

Discussion Paper: 2005/09

# Tests of risk premia in linear factor models

Frank Kleibergen

(revised December 2006)

[www.fee.uva.nl/ke/UvA-Econometrics](http://www.fee.uva.nl/ke/UvA-Econometrics)

**Amsterdam School of Economics**

Department of Quantitative Economics  
Roetersstraat 11  
1018 WB AMSTERDAM  
The Netherlands

UvA  UNIVERSITEIT VAN AMSTERDAM



# Tests of Risk Premia in Linear Factor Models

Frank Kleibergen\*

December 21, 2006

## Abstract

We show that inference on risk premia in linear factor models that is based on the Fama-MacBeth (FM) and generalized least squares (GLS) risk premia estimators is misleading when the  $\beta$ 's are small and/or the number of assets is large. We propose novel statistics, that are based on the maximum likelihood estimator of Gibbons (1982), which remain trustworthy in these cases. We show that the maximum likelihood estimator is straightforward to compute using the eigenvectors of a characteristic polynomial. The inadequacy of the FM and GLS  $t$ /Wald statistics is highlighted in a power and size comparison and using daily portfolio returns from Jagannathan and Wang (1996). The power and size comparison shows that the FM and GLS  $t$ /Wald statistics can be severely size distorted. The monthly portfolio returns from Jagannathan and Wang (1996) reveal a large discrepancy between the 95% confidence sets for the risk premia that result from the different statistics. The FM and GLS  $t$ /Wald statistics imply tight 95% confidence sets for the risk premia on the yield premium and labor income growth that only contain small values while the 95% confidence sets that result from statistics that remain trustworthy in case of small  $\beta$ 's imply much larger 95% confidence sets at considerably higher and therefore more relevant values of the risk premia.

## 1 Introduction

Linear factor models are amongst the most commonly used statistical models in finance, see *e.g.* Lintner (1965), Fama and MacBeth (1973), Gibbons (1982), Shanken (1992), Fama and French (1992,1993,1996) and Jagannathan and Wang (1996,1998). This results as many well-known financial models, like, for example, the capital asset pricing model, imply a linear factor structure for the asset returns. Linear factor models imply risk premia for the different factors that result from a cross-sectional regression of the asset returns on the factor  $\beta$ 's where the  $\beta$ 's are obtained from a time-series regression of the asset returns on the factors. This two step procedure is typically referred to as the Fama-MacBeth (FM) regression, see Fama and MacBeth (1973). Linear regression estimates are sensitive to collinearity of the explanatory variables so the estimates of the risk premia that result from a FM regression are sensitive to collinearity of the  $\beta$ 's. Collinearity of the  $\beta$ 's occurs when they are close or equal to zero and/or when the  $\beta$ -matrix is almost of reduced rank. Kan and Zhang (1999), for example, show that the estimates of the risk premia that result from a FM regression are erroneous when the  $\beta$ 's are zero and the expected asset returns are non-zero.

---

\*Department of Economics, Brown University, Box B, Providence, RI 02912, United States and Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, Email: Frank\_Kleibergen@brown.edu. Homepage: [http://www.econ.brown.edu/fac/Frank\\_Kleibergen](http://www.econ.brown.edu/fac/Frank_Kleibergen). I thank Sung Jae Jun for research assistance and Craig MacKinlay and other participants of the SBFSIF 2 meeting in Quebec City for helpful comments and discussion.

FM regressions provide misleading risk premia estimates in more cases than just the one indicated by Kan and Zhang (1999). We show that FM risk premia estimates are spurious whenever the  $\beta$ 's are relatively small and their spuriousness is further aggravated when the number of assets is large. Hence, even for size-able values of the  $\beta$ 's, the risk premia estimates can be misleading when the number of assets is large. The FM risk premia estimates that are obtained in such settings are typically too small and their 95% confidence sets indicate an erroneous high level of precision. The erroneous level of precision of the confidence sets shows that statistical inference based on Wald  $t$ -statistics is misguided and that the large sample distribution of the  $t$ -statistic differs from the normal one. We therefore propose a number of alternative statistics whose large sample distributions remain unaltered when the  $\beta$ 's are close to zero. Thus these statistics can be used in a trustworthy manner.

