

sury bond portfolios as test assets, following Campbell (1996).<sup>20</sup> Table 3 provides estimates and corresponding standard errors for the market price of risk parameters for the three estimators of the exponentially affine specification: OLS, FGLS, and LML. The last column again provides the Wald statistics and corresponding  $p$ -values for the hypothesis that rows of  $\Lambda_1$  are zero.

While the three estimators deliver similar point estimates, the FGLS and LML variants generally feature smaller standard errors and both give rise to the same conclusions about the significance of individual parameters. Since the LML estimator is asymptotically efficient, we will focus the interpretation of the results on the third panel of Table 3. According to these, the three factors  $RVW$ ,  $DIV$ , and  $RTB$  feature a significant constant component  $\lambda_0$  across all estimators, indicating that the risks they represent are priced unconditionally. In addition, in the case of FGLS,  $TRM$  is also priced unconditionally. Second, for the factors  $RVW$ ,  $DIV$ , and  $RTB$  we reject the null hypothesis that the corresponding rows of  $\Lambda_1$  are equal to zero at all conventional confidence levels. Hence, we find strong statistical support for time variation of risk premia in our version of Campbell's ICAPM model. Looking at the individual elements of  $\Lambda_1$ , we see that  $RVW$ ,  $DIV$ , and  $TRM$  contribute to this time variation while the relative TBill rate,  $RTB$ , does not seem to affect intertemporal risk pricing.

We again assess the economic significance of the estimates based on conditional pricing errors. To facilitate comparison, we report absolute pricing errors of the 20 test assets in excess of the absolute pricing errors of the OLS constant price of risk specification. That is, Table 4 reports the quantity  $\frac{1}{T-1} \sum_{t=1}^{T-1} (|\hat{\alpha}_{i,t+1}| - |\hat{\alpha}_{i,t+1}^{OLS^\mu}|)$  where

$$\hat{\alpha}_{i,t+1} = \left| R_{i,t+1}^e - \hat{\beta}_i' \hat{\Lambda} Z_t + \frac{1}{2} \text{vec} \left( \hat{\beta}_i \hat{\beta}_i' \right)' \text{vec} \left( \hat{\Sigma}_v \right) + \frac{1}{2} \hat{\sigma}_i^2 \right| \quad (20)$$

is the absolute value of the conditional pricing error implied by the respective estimator.<sup>21</sup>

Several conclusions emerge from Table 4. First, as columns 1, 3 and 5 show—for almost all portfolios—pricing errors for the time varying price of risk specifications are smaller than for the constant price of risk specification. Moreover, for many assets, this difference is sizable, with the average reduction implied by the OLS estimator being about 10 basis points per month and the individual improvement reaching 28 basis points per month for the lowest size decile equity portfolio. Hence, allowing for time variation in prices of risk results in considerably more precise predictions of excess returns. Second, independently of the price of risk specification chosen, use of the LML estimator generally lowers pricing errors quite substantially with respect to the OLS and FGLS estimators. Indeed, as the last two columns of Table 4 show, the LML estimators reduce conditional pricing errors by more than 25 basis points per month on average across all assets in the

<sup>20</sup>Specifically, we use 10 size sorted equity portfolios obtained from Kenneth French's website, five industry and rating sorted corporate bond portfolios obtained from Barclay's, and five constant maturity government bond portfolios obtained from CRSP.

<sup>21</sup>Note that  $\Lambda = \lambda_0$  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.

Table 3: Campbell's ICAPM - Market Price of Risk Parameter Estimates

This table reports coefficient estimates and the corresponding standard errors for the market price of risk parameters  $\lambda_0$  and  $\Lambda_1$  in the Campbell ICAPM with *RVW*, *DIV*, *TRM*, and *RTB* 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  $\Lambda_1$  being equal to zero. The sample period is 1973:01-2010:12.

	$\lambda_0$	$\Lambda_1^{RVW}$	$\Lambda_1^{DIV}$	$\Lambda_1^{TRM}$	$\Lambda_1^{RTB}$	$W_{\Lambda_1}$
<b>OLS</b>						
RVW	<b>12.241</b> (1.158)	0.100 (0.053)	0.329 (0.177)	0.416 (0.227)	0.798 (2.443)	9.387 (0.052)
DIV	<b>-0.720</b> (0.316)	-0.012 (0.016)	<b>-0.039</b> (0.019)	<b>-0.082</b> (0.027)	-0.086 (0.218)	<b>10.908</b> (0.028)
TRM	1.524 (1.031)	0.062 (0.051)	0.039 (0.062)	0.077 (0.087)	-0.037 (0.706)	1.836 (0.766)
RTB	<b>-1.093</b> (0.446)	<b>-0.054</b> (0.022)	-0.034 (0.027)	-0.034 (0.037)	0.102 (0.300)	6.003 (0.199)
<b>FGLS</b>						
RVW	<b>11.699</b> (1.079)	<b>0.100</b> (0.050)	0.306 (0.177)	0.418 (0.226)	0.980 (2.436)	<b>9.660</b> (0.047)
DIV	<b>-0.687</b> (0.201)	-0.012 (0.011)	<b>-0.034</b> (0.016)	<b>-0.096</b> (0.022)	-0.192 (0.174)	<b>20.991</b> (0.000)
TRM	<b>1.175</b> (0.500)	0.046 (0.026)	<b>0.081</b> (0.040)	0.005 (0.054)	-0.330 (0.433)	5.514 (0.239)
RTB	<b>-0.887</b> (0.286)	<b>-0.047</b> (0.015)	<b>-0.046</b> (0.022)	-0.019 (0.030)	0.111 (0.233)	<b>10.088</b> (0.039)
<b>LML</b>						
RVW	<b>10.709</b> (0.997)	<b>0.100</b> (0.050)	0.307 (0.166)	<b>0.419</b> (0.212)	0.978 (2.270)	<b>10.203</b> (0.037)
DIV	<b>-0.945</b> (0.184)	-0.013 (0.011)	<b>-0.035</b> (0.016)	<b>-0.096</b> (0.021)	-0.192 (0.166)	<b>22.682</b> (0.000)
TRM	0.788 (0.535)	0.051 (0.031)	0.087 (0.046)	0.012 (0.060)	-0.324 (0.478)	4.600 (0.331)
RTB	<b>-0.714</b> (0.278)	<b>-0.051</b> (0.016)	<b>-0.051</b> (0.023)	-0.024 (0.031)	0.107 (0.237)	<b>9.829</b> (0.043)

constant price of risk specification and by 33 basis points per month in the dynamic specification. On an annualized basis, this amounts to a cross sectional average reduction of conditional pricing errors of more than 3 percent. Hence, even in this relatively small sample, the gains from using the asymptotically efficient LML estimator are economically important. We would like to note that in some empirical applications, the FGLS and LML estimators give rise to unusually large pricing errors for some assets. This appears to be driven by the asset specific weighting of parameters implied by these two estimators as the OLS estimator does not have this feature. This notwithstanding, the results from our illustration of the exponentially affine pricing kernel specification and the corresponding estimators show that time variation in prices of risk is pervasive and that explicitly allowing for it considerably reduces conditional pricing errors.

## 7 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



Table 4: Campbell's ICAPM - Mean Absolute Conditional Pricing Errors

This table reports time series averages of absolute conditional pricing errors implied by our four factor version of Campbell's ICAPM. The set of test assets comprises 10 size sorted stock portfolios obtained from Ken French's website, five industry and rating sorted credit portfolios obtained from Barclay's Capital, and five constant maturity Treasury portfolios obtained from CRSP. All returns are in excess of the one month Tbill also obtained from Ken French's website. Six different estimators are considered. *OLS* denotes the OLS estimator of the exponentially affine pricing model with time varying prices of risk; *FGLS<sup>μ</sup>* and *FGLS* denote the FGLS estimators of the exponentially affine pricing model with constant and time varying prices of risk, respectively; *LML<sup>μ</sup>* and *LML* denote the linearized maximum likelihood estimators of the exponentially affine pricing model with constant and time varying prices of risk, respectively. The sample period is 1952Q1-2010Q2.

	(1)	(2)	(3)	(4)	(5)
	<i>OLS</i>	<i>FGLS<sup>μ</sup></i>	<i>FGLS</i>	<i>LML<sup>μ</sup></i>	<i>LML</i>
<b>size1</b>	-0.279	0.566	0.039	-0.055	-0.296
<b>size2</b>	-0.191	0.656	0.303	-0.314	-0.651
<b>size3</b>	-0.112	0.309	0.144	-0.079	-0.220
<b>size4</b>	-0.129	0.259	0.087	-0.140	-0.279
<b>size5</b>	-0.083	0.187	0.072	-0.022	-0.119
<b>size6</b>	-0.057	0.030	-0.030	0.003	-0.058
<b>size7</b>	-0.066	-0.239	-0.273	-0.302	-0.346
<b>size8</b>	-0.048	0.083	0.029	-0.028	-0.065
<b>size9</b>	0.007	-0.044	-0.063	-0.051	-0.048
<b>size10</b>	0.018	-0.044	-0.033	-0.058	-0.049
<b>igind</b>	-0.015	0.036	-0.003	-0.030	-0.039
<b>igutil</b>	-0.028	0.002	-0.030	-0.017	-0.031
<b>igfin</b>	-0.090	-0.066	-0.132	-0.261	-0.263
<b>aaa</b>	-0.021	-0.000	-0.021	0.004	-0.014
<b>baa</b>	-0.212	-0.216	-0.355	-0.996	-1.031
<b>cmt1</b>	-0.034	-0.057	-0.066	-0.088	-0.093
<b>cmt2</b>	-0.004	0.014	0.002	-0.014	-0.017
<b>cmt5</b>	-0.167	-0.179	-0.281	-0.773	-0.871
<b>cmt7</b>	-0.175	-0.184	-0.283	-0.763	-0.844
<b>cmt10</b>	-0.250	-0.356	-0.441	-1.069	-1.205
<b>Avg</b>	<b>-0.097</b>	<b>0.038</b>	<b>-0.067</b>	<b>-0.253</b>	<b>-0.327</b>

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 illustrations, we have revisited two influential asset pricing models that may be cast as special cases of our generic approach. We find support that time variation of risk premia is present in these models, 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 Affine Model: 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  $(\hat{\Lambda}_{\text{gls}}^\mu, \hat{B}_{4\text{gls}}^\mu)$  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  $T^{-1} \cdot \hat{E}_{\text{ols}} M_t \hat{E}'_{\text{ols}}$  where  $M_t = I_T - T^{-1} \cdot \iota_T \iota_T'$  (see Step 3b in the discussion of Theorem 6).

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

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

### A.2 Affine Model: Maximum Likelihood Estimation when $\Sigma_e$ is Unrestricted

When we relax the diagonality assumption of Assumption 2 (b) and instead only assume that  $\Sigma_e$  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  $(\Psi, B, \Lambda)$ ,  $\hat{\Sigma}_{e,\text{mle}} = \frac{EE'}{T}$  and given  $\Lambda$  and  $\Psi$ ,

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

Thus, if we concentrate the likelihood with respect to  $\Sigma_e$ ,  $B$ , 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,

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

where  $M_R = I_T - R'_e (R_e R'_e)^{-1} R_e$ ,  $D = [I_K \ \Lambda]'$  and  $Y$  is the  $(2K+1) \times T$  matrix  $Y = [\hat{V}'_{\text{ols}} \ Z'_-]'$ . Next, if we define  $G_1 = Y M_R Y'$ ,  $G_2 = Y Y'$  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  $\{\mu_i : i = 1, \dots, K\}$  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,  $\hat{D} = (a_1, \dots, a_K)$ , where  $a_i$  is the eigenvector associated with the eigenvalue  $\mu_i$ . 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  $\hat{D} (\hat{D}_1)^{-1}$  which implies that

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

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

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

### A.3 Exponentially Affine Model: Asymptotic Properties of Regression Estimators

#### A.3.1 Time Varying Prices of Risk

The expressions for the limiting variance of  $\hat{\lambda}_{\text{ols}}^\epsilon$  and  $\hat{\lambda}_{\text{fgls}}^\epsilon$  of Theorem 6 are given by,

$$\mathcal{V}_{\hat{\lambda}_{\text{ols}}^\epsilon} = \mathcal{V}_{\hat{\lambda}_{\text{ols}}} + (\Upsilon^{-1} \Upsilon_1 \Upsilon'_1 \Upsilon^{-1} \otimes Q_{\text{ols}}^\epsilon) + \mathcal{C}_{\text{ols}}^\epsilon + \mathcal{C}_{\text{ols}}^{\epsilon'}$$
 (21)

$$\mathcal{V}_{\hat{\lambda}_{\text{fgls}}^\epsilon} = \mathcal{V}_{\hat{\lambda}_{\text{fgls}}} + (\Upsilon^{-1} \Upsilon_1 \Upsilon'_1 \Upsilon^{-1} \otimes Q_{\text{fgls}}^\epsilon) + \mathcal{C}_{\text{fgls}}^\epsilon + \mathcal{C}_{\text{fgls}}^{\epsilon'}$$
 (22)

where

$$\Xi_1 = \text{bdiag}_{[N, N], [1, 1]} (\sigma_i^2 \cdot \beta'_i \Sigma_v \beta_i), \quad \Xi_2 = \text{bdiag}_{[KN, N], [K, 1]} (\sigma_i^2 \cdot \Sigma_v \beta_i),$$

$$\begin{aligned} Q_{\text{ols}}^\epsilon &= \frac{1}{2} (BB')^{-1} B \Sigma_e^2 B' (BB')^{-1} + (BB')^{-1} B \Xi_1 B' (BB')^{-1} \\ &\quad + \frac{1}{4} (BB')^{-1} B J_B (I_{K^2} + \kappa_{KK}) (\Sigma_v \otimes \Sigma_v) J'_B B' (BB')^{-1} \\ \mathcal{C}_{\text{ols}}^\epsilon &= - \left( \Lambda' \Sigma_v^{-1} \otimes (BB')^{-1} B \right) \kappa_{KN} \Xi_2 \left( \Upsilon'_1 \Upsilon^{-1} \otimes B' (BB')^{-1} \right), \\ Q_{\text{fgls}}^\epsilon &= \frac{1}{2} (B \Sigma_e^{-1} B')^{-1} B B' (B \Sigma_e^{-1} B')^{-1} + (B \Sigma_e^{-1} B')^{-1} B \Sigma_e^{-1} \Xi_1 \Sigma_e^{-1} B' (B \Sigma_e^{-1} B')^{-1} \\ &\quad + \frac{1}{4} (B \Sigma_e^{-1} B')^{-1} B \Sigma_e^{-1} J_B (I_{K^2} + \kappa_{KK}) (\Sigma_v \otimes \Sigma_v) J'_B \Sigma_e^{-1} B' (B \Sigma_e^{-1} B')^{-1} \\ \mathcal{C}_{\text{fgls}}^\epsilon &= - \left( \Lambda' \Sigma_v^{-1} \otimes (B \Sigma_e^{-1} B')^{-1} B \Sigma_e^{-1} \right) \kappa_{KN} \Xi_2 \left( \Upsilon'_1 \Upsilon^{-1} \otimes \Sigma_e^{-1} B' (B \Sigma_e^{-1} B')^{-1} \right). \end{aligned}$$

### A.3.2 Constant Prices of Risk

Define the analogous estimators to those of Theorem 6 as,

$$\begin{aligned}\hat{\Lambda}_{\text{ols}}^{\epsilon,\mu} &= \left(\hat{B}_{\text{ols}}\hat{B}'_{\text{ols}}\right)^{-1} \hat{B}_{\text{ols}} \left[\bar{r}_e^{\epsilon,\mu} + \frac{1}{2}\hat{J}_{B,\text{ols}} \text{vec}(\hat{\Sigma}_{v,\text{ols}}) + \frac{1}{2}\hat{d}_{e,\text{sur}}\right], \\ \hat{\Lambda}_{\text{fgls}}^{\epsilon,\mu} &= \left(\hat{B}_{\text{ols}}\hat{\Sigma}_{e,\text{ols}}^{-1}\hat{B}'_{\text{ols}}\right)^{-1} \hat{B}_{\text{ols}}\hat{\Sigma}_{e,\text{ols}}^{-1} \left[\bar{r}_e^{\epsilon,\mu} + \frac{1}{2}\hat{J}_{B,\text{ols}} \text{vec}(\hat{\Sigma}_{v,\text{ols}}) + \frac{1}{2}\hat{d}_{e,\text{sur}}\right],\end{aligned}$$

where  $\bar{r}_e^{\epsilon,\mu} = r_e^{\epsilon,\mu} \iota_T / T$  and  $\hat{\Sigma}_{e,\text{ols}}$  is constructed assuming constant prices of risk. These estimators satisfy

$$\sqrt{T}(\hat{\lambda}_{\text{ols}}^{\epsilon,\mu} - \lambda) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\hat{\lambda},\text{ols}}^{\epsilon,\mu}), \quad \sqrt{T}(\hat{\lambda}_{\text{fgls}}^{\epsilon,\mu} - \lambda) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\hat{\lambda},\text{fgls}}^{\epsilon,\mu}).$$