The paper is organized as follows. In the second section, we discuss the FM and Generalized Least Squares (GLS) risk premia estimators and construct their large sample distributions when the  $\beta$ 's are small. These large sample distributions are non-standard and imply that Wald statistics, like, for example, the  $t$ -statistic, are misleading when the  $\beta$ 's are small and/or when the number of assets is large. The third section proposes a number of novel statistics whose large sample distributions remain unaltered when the  $\beta$ 's are small and/or the number of assets is large. These statistics are based on the Generalized Method of Moment (GMM) and instrumental variable statistics that are proposed in Kleibergen (2002,2005a) and Moreira (2003). They lead to inference that is centered around the maximum likelihood estimator (MLE) of Gibbons (1982) and we show that the MLE results in a straightforward manner from the eigenvectors of a characteristic polynomial thus avoiding numerical optimization. The fourth section conducts a size and power comparison. It illustrates the size distortions of the Wald  $t$ -statistics. The fifth section generalizes the test statistics to allow for tests on some of the risk premia. The sixth section uses data from Jagannathan and Wang (1996) to estimate a linear factor model for monthly portfolio returns. It shows that, because of the large number of portfolios and the relatively small  $\beta$ 's of the yield premium and labor income growth that, the FM and GLS risk premia estimators for the risk premia on the yield premium and labor income growth are downward biased and lead to small 95% confidence sets at rather low values. The 95% confidence sets of the statistics that remain trustworthy in case of many portfolios and small  $\beta$ 's indicate much larger 95% confidence sets at considerably higher values of the risk premia. This shows the importance of usage of these statistics for inference on the risk premia.

We use the following notation throughout the paper:  $E(a)$  is the mean of a random variable  $a$ ,  $\text{var}(a)$  is the variance and  $\text{cov}(a, b)$  is the covariance between  $a$  and  $b$ ;  $\text{vec}(A)$  stands for the column vectorization of the  $T \times n$  dimensional matrix  $A$ ,  $\text{vec}(A) = (a'_1 \dots a'_n)'$  when  $A = (a_1 \dots a_n)$ ,  $\text{vecinv}_n((a'_1 \dots a'_n)') = A$ ,  $P_A = A(A'A)^{-1}A'$  and  $M_A = I_T - P_A$  for a full rank matrix  $A$  and the  $T \times T$  identity matrix  $I_T$ ,  $\iota_n$  is a  $n \times 1$  vector of ones, " $\xrightarrow{p}$ " indicates convergence in probability, " $\xrightarrow{d}$ " indicates convergence in distribution and " $\stackrel{a}{=}$ " implies equality in large samples.

## 2 Risk Premia estimators and small $\beta$ 's

We analyze the FM estimator in a GMM setting, see *e.g.* Hansen (1982). We therefore specify the moments of the factor related (excess) assets returns<sup>1</sup> as, see *e.g.* Cochrane (2001),

$$\begin{aligned} E(R_t) &= \iota_n \lambda_1 + \beta \lambda_F \\ \text{cov}(R_t, F_t) &= \beta \text{var}(F_t) \\ E(F_t) &= \mu_F, \end{aligned} \tag{1}$$

---

<sup>1</sup>All asset returns are in deviation from the riskless return.