The asymptotic variance formulas are,

$$\begin{aligned}\mathcal{V}_{\hat{\lambda},\text{ols}}^{\epsilon,\mu} &= \mathcal{V}_{\hat{\lambda},\text{ols}}^{\mu} + Q_{\text{ols}}^{\epsilon,\mu} + C_{\text{ols}}^{\epsilon,\mu} + C_{\text{ols}}^{\epsilon,\mu'}, \\ \mathcal{V}_{\hat{\lambda},\text{fgls}}^{\epsilon,\mu} &= \mathcal{V}_{\hat{\lambda},\text{fgls}}^{\mu} + Q_{\text{fgls}}^{\epsilon,\mu} + C_{\text{fgls}}^{\epsilon,\mu} + C_{\text{fgls}}^{\epsilon,\mu'}\end{aligned}$$

where  $Q_{\text{ols}}^{\epsilon,\mu} = Q_{\text{ols}}^{\epsilon}$  and  $Q_{\text{fgls}}^{\epsilon,\mu} = Q_{\text{fgls}}^{\epsilon}$  and

$$\begin{aligned}C_{\text{ols}}^{\epsilon,\mu} &= -\left(\lambda'_0 \Sigma_v^{-1} \otimes (BB')^{-1} B\right) \kappa_{KN} \Xi_2 B' (BB')^{-1}, \\ C_{\text{fgls}}^{\epsilon,\mu} &= -\left(\lambda'_0 \Sigma_v^{-1} \otimes (B\Sigma_e^{-1} B')^{-1} B\Sigma_e^{-1}\right) \kappa_{KN} \Xi_2 \Sigma_e^{-1} B' (B\Sigma_e^{-1} B')^{-1}.\end{aligned}$$

## A.4 Exponentially Affine Model: Score Vector and Information Matrix

### A.4.1 Time Varying Prices of Risk

The (scaled) score vector,  $\dot{\ell}^{\epsilon} = \dot{\ell}^{\epsilon}(\theta^{\epsilon})$ , is the  $(2K(K+1) + NK + K^2 + N) \times 1$  vector with elements,

$$\begin{aligned}\left[\dot{\ell}^{\epsilon}\right]_1 &= \frac{\partial \ell^{\epsilon}(\theta^{\epsilon})}{\partial \psi} = T^{-1} \cdot [\text{vec}(\Sigma_v^{-1} V Z'_-) - \text{vec}(B\Sigma_e^{-1} E^{\epsilon} Z'_-)] \\ \left[\dot{\ell}^{\epsilon}\right]_2 &= \frac{\partial \ell^{\epsilon}(\theta^{\epsilon})}{\partial b} = T^{-1} \cdot [\text{vec}((\Lambda Z_- + V) E^{\epsilon'} \Sigma_e^{-1}) - \text{vec}(\Sigma_v B \Sigma_e^{-1}) \odot (E^{\epsilon} \iota_T \otimes \iota_K)] \\ \left[\dot{\ell}^{\epsilon}\right]_3 &= \frac{\partial \ell^{\epsilon}(\theta^{\epsilon})}{\partial \lambda} = T^{-1} \cdot \text{vec}(B\Sigma_e^{-1} E^{\epsilon} Z'_-) \\ \left[\dot{\ell}^{\epsilon}\right]_4 &= \frac{\partial \ell^{\epsilon}(\theta^{\epsilon})}{\partial \sigma_v} = T^{-1} \cdot \frac{1}{2} [\text{vec}(-T \cdot \Sigma_v^{-1} + \Sigma_v^{-1} V V' \Sigma_v^{-1}) - J'_B \Sigma_e^{-1} E^{\epsilon} \iota_T] \\ \left[\dot{\ell}^{\epsilon}\right]_5 &= \frac{\partial \ell^{\epsilon}(\theta^{\epsilon})}{\partial d_e} = T^{-1} \cdot \left[-\frac{T}{2} d_{e,\text{inv}} + \frac{1}{2} (d_{e,\text{inv}} \odot d_{e,\text{inv}}) \odot ((E^{\epsilon} \odot E^{\epsilon}) \iota_T) - \frac{1}{2} (d_{e,\text{inv}} \odot E^{\epsilon} \iota_T)\right]\end{aligned}$$

where  $d_{e,\text{inv}} = (\sigma_1^{-2}, \sigma_2^{-2}, \dots, \sigma_N^{-2})'$  and  $E^\epsilon = E(\Psi, B, \Lambda, \Sigma_v, d_e)$ ,  $V = V(\Psi)$  (see equations (7) and (16)). The information matrix,  $\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)$ , is the square, symmetric matrix with elements,

$$\begin{aligned} [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{11} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \psi \partial \psi'} = (\Upsilon \otimes (B \Sigma_e^{-1} B' + \Sigma_v^{-1})) \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{12} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \psi \partial b'} = \left[ \sigma_1^{-2} ((\Upsilon_1 \beta_1' \Sigma_v - \Upsilon \Lambda') \otimes \beta_1) \quad \dots \quad \sigma_N^{-2} ((\Upsilon_1 \beta_N' \Sigma_v - \Upsilon \Lambda') \otimes \beta_N) \right] \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{13} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \psi \partial \lambda'} = -(\Upsilon \otimes B \Sigma_e^{-1} B') \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{14} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \psi \partial \sigma_v'} = \frac{1}{2} (\Upsilon_1 \otimes B \Sigma_e^{-1} J_B) \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{15} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \psi \partial d_e'} = \frac{1}{2} (\Upsilon_1 \otimes B \Sigma_e^{-1}) \end{aligned}$$

and

$$\begin{aligned} [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{22} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial b \partial b'} = \text{bdiag}_{[KN, KN], [K, K]} (\sigma_i^{-2} [\Lambda \Upsilon \Lambda' + \Sigma_v + \Sigma_v \beta_i \beta_i' \Sigma_v - \Sigma_v \beta_i \Upsilon_1' \Lambda' - \Lambda \Upsilon_1 \beta_i' \Sigma_v]) \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{23} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial b \partial \lambda'} = -([\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{12})' \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{24} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial b \partial \sigma_v'} = \begin{bmatrix} \frac{1}{2} \sigma_1^{-2} ((\Sigma_v \beta_1 - \Lambda \Upsilon_1) \beta_1' \otimes \beta_1') \\ \vdots \\ \frac{1}{2} \sigma_N^{-2} ((\Sigma_v \beta_N - \Lambda \Upsilon_1) \beta_N' \otimes \beta_N') \end{bmatrix} \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{25} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial b \partial d_e'} = \text{bdiag}_{[KN, N], [K, 1]} \left( \frac{1}{2 \sigma_i^2} (\Sigma_v \beta_i - \Lambda \Upsilon_1) \right) \end{aligned}$$

and

$$\begin{aligned} [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{33} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \lambda \partial \lambda'} = -[\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{13} \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{34} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \lambda \partial \sigma_v'} = -[\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{14} \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{35} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \lambda \partial d_e'} = -[\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{15} \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{44} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \sigma_v \partial \sigma_v'} = \frac{1}{2} (\Sigma_v^{-1} \otimes \Sigma_v^{-1}) + \frac{1}{4} J_B' \Sigma_e^{-1} J_B \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{45} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \sigma_v \partial d_e'} = \frac{1}{4} J_B' \Sigma_e^{-1} \\ [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{55} &= \frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial d_e \partial d_e'} = \frac{1}{4} \text{bdiag}_{[N, N], [1, 1]} \left( 2 (\sigma_i^{-2})^2 + \sigma_i^{-2} \right) \end{aligned}$$