with  $R_t$  : the  $n \times 1$  vector of (excess) asset returns,  $F_t$  : the  $k \times 1$  vector of factors,  $\lambda_1$  : the zero- $\beta$  return,  $\lambda_F$  : the  $k \times 1$  vector of factor risk premia,  $\beta$  : the  $n \times k$  matrix of asset  $\beta$ 's,  $\mu_F$  : the  $k \times 1$  vector of factor means and  $t$  : the time index of the observation,  $t = 1, \dots, T$ . The moment conditions (1) result from a linear factor model for the asset returns and factors,

$$\begin{aligned} R_t &= \iota_n \lambda_1 + \beta(F_t - \mu_F + \lambda_F) + u_t \\ F_t &= \mu_F + v_t \end{aligned} \quad (2)$$

with  $u_t$  and  $v_t$ ,  $t = 1, \dots, T$  :  $n \times 1$  and  $k \times 1$  vectors of disturbances. These vectors of disturbances are identically and independently distributed over time with mean zero and  $n \times n$  and  $k \times k$  dimensional covariance matrices  $\Lambda$  (for  $u_t$ ) and  $V_{FF}$  (for  $v_t$ ). Because  $\beta = \text{cov}(R_t, F_t) \text{var}(F_t)^{-1}$ , the covariance between the disturbances  $u_t$  and  $v_t$  is equal to  $\beta V_{FF}$ . The linear factor model (2) can as well be specified as

$$\left. \begin{aligned} R_t &= \iota_n \lambda_1 + \beta(F_t - \bar{F} + \bar{F} - \mu_F + \lambda_F) + u_t \\ F_t &= \mu_F + v_t \end{aligned} \right\} \Leftrightarrow \left\{ \begin{aligned} R_t &= \iota_n \lambda_1 + \beta(\bar{F}_t + \lambda_F) + \varepsilon_t \\ F_t &= \mu_F + v_t \end{aligned} \right. \quad (3)$$

with  $\bar{F}_t = F_t - \bar{F}$ ,  $\bar{F} = \frac{1}{T} \sum_{t=1}^T F_t$ ,  $\varepsilon_t = u_t + \beta \bar{v}$  and  $\bar{v} = \frac{1}{T} \sum_{t=1}^T v_t$ . Because of the specification of  $\beta$ , weighted averages of  $\varepsilon_t$  and  $v_t$  are uncorrelated in large samples. We therefore use the latter specification in (3) to make an assumption about the behavior of the disturbances of the linear factor model (2) in large samples.

**Assumption 1.** *When the number of time series observations  $T$  becomes large,*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{pmatrix} \begin{pmatrix} 1 \\ F_t \end{pmatrix} \otimes (R_t - \iota_n \lambda_1 - \beta(\bar{F}_t + \lambda_F)) \\ F_t - \mu_f \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \varphi_R \\ \varphi_\beta \\ \varphi_F \end{pmatrix}, \quad (4)$$

with  $\varphi_R : n \times 1$ ,  $\varphi_\beta : nk \times 1$ ,  $\varphi_F : k \times 1$ ,  $(\varphi'_R \vdots \varphi'_\beta \vdots \varphi'_F)' \sim N(0, V)$  and

$$V = \begin{pmatrix} (Q \otimes \Omega) & 0 \\ 0 & V_{FF} \end{pmatrix}, \quad (5)$$

with  $Q = \begin{pmatrix} Q_{11} \vdots Q_{1F} \\ Q_{F1} \vdots Q_{FF} \end{pmatrix} = \begin{pmatrix} 1 \vdots \mu'_F \\ \mu_F \vdots V_{FF} + \mu_F \mu'_F \end{pmatrix} : (k+1) \times (k+1)$ ,  $\Omega : n \times n$ ,  $V_{FF} : k \times k$ ,  $V_{FF} = \text{var}(F_t)$  and  $\Omega = \Lambda - \beta V_{FF} \beta'$ .

Assumption 1 is a central limit theorem for the disturbances of the linear factor model (3) interacted with a constant and the factors. It holds under rather mild conditions, like, for example, Assumptions 1 and 2 from Shanken (1992). Because of the independence of the disturbances over time and their finite variance, Assumption 1 holds under the linear factor model (2) with a constant covariance matrix as well. Appendix A shows the implications of Assumption 1 for a central limit theorem for the original specification of the linear factor model (2).