In the exponentially affine case, the form of the inverse information matrix is not as straightforward as in the affine case. However, we can still characterize the inverse information matrix. For simplicity of notation define  $\mathcal{I}_{ab}^\epsilon = [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)]_{ab}$  and  $\mathcal{H}_{ab}^\epsilon = [\mathcal{I}^\epsilon(\theta^\epsilon; \Upsilon)^{-1}]_{ab}$ . The first row has the elements,

$$\mathcal{H}_{11}^\epsilon = (\Upsilon^{-1} \otimes \Sigma_v), \quad \mathcal{H}_{12}^\epsilon = 0_{K(K+1) \times NK}, \quad \mathcal{H}_{13}^\epsilon = \mathcal{H}_{11}^\epsilon,$$

$$\mathcal{H}_{14}^\epsilon = 0_{K(K+1) \times K^2}, \quad \mathcal{H}_{15}^\epsilon = 0_{K(K+1) \times N}.$$

Next define  $\mathcal{S}_{ab} = \mathcal{I}_{ab}^\epsilon - \mathcal{I}_{1a}^{\epsilon'} (\mathcal{I}_{33}^\epsilon)^{-1} \mathcal{I}_{1b}^\epsilon$  along with

$$\begin{aligned} \mathcal{R}_{15} &= \mathcal{S}_{15} - \mathcal{S}_{14} \mathcal{S}_{44}^{-1} \mathcal{S}_{45}, & \mathcal{R}_{22} &= \mathcal{S}_{22} - \mathcal{S}_{24} \mathcal{S}_{44}^{-1} \mathcal{S}_{24}' \\ \mathcal{R}_{24} &= \mathcal{S}_{24} - \mathcal{S}_{25} \mathcal{S}_{55}^{-1} \mathcal{S}_{45}', & \mathcal{R}_{25} &= \mathcal{S}_{25} - \mathcal{S}_{24} \mathcal{S}_{44}^{-1} \mathcal{S}_{45} \\ \mathcal{R}_{44} &= \mathcal{S}_{44} - \mathcal{S}_{45} \mathcal{S}_{55}^{-1} \mathcal{S}_{45}', & \mathcal{R}_{55} &= \mathcal{S}_{55} - \mathcal{S}_{45}' \mathcal{S}_{44}^{-1} \mathcal{S}_{45} \end{aligned}$$

Then,

$$\begin{aligned} \mathcal{H}_{22}^\epsilon &= [\mathcal{R}_{22} - \mathcal{R}_{25} \mathcal{R}_{55}^{-1} \mathcal{R}_{25}']^{-1} & \mathcal{H}_{25}^\epsilon &= [-\mathcal{R}_{55}^{-1} \mathcal{R}_{25}' \mathcal{H}_{22}^\epsilon]^\prime \\ \mathcal{H}_{24}^\epsilon &= [-\mathcal{S}_{44}^{-1} (\mathcal{S}_{24}' \mathcal{H}_{22}^\epsilon + \mathcal{S}_{45} \mathcal{H}_{25}^{\epsilon'})]^\prime & \mathcal{H}_{23}^\epsilon &= [\mathcal{I}_{33}^{-1} (\mathcal{I}_{12} \mathcal{H}_{22}^\epsilon + \mathcal{I}_{14} \mathcal{H}_{24}^{\epsilon'} + \mathcal{I}_{15} \mathcal{H}_{25}^{\epsilon'})]^\prime \\ \mathcal{H}_{44}^\epsilon &= \mathcal{R}_{44}^{-1} (\mathcal{I}_{K^2} - \mathcal{R}_{24}' \mathcal{H}_{24}^\epsilon) & \mathcal{H}_{45}^\epsilon &= [-\mathcal{S}_{55}^{-1} (\mathcal{S}_{25}' \mathcal{H}_{24}^\epsilon + \mathcal{S}_{45}' \mathcal{H}_{44}^\epsilon)]^\prime \\ \mathcal{H}_{35}^\epsilon &= [-\mathcal{R}_{55}^{-1} (\mathcal{R}_{15}' \mathcal{H}_{11}^\epsilon + \mathcal{R}_{25}' \mathcal{H}_{23}^\epsilon)]^\prime & \mathcal{H}_{34}^\epsilon &= [-\mathcal{S}_{44}^{-1} (\mathcal{S}_{14}' \mathcal{H}_{11}^\epsilon + \mathcal{S}_{24}' \mathcal{H}_{23}^\epsilon + \mathcal{S}_{45}' \mathcal{H}_{35}^{\epsilon'})]^\prime \\ \mathcal{H}_{33}^\epsilon &= \mathcal{I}_{33}^{-1} (\mathcal{I}_{11} \mathcal{H}_{11}^\epsilon + \mathcal{I}_{12} \mathcal{H}_{23}^\epsilon + \mathcal{I}_{14} \mathcal{H}_{34}^{\epsilon'} + \mathcal{I}_{15} \mathcal{H}_{35}^{\epsilon'}) \\ \mathcal{H}_{55}^\epsilon &= \mathcal{I}_{55}^{-1} (\mathcal{I}_N - (\mathcal{I}_{25}' \mathcal{H}_{25}^\epsilon - \mathcal{I}_{15}' \mathcal{H}_{35}^\epsilon + \mathcal{I}_{45}' \mathcal{H}_{45}^\epsilon)) \end{aligned}$$

Note that the ordering of elements reflects the dependence on other elements (e.g., to obtain the (2, 3) element of  $\mathcal{H}^\epsilon$  requires the (2, 2), (2, 5), and (2, 4) elements of  $\mathcal{H}^\epsilon$ ).

#### A.4.2 Constant Prices of Risk

Define  $\theta^{\epsilon, \mu} = (\psi', b', \lambda_0', \sigma_v', d_e')'$ . The (scaled) score vector,  $\dot{\ell}^{\epsilon, \mu} = \dot{\ell}^{\epsilon, \mu}(\theta^{\epsilon, \mu})$ , is the  $(2K + N)(K + 1) \times 1$  vector with elements,

$$\begin{aligned} [\dot{\ell}^{\epsilon, \mu}]_1 &= \frac{\partial \ell^{\epsilon, \mu}(\theta^{\epsilon, \mu})}{\partial \psi} = T^{-1} \cdot [-\text{vec}(B \Sigma_e^{-1} E^{\epsilon, \mu} Z'_-) + \text{vec}(\Sigma_v^{-1} V Z'_-)] \\ [\dot{\ell}^{\epsilon, \mu}]_2 &= \frac{\partial \ell^{\epsilon, \mu}(\theta^{\epsilon, \mu})}{\partial b} = T^{-1} \cdot [\text{vec}((\lambda_0 \iota_T' + V) E^{\epsilon, \mu} \Sigma_e^{-1}) - \text{vec}(\Sigma_v B \Sigma_e^{-1}) \odot (E^{\epsilon, \mu} \iota_T \otimes \iota_K)] \\ [\dot{\ell}^{\epsilon, \mu}]_3 &= \frac{\partial \ell^{\epsilon, \mu}(\theta^{\epsilon, \mu})}{\partial \lambda_0} = T^{-1} \cdot \text{vec}(B \Sigma_e^{-1} E^{\epsilon, \mu} \iota_T) \\ [\dot{\ell}^{\epsilon, \mu}]_4 &= \frac{\partial \ell^{\epsilon, \mu}(\theta^{\epsilon, \mu})}{\partial \sigma_v} = T^{-1} \cdot \frac{1}{2} [\text{vec}(-T \cdot \Sigma_v^{-1} + \Sigma_v^{-1} V V' \Sigma_v^{-1}) - J_B' \Sigma_e^{-1} E^{\epsilon, \mu} \iota_T] \\ [\dot{\ell}^{\epsilon, \mu}]_5 &= \frac{\partial \ell^{\epsilon, \mu}(\theta^{\epsilon, \mu})}{\partial d_e} = T^{-1} \cdot \left[ -\frac{T}{2} d_{e, \text{inv}} + \frac{1}{2} (d_{e, \text{inv}} \odot d_{e, \text{inv}}) \odot ((E^{\epsilon, \mu} \odot E^{\epsilon, \mu}) \iota_T) - \frac{1}{2} (d_{e, \text{inv}} \odot E^{\epsilon, \mu} \iota_T) \right] \end{aligned}$$