For time series observations of  $R_t$  and  $F_t$ ,  $t = 1, \dots, T$ , the FM two-pass methodology estimates the risk premia using a least squares time-series regression estimate of  $\beta$ ,  $\hat{\beta} = \sum_{t=1}^T \bar{R}_t \bar{F}_t' \left[ \sum_{j=1}^T \bar{F}_j \bar{F}_j' \right]^{-1}$ , with  $\bar{R}_t = R_t - \bar{R}$ ,  $\bar{R} = \frac{1}{T} \sum_{t=1}^T R_t$ . Given  $\hat{\beta}$ , estimators for  $\lambda_1$  and  $\lambda_k$  are obtained by regressing the average value of  $R_t$ ,  $\bar{R}$ , on  $(\iota_n \vdots \hat{\beta})$ ,

$$\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} = \left[ (\iota_n \vdots \hat{\beta})' (\iota_n \vdots \hat{\beta}) \right]^{-1} (\iota_n \vdots \hat{\beta})' \bar{R}. \quad (6)$$

**Lemma 1.** *When Assumption 1 holds, the average returns,  $\bar{R}$ , the least squares estimator of the assets  $\beta$ 's,  $\hat{\beta}$ , and the average factors,  $\bar{F}$ , have a joint large sample distribution that reads*

$$\sqrt{T} \begin{pmatrix} \bar{R} - \iota_n \lambda_1 - \beta \lambda_F \\ \text{vec}(\hat{\beta} - \beta) \\ \bar{F} - \mu_f \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_R \\ \psi_\beta \\ \psi_F \end{pmatrix}, \quad (7)$$

where  $\psi_R$ ,  $\psi_\beta$  and  $\psi_F$  are  $n \times 1$ ,  $nk \times 1$  and  $k \times 1$  independent normal distributed random vectors with mean zero and covariance matrices,  $\Omega$ ,  $V_{FF}^{-1} \otimes \Omega$  and  $V_{FF}$ .

**Proof.** see Appendix B. ■

Lemma 1 corresponds with Lemma 1 of Shanken (1992). We use Lemma 1 to derive the properties of the FM risk premia estimator.

When the  $\beta$ 's have size-able values, the FM risk premia estimator has a normal limiting distribution so Wald  $t$ -tests perform adequately. Kan and Zhang (1999) show that the large sample distribution of the FM risk premia estimator is notably different when the  $\beta$ 's are equal to zero and  $E(\bar{R})$  equals some  $n \times 1$  vector of constants  $c$ . The behavior of the FM risk premia estimator differs in other cases as well. It results from the dependence of the FM risk premia estimator on the  $\beta$ 's. This dependence is similar to the dependence of the two stage least squares estimator on the quality of the instruments in econometrics. There is an extensive literature in econometrics that discusses the behavior of the two stage least squares estimator under limit sequences where the quality of the instruments depends on the sample size, see *e.g.* Phillips (1989), Bekker (1994) and Staiger and Stock (1997). Theorem 1 states the behavior of the FM risk premia estimator in three similar cases for the  $\beta$ 's. The first case corresponds with Kan and Zhang (1999) and has a zero value of the  $\beta$ -vector. The second case has relatively small but non-zero elements of the  $\beta$ -vector. We therefore specify the  $\beta$ 's such that they decrease with the sample size which is analogous to the weak instruments of Staiger and Stock (1999) so we refer to this case as weak  $\beta$ 's. The third case has all  $\beta$ 's equal to zero except for a fixed fraction of the assets.

**Theorem 1.** *When Assumptions 1 and 2 hold, the behavior of the FM risk premia estimator (6) in large samples is characterized by:*

1. *When  $\beta = 0$  and*

(a)  $E(\bar{R}) = \iota_n \lambda_1$  :

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \lambda_1) \\ \hat{\lambda}_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ [\Psi'_\beta M_{\iota_n} \Psi_\beta]^{-1} \Psi'_\beta M_{\iota_n} \end{pmatrix} \psi_R, \quad (8)$$

*and the order of finite moments of the limit behavior of  $\hat{\lambda}_F$  is equal to  $n - k$  while the mean of the limit behavior of  $\hat{\lambda}_F$  equals zero when  $n > k$ .*

(b)  $E(\bar{R}) = c$  :

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \frac{1}{a} b' c) \\ \hat{\lambda}_F - \sqrt{T} [\Psi'_\beta M_{\iota_n} \Psi_\beta]^{-1} \Psi'_\beta M_{\iota_n} c \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ [\Psi'_\beta M_{\iota_n} \Psi_\beta]^{-1} \Psi'_\beta M_{\iota_n} \end{pmatrix} \psi_R \quad (9)$$

where  $\Psi_\beta = \text{vecinv}_k(\psi_\beta)$ ,  $a = \iota'_n M_{\Psi_\beta} \iota_n$ ,  $b = M_{\Psi_\beta} \iota_n$ .

2. *When  $\beta$  has a weak value such that  $\beta = \beta_T = \frac{1}{\sqrt{T}} B$ , with  $B$  a fixed full rank  $n \times k$  matrix, and*

(a)  $E(\bar{R}) = \iota_n \lambda_1 + \beta \lambda_F$  :

$$\begin{aligned} & \begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \lambda_1) \\ \hat{\lambda}_F - \lambda_F \end{pmatrix} \xrightarrow{d} \\ & \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ [(B + \Psi_\beta)' M_{\iota_n} (B + \Psi_\beta)]^{-1} (B + \Psi_\beta)' M_{\iota_n} \end{pmatrix} (\psi_R - \Psi_\beta \lambda_F). \end{aligned} \quad (10)$$

and the order of finite moments of the limit behavior of  $\hat{\lambda}_F$  is equal to  $n - k$ .

(b)  $E(\bar{R}) = c$  :

$$\begin{aligned} & \begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \frac{1}{a} b' c) \\ \hat{\lambda}_F - \sqrt{T} [(B + \Psi_\beta)' M_{\iota_n} (B + \Psi_\beta)]^{-1} (B + \Psi_\beta)' M_{\iota_n} c \end{pmatrix} \\ & \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ [(B + \Psi_\beta)' M_{\iota_n} (B + \Psi_\beta)]^{-1} (B + \Psi_\beta)' M_{\iota_n} \end{pmatrix} \psi_R, \end{aligned} \quad (11)$$

where  $\Psi_\beta = \text{vecinv}_k(\psi_\beta)$ ,  $a = \iota_n' M_{B+\Psi_\beta} \iota_n$ ,  $b = M_{B+\Psi_\beta} \iota_n$ .

3. When  $\beta = \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}$ , with  $\beta_1 : n_1 \times k$ ,  $n_1$  is fixed and  $\frac{n}{T} \rightarrow c$ , the bias of  $\hat{\lambda}_F$  equals

$$(\beta_1' \beta_1 + W)^{-1} W (\mu_F - \lambda_F), \quad (12)$$

with  $W = [\lim_{n \rightarrow \infty} \text{tr}(\frac{1}{n} \Omega)] V_{FF}^{-1}$ .