where

$$E^\epsilon = E(\Psi, B, \lambda_0, \Sigma_v, d_e) = r_e^{\epsilon, \mu} + \frac{1}{2} J_B \text{vec}(\Sigma_v) \iota_T + \frac{1}{2} d_e \iota_T' - B' \lambda_0 \iota_T',$$

and  $V = V(\Psi)$  (see equation (7)). The information matrix,  $\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)$ , is the square, symmetric matrix with elements,

$$\begin{aligned} [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{11} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial \psi \partial \psi'} = (\Upsilon \otimes (B \Sigma_e^{-1} B' + \Sigma_v^{-1})) \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{12} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial \psi \partial b'} = \left[ \sigma_1^{-2} (\Upsilon_1 (\Sigma_v \beta_1 - \lambda_0)' \otimes \beta_1) \quad \cdots \quad \sigma_N^{-2} (\Upsilon_1 (\Sigma_v \beta_N - \lambda_0)' \otimes \beta_N) \right] \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{13} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial \psi \partial \lambda_0'} = -(\Upsilon_1 \otimes B \Sigma_e^{-1} B') \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{14} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial \psi \partial \sigma_v'} = \frac{1}{2} (\Upsilon_1 \otimes B \Sigma_e^{-1} J_B) \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{15} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial \psi \partial d_e'} = \frac{1}{2} (\Upsilon_1 \otimes B \Sigma_e^{-1}) \end{aligned}$$

and

$$\begin{aligned} [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{22} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial b \partial b'} = \text{bdiag}_{[KN,KN],[K,K]} (\sigma_i^{-2} [\lambda_0 \lambda_0' + \Sigma_v + \Sigma_v \beta_i \beta_i' \Sigma_v - \Sigma_v \beta_i \lambda_0' - \lambda_0 \beta_i' \Sigma_v]) \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{23} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial b \partial \lambda_0'} = \begin{bmatrix} -\sigma_1^{-2} (\Sigma_v \beta_1 - \lambda_0) \beta_1' \\ \vdots \\ -\sigma_N^{-2} (\Sigma_v \beta_N - \lambda_0) \beta_N' \end{bmatrix} \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{24} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial b \partial \sigma_v'} = \begin{bmatrix} \frac{1}{2} \sigma_1^{-2} ((\Sigma_v \beta_1 - \lambda_0) \beta_1' \otimes \beta_1') \\ \vdots \\ \frac{1}{2} \sigma_N^{-2} ((\Sigma_v \beta_N - \lambda_0) \beta_N' \otimes \beta_N') \end{bmatrix} \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{25} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial b \partial d_e'} = \text{bdiag}_{[KN,N],[K,1]} \left( \frac{1}{2 \sigma_i^2} (\Sigma_v \beta_i - \lambda_0) \right) \end{aligned}$$

and

$$\begin{aligned} [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{33} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial \lambda_0 \partial \lambda_0'} = B \Sigma_e^{-1} B' \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{34} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial \lambda_0 \partial \sigma_v'} = -\frac{1}{2} B \Sigma_e^{-1} J_B \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{35} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial \lambda_0 \partial d_e'} = -\frac{1}{2} B \Sigma_e^{-1} \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{44} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial \sigma_v \partial \sigma_v'} = \frac{1}{2} (\Sigma_v^{-1} \otimes \Sigma_v^{-1}) + \frac{1}{4} J_B' \Sigma_e^{-1} J_B \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{45} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial \sigma_v \partial d_e'} = \frac{1}{4} J_B' \Sigma_e^{-1} \\ [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{55} &= \frac{\partial^2 \ell^{\epsilon,\mu}(\theta^{\epsilon,\mu})}{\partial d_e \partial d_e'} = \frac{1}{4} \text{bdiag}_{[N,N],[1,1]} \left( 2 (\sigma_i^{-2})^2 + \sigma_i^{-2} \right) \end{aligned}$$

As for time varying prices of risk, the form of the inverse information matrix is not as straightforward as in the affine case. However, we can still characterize the inverse information matrix. For simplicity of notation

define  $\mathcal{I}_{ab}^{\epsilon,\mu} = [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)]_{ab}$  and  $\mathcal{H}_{ab}^{\epsilon,\mu} = [\mathcal{I}^{\epsilon,\mu}(\theta^{\epsilon,\mu}; \Upsilon)^{-1}]_{ab}$ . The first row has the elements,

$$\mathcal{H}_{11}^{\epsilon,\mu} = [((\Upsilon - \Upsilon_1 \Upsilon_1') \otimes B \Sigma_e^{-1} B') + (\Upsilon \otimes \Sigma_v^{-1})]^{-1}, \quad \mathcal{H}_{12}^{\epsilon,\mu} = 0_{K(K+1) \times NK}, \quad \mathcal{H}_{13}^{\epsilon,\mu} = \mathcal{H}_{11}^{\epsilon,\mu} (\Upsilon_1 \otimes I_K),$$

$$\mathcal{H}_{14}^{\epsilon,\mu} = 0_{K(K+1) \times K^2}, \quad \mathcal{H}_{15}^{\epsilon,\mu} = 0_{K(K+1) \times N}.$$

Next define  $\mathcal{S}_{ab}^{\mu} = \mathcal{I}_{ab}^{\epsilon,\mu} - \mathcal{I}_{a3}^{\epsilon,\mu} (\mathcal{I}_{33}^{\epsilon,\mu})^{-1} \mathcal{I}_{3b}^{\epsilon,\mu}$  where  $\mathcal{I}_{ab}^{\epsilon,\mu} = \mathcal{I}_{ba}^{\mu'}$  and define  $\mathcal{R}_{15}^{\mu}$ ,  $\mathcal{R}_{22}^{\mu}$ , etc. as in the previous section with  $\mathcal{S}_{ab}$  replaced by  $\mathcal{S}_{ab}^{\mu}$ . Then,

$$\begin{aligned} \mathcal{H}_{22}^{\epsilon,\mu} &= [\mathcal{R}_{22}^{\mu} - \mathcal{R}_{25}^{\mu} (\mathcal{R}_{55}^{\mu})^{-1} \mathcal{R}_{25}^{\mu}]^{-1} & \mathcal{H}_{25}^{\epsilon,\mu} &= \left( -(\mathcal{R}_{55}^{\mu})^{-1} \mathcal{R}_{25}^{\mu} \mathcal{H}_{22}^{\epsilon,\mu} \right)' \\ \mathcal{H}_{24}^{\epsilon,\mu} &= \left( -(\mathcal{S}_{44}^{\mu})^{-1} [\mathcal{S}_{24}^{\mu'} \mathcal{H}_{22}^{\epsilon,\mu} + \mathcal{S}_{45}^{\mu} \mathcal{H}_{25}^{\epsilon,\mu'}] \right)' & \mathcal{H}_{23}^{\epsilon,\mu} &= \left( -(\mathcal{I}_{33}^{\mu})^{-1} [\mathcal{I}_{23}^{\mu'} \mathcal{H}_{22}^{\epsilon,\mu} + \mathcal{I}_{34}^{\mu} \mathcal{H}_{24}^{\epsilon,\mu'} + \mathcal{I}_{35}^{\mu} \mathcal{H}_{25}^{\epsilon,\mu'}] \right)' \\ \mathcal{H}_{44}^{\epsilon,\mu} &= (\mathcal{R}_{44}^{\mu})^{-1} [I_{K^2} - \mathcal{R}_{24}^{\mu'} \mathcal{H}_{24}^{\epsilon,\mu}] & \mathcal{H}_{45}^{\epsilon,\mu} &= \left( -(\mathcal{S}_{55}^{\mu})^{-1} [\mathcal{S}_{25}^{\mu'} \mathcal{H}_{24}^{\epsilon,\mu} + \mathcal{S}_{45}^{\mu} \mathcal{H}_{44}^{\epsilon,\mu}] \right)' \end{aligned}$$

$$\mathcal{H}_{35}^{\epsilon,\mu} = \left( -(\mathcal{R}_{55}^{\mu})^{-1} [\mathcal{R}_{15}^{\mu'} \mathcal{H}_{13}^{\epsilon,\mu} + \mathcal{R}_{25}^{\mu'} \mathcal{H}_{23}^{\epsilon,\mu} + [\mathcal{I}_{35}^{\mu'} - \mathcal{S}_{45}^{\mu'} (\mathcal{S}_{44}^{\mu})^{-1} \mathcal{I}_{34}^{\mu'}] (\mathcal{I}_{33}^{\mu})^{-1}] \right)'$$