**Proof.** see Appendix C. ■

Theorem 1 shows that the large sample distribution of the FM risk premia estimator differs substantially from normality in case of zero, weak or many  $\beta$ 's. First, Theorem 1 shows that zero or weak values of  $\beta$  alter the convergence rate of the risk premia estimator. When the moment conditions apply, so  $E(\bar{R}) = \iota_n \lambda_1$  (1a) or  $E(\bar{R}) = \iota_n \lambda_1 + \beta \lambda_F$  (2a), Theorem 1 shows that the risk premia estimator  $\hat{\lambda}_F$  converges to a random variable instead of the convergence at rate  $\sqrt{T}$  to the true value  $\lambda_F$  in case of size-able values of  $\beta$ . When the moment conditions do not apply, so  $E(\bar{R}) = c$  in (1b) and (2b), the risk premia estimator  $\hat{\lambda}_F$  diverges at rate  $\sqrt{T}$  both in case of zero or weak  $\beta$ 's. This extends Kan and Zhang (1999) who show that the risk premia estimator  $\hat{\lambda}_F$  diverges when  $\beta$  equals zero and  $E(\bar{R}) = c$  which corresponds with (1b). Second, Theorem 1 shows that the random variable where the risk premia estimator  $\hat{\lambda}_F$  converges to depends on the number of assets since the number of finite moments of this random variable depends on the number of assets. Hence, when we add assets with zero  $\beta$ 's, the density of the risk premia estimator becomes more concentrated. The dependence on the number of assets is further indicated in the third case which shows that the FM risk premia estimator is biased towards  $\mu_F$  when the number of assets is large and the  $\beta$ 's are only non-zero for a relatively small number of assets. The expression of the bias shows that the FM estimator is more biased towards  $\mu_F$  (the factor means) for smaller values of  $\beta$ .

Although we do not observe the stylized settings from Theorem 1 in practice, Theorem 1 shows that the traditional asymptotic theory, that is used to conduct tests on the risk premia using the FM estimator, becomes unreliable when the  $\beta$ 's are rather small and/or only a fraction of the  $\beta$ 's differs from zero. To illustrate the consequences of Theorem 1, we compute the empirical distribution of the risk premia estimator  $\hat{\lambda}_F$  for some simulated data-sets. We therefore simulate data from the model,

$$R_t = c + \beta F_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (13)$$

Panel 1: Shapes of densities of  $\hat{\lambda}_F$  for small or zero values of  $\beta$ .

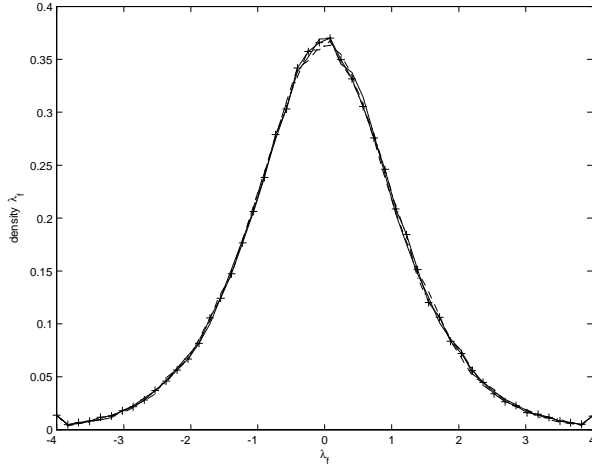


Figure 1.1: Density of  $\hat{\lambda}_F$  when  $\beta = 0$ ,  $\lambda_1 = 1.24$  and  $T = 330$  (solid line), 990 (dashed), 1880 (dashed-dotted), 3300 (solid with plusses)

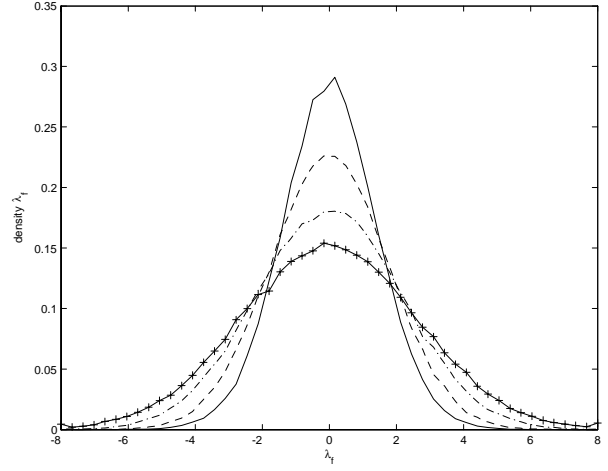


Figure 1.2: Density of  $\hat{\lambda}_F$  when  $\beta = 0$ ,  $\lambda_1 = 1.24$  and  $E(\bar{R}) = c \sim N(0, 0.25^2 I_n)$ ,  $T = 330$  (solid line), 990 (dashed), 1880 (dashed-dotted), 3300 (solid with plusses)

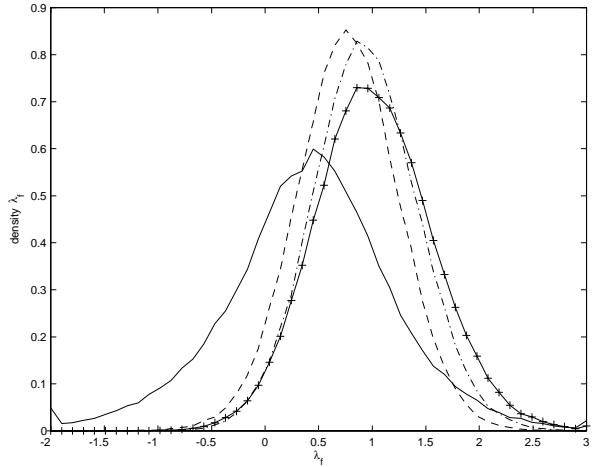


Figure 1.3: Density of  $\hat{\lambda}_F$  with  $\lambda_F = 1$ ,  $\lambda_1 = 1.24$ ,  $\beta = (0.3 \ 0 \ \dots \ 0)'$  and  $n = 2$  (solid with plusses), 10 (dashed-dotted), 24 (dashed), 100 (solid line)

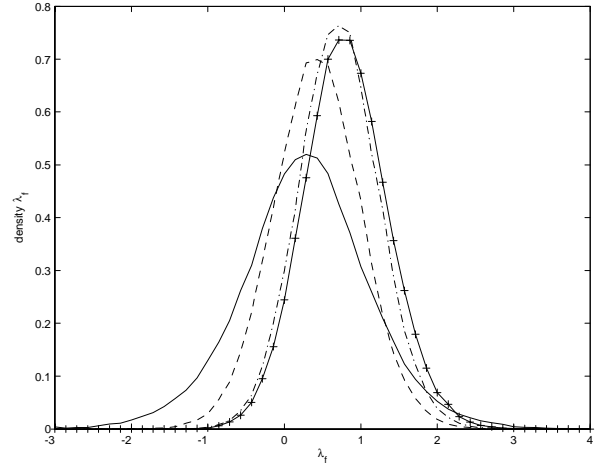


Figure 1.4: Density  $\hat{\lambda}_F$  with  $E(\bar{R}) = (t_n \lambda_1 + \beta, +c)$  with  $\lambda_1 = 1.24$ ,  $\beta = (0.3 \ 0 \ \dots \ 0)'$ ,  $c \sim N(0.04 I_n)$ , and  $n = 2$  (solid with plusses), 10 (dashed-dotted), 24 (dashed), 100 (solid line)

with  $E(F_t) = 0$  and where the  $\varepsilon_t$ 's are independent realizations of  $N(0, \Omega)$  distributed random variables. The parameter settings that we use for the simulated data-sets are obtained from Jagannathan and Wang (1996). Unlike other factor models like those in, for example, Fama and French (1992,1993,1996), the factor model in Jagannathan and Wang (1996) involves factors that do not only consist of portfolio returns, *i.e.* the yield premium and labor income growth. We can therefore estimate the non-zero risk premia on these factors which explains why we use this data-set to illustrate the convergence issues associated with the FM estimator and test statistics.

The data-set of Jagannathan and Wang (1996) consists of monthly observations from July 1963-December 1990 of the return on hundred size and  $\beta$  sorted portfolios so  $n = 100$  and  $T = 330$ . In this simulation experiment, the (demeaned) return on the value weighted portfolio is the only factor. The

covariance matrix  $\Omega$  results from regressing the size and  $\beta$ -sorted portfolio returns on a constant and the return on the value weighted portfolio. Figures 1.1-1.4 in Panel 1 show the different elements of Theorem 1. The value of  $\lambda_1$  that is used for the simulated data-sets results from the FM-estimator (6) when applied to the data from Jagannathan and Wang (1996).