$$\mathcal{H}_{34}^{\epsilon,\mu} = \left( -(\mathcal{S}_{44}^{\mu})^{-1} [\mathcal{S}_{14}^{\mu'} \mathcal{H}_{13}^{\epsilon,\mu} + \mathcal{S}_{24}^{\mu'} \mathcal{H}_{23}^{\epsilon,\mu} + \mathcal{S}_{45}^{\mu} \mathcal{H}_{35}^{\epsilon,\mu'} + \mathcal{I}_{34}^{\mu'} (\mathcal{I}_{33}^{\mu})^{-1}] \right)'$$

$$\mathcal{H}_{33}^{\epsilon,\mu} = \mathcal{I}_{33}^{-1} [I_K - \mathcal{I}_{13}^{\mu'} \mathcal{H}_{13}^{\epsilon,\mu} - \mathcal{I}_{23}^{\mu'} \mathcal{H}_{23}^{\epsilon,\mu} - \mathcal{I}_{34}^{\mu} \mathcal{H}_{34}^{\epsilon,\mu'} - \mathcal{I}_{35}^{\mu} \mathcal{H}_{35}^{\epsilon,\mu'}]$$

$$\mathcal{H}_{55}^{\epsilon,\mu} = (\mathcal{I}_{55}^{\mu})^{-1} [I_N - \mathcal{I}_{25}^{\mu'} \mathcal{H}_{25}^{\epsilon,\mu} - \mathcal{I}_{35}^{\mu'} \mathcal{H}_{35}^{\epsilon,\mu} - \mathcal{I}_{45}^{\mu'} \mathcal{H}_{45}^{\epsilon,\mu}]$$

## A.5 Affine Model/Exponentially Affine Model: Inference on $B' \Lambda$

Suppose that we have two generic estimators of  $B$  and  $\Lambda$ , 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

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

where

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

When prices of risk are constant,

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

where

$$\begin{aligned} \bar{\mathcal{V}}_{b\lambda}^{\mu} &= (\lambda'_0 \otimes I_N) \kappa_{KN} \bar{\mathcal{V}}_b^{\mu} \kappa'_{KN} (\lambda_0 \otimes I_N) + B' \bar{\mathcal{V}}_\lambda^{\mu} B \\ &\quad + (\lambda'_0 \otimes I_N) \kappa_{KN} \bar{\mathcal{C}}_{b,\lambda}^{\mu} B + B' \bar{\mathcal{C}}_{b,\lambda}^{\mu'} \kappa'_{KN} (\lambda_0 \otimes I_N). \end{aligned}$$



## 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  $\Omega_{ez}$  and  $\Omega_{ev}$ ; define  $\Omega_{vi} = V\iota_T/\sqrt{T}$  and similarly for  $\Omega_{ei}$ ; define  $\Omega_{vv} = VV'/T$  and similarly for  $\Omega_{ee}$ . Recall that  $\hat{\Upsilon} = Z_-Z'_-/T$ . Throughout the Appendix let  $M_i = I_T - T^{-1} \cdot \iota_T \iota'_T$ ,  $M_Z = I_T = Z'_-(Z_-Z'_-)^{-1}Z_-$  and  $\hat{V}_{ols} = XM_Z$ .

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

- (i)  $\Omega_{vv}, \Omega_{ee}, \Omega_{vz}, \Omega_{ez}, \Omega_{ev}$  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,  $T^{-1} \cdot \hat{E}M_i\hat{E}' = \Omega_{ee} + O_p(T^{-1})$ , where  $M_i = I_T - T^{-1} \cdot \iota_T \iota'_T$ ;
- (v)  $\sqrt{T}(\hat{\Psi}_{ols} - \Psi) = \Delta_{\Psi,ols} + o_p(1)$  where  $\Delta_{\Psi,ols} = \Omega_{vz}\Upsilon^{-1}$ ;
- (vi)  $\sqrt{T}(\hat{B}_{ols} - B) = \Delta_{B,ols} + o_p(1)$  where  $\Delta_{B,ols} = \Sigma_v^{-1}\Omega'_{ev}$ .

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

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

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

$$T^{-1} \cdot \hat{E}M_i\hat{E}' = T^{-1} \cdot EM_iE' + T^{-1} \cdot \left[ (\hat{E} - E) M_iE' + EM_i (\hat{E} - E)' + (\hat{E} - E) M_i (\hat{E} - E)' \right].$$

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

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

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

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

Thus,

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

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

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

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

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

**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

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

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

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

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

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

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

$$\begin{aligned} \mathcal{T}_{\Lambda,\Gamma,1} &= \Lambda, \\ \mathcal{T}_{\Lambda,\Gamma,2} &= V Z'_- (Z_- Z'_-)^{-1}, \\ \mathcal{T}_{\Lambda,\Gamma,3} &= - \left( \hat{B}_{\text{ols}} \hat{\Gamma} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} \hat{\Gamma} (\hat{B}_{\text{ols}} - B)' \Lambda, \\ \mathcal{T}_{\Lambda,\Gamma,4} &= - \left( \hat{B}_{\text{ols}} \hat{\Gamma} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} \hat{\Gamma} (\hat{B}_{\text{ols}} - B)' V Z'_- (Z_- Z'_-)^{-1}, \\ \mathcal{T}_{\Lambda,\Gamma,5} &= \left( \hat{B}_{\text{ols}} \hat{\Gamma} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} \hat{\Gamma} E Z'_- (Z_- Z'_-)^{-1}. \end{aligned}$$

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

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

and

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

Using Lemma 1 again we have,

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

where

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

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

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

where

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

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

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

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

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

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

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

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

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

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

$$\mathcal{T}_{B, \Gamma, 1, 1} = -[\Lambda \Upsilon \Lambda' + \Sigma_v]^{-1} \Lambda \Upsilon \Delta'_{\Lambda, \Gamma} B + o_p(1), \quad \mathcal{T}_{B, \Gamma, 1, 2} = [\Lambda \Upsilon \Lambda' + \Sigma_v]^{-1} \Lambda \Omega'_{vz} B + o_p(1),$$

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

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

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

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

$$1. \sqrt{T}(\hat{\lambda}_0^\mu(\hat{\Gamma}) - \lambda_0) = \Delta_{\lambda, \Gamma}^\mu + o_p(1) \text{ where } \Delta_{\lambda, \Gamma}^\mu = \Omega_{vl} + (B\Gamma B')^{-1} B\Gamma \Omega_{el} - (B\Gamma B')^{-1} B\Gamma \Omega_{ev} \Sigma_v^{-1} \lambda_0,$$

$$2. \sqrt{T}(\hat{B}^\mu(\hat{\Gamma}) - B) = \Delta_{B, \Gamma}^\mu + o_p(1) \text{ where } \Delta_{B, \Gamma}^\mu = [\Sigma_v + \lambda_0 \lambda_0']^{-1} (\lambda_0 \Omega'_{vl} B + \lambda_0 \Omega'_{el} + \Omega'_{ev} - \lambda_0 \Delta'_{\lambda, \Gamma} B).$$

**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  $\delta_{\Psi, \text{ols}} = (\Upsilon^{-1} \otimes I_K) \omega_{vz}$  and so by Lemma 2 and Slutsky's Lemma  $\delta_{\Psi, \text{ols}} \rightarrow_d \mathcal{N}(0, (\Upsilon^{-1} \otimes \Sigma_v))$ . Similarly,  $\delta_{B, \text{ols}} = (I_N \otimes \Sigma_v^{-1}) \kappa_{NK} \omega_{ev}$  and so by Lemma 2, Slutsky's Lemma and standard properties of the commutation matrix  $\delta_{B, \text{ols}} \rightarrow_d \mathcal{N}(0, \Sigma_e \otimes \Sigma_v^{-1})$ . Next, by Lemma 3 we have,

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

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

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

and so the result follows by setting  $\Gamma$  to  $I_N$  or  $\Sigma_e^{-1}$ .

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

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

or equivalently,

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

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

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

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

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

where

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

By similar calculations as above we have that

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

and

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

Cancelling the common term, simplifying the resulting expression, and setting  $\Gamma$  to  $I_N$  or  $\Sigma_e^{-1}$  yields the result.

**Remark 4** *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*

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

*We still have that  $\mathbb{E}[\omega_{zv}\omega'_{ez}] = 0$  but  $\mathbb{E}[\omega_{zv}\omega'_{ev}]$  and  $\mathbb{E}[\omega_{ez}\omega'_{ev}]$  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,

$$\begin{aligned}\frac{\partial \ell(\theta; \sigma_e, \sigma_v)}{\partial \psi'} &= - \text{vec}(E)' (I_T \otimes \Sigma_e^{-1}) \frac{\partial \text{vec}(E)}{\partial \psi'} - \text{vec}(V)' (I_T \otimes \Sigma_v^{-1}) \frac{\partial \text{vec}(V)}{\partial \psi'} \\ &= - \text{vec}(B\Sigma_e^{-1}EZ'_-)' + \text{vec}(\Sigma_v^{-1}VZ'_-)' .\end{aligned}$$

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,

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

and so by Lemma 1,

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

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.

**Proof of Theorem 6.** For simplicity define,

$$\hat{\Lambda}^\epsilon(\Gamma) = \left( \hat{B}_{\text{ols}} \Gamma \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} \Gamma \left[ r_e^\epsilon + \frac{1}{2} \hat{J}_{B, \text{ols}} \text{vec}(\hat{\Sigma}_{v, \text{ols}}) \iota'_T + \frac{1}{2} \hat{d}_{e, \text{sur}} \iota'_T \right] Z'_- (Z_- Z'_-)^{-1}.$$

Suppose that  $\hat{\Gamma} - \Gamma = o_p(1)$  and note that by similar steps as in the proof of Lemma 1 (iv) we can show that  $\hat{\Sigma}_{e, \text{ols}}^{-1} - \Sigma_e^{-1} = o_p(1)$ . Then, similar to the proof of Lemma 3,

$$\hat{\Lambda}^\epsilon(\hat{\Gamma}) = \mathcal{T}_{\Lambda, \Gamma, 1}^\epsilon + \mathcal{T}_{\Lambda, \Gamma, 2}^\epsilon + \mathcal{T}_{\Lambda, \Gamma, 3}^\epsilon + o_p(T^{-1/2}),$$

where

$$\begin{aligned} \mathcal{T}_{\Lambda, \Gamma, 1}^\epsilon &= T^{-1/2} \cdot \Delta_{\Lambda, \Gamma}, \\ \mathcal{T}_{\Lambda, \Gamma, 2}^\epsilon &= \frac{1}{2} \left( \hat{B}_{\text{ols}} \hat{\Gamma} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} \hat{\Gamma} \left[ \hat{J}_{B, \text{ols}} \text{vec}(\hat{\Sigma}_{v, \text{ols}}) - J_B \text{vec}(\Sigma_v) \right] \hat{\Upsilon}_1 \hat{\Upsilon}_1^{-1}, \\ \mathcal{T}_{\Lambda, \Gamma, 3}^\epsilon &= \frac{1}{2} \left( \hat{B}_{\text{ols}} \hat{\Gamma} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} \hat{\Gamma} \left[ \hat{d}_{e, \text{sur}} - d_e \right] \hat{\Upsilon}_1 \hat{\Upsilon}_1^{-1}. \end{aligned}$$

Let us deal with  $\mathcal{T}_{\Lambda, \Gamma, 2}^\epsilon$  first. The  $i$ th row of  $\hat{J}_{B, \text{ols}} \text{vec}(\hat{\Sigma}_{v, \text{ols}}) - J_B \text{vec}(\Sigma_v)$  is

$$\hat{\beta}'_{i, \text{ols}} \hat{\Sigma}_v \hat{\beta}_{i, \text{ols}} = \beta'_i \Sigma_v \beta_i + \left( \hat{\beta}_{i, \text{ols}} - \beta_i \right)' \Sigma_v \beta_i + \beta'_i \Sigma_v \left( \hat{\beta}_{i, \text{ols}} - \beta_i \right) + \beta'_i \left( \hat{\Sigma}_v - \Sigma_v \right) \beta_i + O_p(T^{-1}).$$

Thus,  $\mathcal{T}_{\Lambda, \Gamma, 2}^\epsilon = \mathcal{T}_{\Lambda, \Gamma, 2, 1}^\epsilon + \mathcal{T}_{\Lambda, \Gamma, 2, 2}^\epsilon + o_p(T^{-1/2})$ , where

$$\mathcal{T}_{\Lambda, \Gamma, 2, 1}^\epsilon = (B \Gamma B')^{-1} B \Gamma \begin{bmatrix} \beta'_1 \Sigma_v \left( \hat{\beta}_{1, \text{ols}} - \beta_1 \right) \\ \vdots \\ \beta'_N \Sigma_v \left( \hat{\beta}_{N, \text{ols}} - \beta_N \right) \end{bmatrix} \Upsilon'_1 \Upsilon_1^{-1},$$

and

$$\mathcal{T}_{\Lambda, \Gamma, 2, 2}^\epsilon = \frac{1}{2} (B \Gamma B')^{-1} B \Gamma J_B \text{vec} \left( \hat{\Sigma}_v - \Sigma_v \right) \Upsilon'_1 \Upsilon_1^{-1}.$$

We will deal with  $\mathcal{T}_{\Lambda, \Gamma, 2, 2}^\epsilon$  first. By Lemma 1

$$\mathcal{T}_{\Lambda, \Gamma, 2, 2}^\epsilon = \frac{1}{2} (B \Gamma B')^{-1} B \Gamma J_B \text{vec} \left( \hat{\Sigma}_v - \Sigma_v \right) \Upsilon'_1 \Upsilon_1^{-1} = \frac{1}{2} (B \Gamma B')^{-1} B \Gamma J_B \text{vec} \left( \Omega_{vv} - \Sigma_v \right) \Upsilon'_1 \Upsilon_1^{-1} + o_p(T^{-1/2}).$$

Next let us deal with  $\mathcal{T}_{\Lambda, \Gamma, 2, 1}^\epsilon$ . Note that by Lemma 1

$$\sqrt{T} \beta'_i \Sigma_v \left( \hat{\beta}_{i, \text{ols}} - \beta_i \right) = \beta'_i \frac{VE^i}{\sqrt{T}} + o_p(1),$$

where  $(E^i)'$  is the  $i$ th row of the matrix  $E$ . Thus,

$$\sqrt{T} \mathcal{T}_{\Lambda, \Gamma, 2, 1}^\epsilon = (B\Gamma B')^{-1} B\Gamma \xi_{ve} \Upsilon'_1 \Upsilon^{-1} + o_p(1),$$

where

$$\xi_{ve} = \begin{bmatrix} \beta'_1 \frac{VE^1}{\sqrt{T}} \\ \vdots \\ \beta'_N \frac{VE^N}{\sqrt{T}} \end{bmatrix}.$$

Next let us consider  $\mathcal{T}_{\Lambda, \Gamma, 3}^\epsilon$ . By similar steps as in the proof of Lemma 1 (*iv*), we can show that  $\hat{d}_{e, \text{sur}} - d_e = \Delta_d + o_p(T^{-1/2})$  where  $\Delta_d$  is comprised of the diagonal elements of  $\Omega_{ee} - \Sigma_e$ . Thus,

$$\mathcal{T}_{\Lambda, \Gamma, 3}^\epsilon = \frac{1}{2} \left( \hat{B}_{\text{ols}} \hat{\Gamma} \hat{B}'_{\text{ols}} \right)^{-1} \hat{B}_{\text{ols}} \hat{\Gamma} \left[ \hat{d}_{e, \text{sur}} - d_e \right] \hat{\Upsilon}_1 \hat{\Upsilon}^{-1} = \frac{1}{2} (B\Gamma B')^{-1} B\Gamma \Delta_d \Upsilon_1 \Upsilon^{-1} + o_p(T^{-1/2}).$$

Combining these results yields

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

where

$$\begin{aligned} \Delta_{\Lambda, \Gamma}^\epsilon &= \Delta_{\Lambda, \Gamma} + (B\Gamma B')^{-1} B\Gamma \xi_{ve} \Upsilon'_1 \Upsilon^{-1} + \frac{1}{2} (B\Gamma B')^{-1} B\Gamma J_B \text{vec}(\Omega_{vv} - \Sigma_v) \Upsilon'_1 \Upsilon^{-1} \\ &\quad + \frac{1}{2} (B\Gamma B')^{-1} B\Gamma \Delta_d \Upsilon_1 \Upsilon^{-1}. \end{aligned}$$

By Lemma 3,

$$\Delta_{\Lambda, \Gamma} = \Delta_{\Psi, \text{ols}} + (B\Gamma B')^{-1} B\Gamma \Omega_{ez} \Upsilon^{-1} - (B\Gamma B')^{-1} B\Gamma \Omega_{ev} \Sigma_v^{-1} \Lambda.$$

To deal with the variance and sole asymptotic covariance term note that

$$\lim_{T \rightarrow \infty} \mathbb{E} [\xi_{ve} \xi'_{ve}] = \Xi_1, \tag{23}$$

and

$$\lim_{T \rightarrow \infty} \mathbb{E} [\omega_{ev} \xi'_{ve}] = \kappa_{KN} \Xi_2. \tag{24}$$

From the proof of Theorem 1,

$$\lim_{T \rightarrow \infty} \mathbb{E} [\delta_{\Lambda, \Gamma} \delta'_{\Lambda, \Gamma}] = \mathcal{V}_{\hat{\Lambda}, \Gamma}.$$

The next variance term to consider

$$\text{vec} \left( (B\Gamma B')^{-1} B\Gamma \xi_{ve} \Upsilon'_1 \Upsilon^{-1} \right) \xrightarrow{d} \mathcal{N} \left( 0, \Upsilon^{-1} \Upsilon_1 \Upsilon'_1 \Upsilon^{-1} \otimes (B\Gamma B')^{-1} B\Gamma \Xi_1 \Gamma B' (B\Gamma B')^{-1} \right),$$

by equation (23). Next since  $\sqrt{T} \text{vec}(\Omega_{vv} - \Sigma_v) \rightarrow_d \mathcal{N}(0, (I_{K^2} + \kappa_{K, K})(\Sigma_v \otimes \Sigma_v))$  then,

$$\text{vec} \left( \frac{1}{2} (B\Gamma B')^{-1} B\Gamma J_B \text{vec}(\Omega_{vv} - \Sigma_v) \Upsilon'_1 \Upsilon^{-1} \right) \xrightarrow{d} \mathcal{N} \left( 0, \mathcal{V}_{T, \Gamma, 2, 2}^\epsilon \right),$$

where

$$\mathcal{V}_{T,\Gamma,2,2}^\epsilon = \frac{1}{4} \left( \Upsilon^{-1} \Upsilon_1 \Upsilon_1' \Upsilon^{-1} \otimes (B\Gamma B')^{-1} B\Gamma J_B (I_{K^2} + \kappa_{KK}) (\Sigma_v \otimes \Sigma_v) J_B' \Gamma B' (B\Gamma B')^{-1} \right).$$

Finally, since  $\sqrt{T} \text{vec}(\Omega_{ee} - \Sigma_e) \rightarrow_d \mathcal{N}(0, (I_{N^2} + \kappa_{N,N}) (\Sigma_e \otimes \Sigma_e))$ , then

$$\text{vec} \left( \frac{1}{2} (B\Gamma B')^{-1} B\Gamma \Delta_d \Upsilon_1 \Upsilon^{-1} \right) \xrightarrow{d} \mathcal{N} \left( \frac{1}{2} \Upsilon^{-1} \Upsilon_1 \Upsilon_1' \Upsilon^{-1} \otimes (B\Gamma B')^{-1} B\Gamma \Sigma_e^2 \Gamma B' (B\Gamma B')^{-1} \right).$$

The sole asymptotic covariance term is,

$$\begin{aligned} & \lim_{T \rightarrow \infty} \mathbb{E} \left[ - \left( \Lambda \Sigma_v^{-1} \otimes (B\Gamma B')^{-1} B\Gamma \right) \Omega_{ev} \xi'_{ve} \left( \Upsilon^{-1} \Upsilon_1 \otimes (B\Gamma B')^{-1} B\Gamma \right)' \right] \\ &= - \left( \Lambda \Sigma_v^{-1} \otimes (B\Gamma B')^{-1} B\Gamma \right) \kappa_{KN} \Xi_2 \left( \Upsilon_1' \Upsilon^{-1} \otimes \Gamma B' (B\Gamma B')^{-1} \right), \end{aligned}$$

by equation (24).

**Proof of Theorem 7.** Similar to the proof of Theorem 4, we may rely on standard matrix calculus results. Note that the first and third elements of the score are unchanged from the affine model. For the second element of the score,

$$\frac{\partial \ell^\epsilon(\theta^\epsilon)}{\partial \beta_i} = -\frac{1}{\sigma_i^2} \sum_{t=1}^T e_{i,t} (\Sigma_v \beta_i - (\Lambda Z_{t-1} + v_t)),$$

which may be stacked to form

$$\frac{\partial \ell^\epsilon(\theta^\epsilon)}{\partial b} = \text{vec}((\Lambda Z_- + V) E' \Sigma_e^{-1}) - [\text{vec}(\Sigma_v B \Sigma_e^{-1}) \odot (E^\epsilon \iota_T \otimes \iota_K)].$$

By similar steps we may obtain the fourth and fifth elements of the score vector. Similarly, The (1,1), (1,3) and (3,3) elements of the information matrix are unchanged from the affine model. The (2,2) element is,

$$\frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \beta_i \partial \beta_i'} = - \sum_{t=1}^T \frac{1}{\sigma_i^2} (\Sigma_v \beta_i - (\Lambda Z_{t-1} + v_t)) (\Sigma_v \beta_i - (\Lambda Z_{t-1} + v_t))' - \sum_{t=1}^T \frac{1}{\sigma_i^2} e_{i,t}^2 \frac{\partial (\Sigma_v \beta_i)'}{\partial \beta_i},$$

and

$$\frac{\partial^2 \ell^\epsilon(\theta^\epsilon)}{\partial \beta_i \partial \beta_j'} = 0, \quad i \neq j.$$

Stacking these results and utilizing Lemma 1 yields,

$$\begin{aligned} & \lim_{T \rightarrow \infty} \left( -\frac{1}{T} \right) \mathbb{E} \left[ \frac{\partial^2 \ell^\epsilon(B, \Lambda, \Psi, \Sigma_v; \Sigma_e)}{\partial b \partial b'} \right] \\ &= \text{bdiag}_{[NK, NK], [K, K]} (\sigma_i^{-2} [\Lambda \Upsilon \Lambda' + \Sigma_v + \Sigma_v \beta_i \beta_i' \Sigma_v - \Sigma_v \beta_i \Upsilon_1' \Lambda' - \Lambda \Upsilon_1 \beta_i' \Sigma_v]). \end{aligned}$$

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 2.** This follows by Theorem 7 and standard properties of one step estimators. See, for example, van der Vaart (1998).



## References

- 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.
- BURNSIDE, A. C. (2009): “The Cross-section of Foreign Currency Risk Premia and Consumption Growth Risk: A Comment,” Working paper, Duke University, Department of Economics.
- BURNSIDE, C. (2010): “Identification and Inference in Linear Stochastic Discount Factor Models,” NBER Working Paper 16634, National Bureau of Economic Research, Inc.
- CAMPBELL, J. Y. (1987): “Stock Returns and the Term Structure,” *Journal of Financial Economics*, 18(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 (1985a): “An Intertemporal General Equilibrium Model of Asset Prices,” *Econometrica*, 53(2), 363–384.
- (1985b): “A Theory of the Term Structure of Interest Rates,” *Econometrica*, 53(2), 385–407.
- DAI, Q., AND K. SINGLETON (2000): “Specification Analysis of Affine Term Structure Models,” *Journal of Finance*, 55(5), 1943–1978.
- (2002): “Expectations 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.
- (2010): "Reality Checks for and of Factor Pricing," Discussion paper, Brown University.
- 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.
- PIAZZESI, M. (2003): "Affine Term Structure Models," *Handbook of Financial Econometrics*.
- 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.
- VASICEK, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5(2), 177–188.
- WHITE, H. (2001): *Asymptotic Theory for Econometricians*. Academic Press, San Diego, Revised edition.