Figure 1.1 contains the sampling density of  $\hat{\lambda}_F$  for different values of the sample size  $T$ . It illustrates Case 1a of Theorem 1. Figure 1.1 shows that  $\hat{\lambda}_F$  converges to a random variable when  $\beta = 0$ , the model is correctly specified and  $T$  converges to infinity since it reflects no convergence of the sampling distribution towards a point mass. The density of  $\hat{\lambda}_F$  is centered around zero because the mean of the demeaned return on the value weighted portfolio is equal to zero. Since  $\beta$  is equal to zero,  $\lambda_F$  is not identified so the density of  $\hat{\lambda}_F$  is centered around a value of  $\lambda_F$  that contains no information about  $\lambda_F$ .

Figure 1.2 illustrates the divergence of  $\hat{\lambda}_F$  as indicated by Case 1b of Theorem 1. In Figure 1.2,  $\beta = 0$  and  $E(\bar{R})$  equals a realization of a normal distributed random variable with mean zero and covariance matrix  $(0.25)^2 I_n$ . Figure 1.2 shows that the variance of  $\hat{\lambda}_F$  rises when the sample size increases. Hence,  $\hat{\lambda}_F$  converges to (plus or minus) infinity when the sample size converges to infinity. Figure 1.2 coincides with Kan and Zhang (1999).

Figure 1.3 shows the influence of adding assets whose  $\beta$ 's equal zero to a factor model whose assets have rather small values of  $\beta$ . The  $\beta$ 's are therefore specified such that only the first one is small and non-zero,  $\beta_1 = 0.3$ . All other elements of  $\beta$  are equal to zero,  $\beta_2 = \dots = \beta_n = 0$ . Figure 1.3 shows the sampling density of  $\hat{\lambda}_F$  for various values of  $n$  while only the  $\beta$  of the first asset is unequal to zero. Figure 1.3 thus corresponds with Case 3 of Theorem 1. For small values of  $n$ ,  $n = 2$ , the density of  $\hat{\lambda}_F$  is centered around the true value of  $\lambda_F$  that is equal to one. When we increase the number of assets  $n$  by adding assets with a zero value of  $\beta$ , the density of  $\hat{\lambda}_F$  becomes centered around zero which equals  $\mu_F$ . Figure 1.3 illustrates the issue of estimating a risk premia for a factor that matters only for a fraction of the assets. Figure 1.3 shows that this estimate becomes more biased when we add assets for which the factor is irrelevant.

Figure 1.4 shows the consequences of misspecification for the parameter setting that is used in Figure 1.3. We therefore added some misspecification to the asset returns,  $E(\bar{R}) = \iota_n \lambda_1 + \beta + c$ , where  $c$  is a realization of a  $N(0, 0.04 I_n)$  distributed random vector. All remaining parameter values are identical to the ones that are used for Figure 1.3. Figure 1.4 shows that the density of  $\hat{\lambda}_F$  moves towards zero and that its' variance increases when we add assets with zero  $\beta$ 's and the model is misspecified. Figure 1.4 reveals a combined effect of the phenomena present in Figures 1.2 and 1.3.

The odd behavior of the FM risk premia estimator for small or zero values of the  $\beta$ 's holds for the GLS risk premia estimator,

$$\begin{pmatrix} \tilde{\lambda}_1 \\ \tilde{\lambda}_F \end{pmatrix} = \left[ (\iota_n \vdots \hat{\beta})' \hat{\Omega}^{-1} (\iota_n \vdots \hat{\beta}) \right]^{-1} (\iota_n \vdots \hat{\beta})' \hat{\Omega}^{-1} \bar{R}, \quad (14)$$

where  $\hat{\Omega}$  is a least squares covariance matrix estimator of  $\Omega$ ,  $\hat{\Omega} = \frac{1}{T-k-1} \sum_{t=1}^T (\bar{R}_t - \hat{\beta} \bar{F}_t)(\bar{R}_t - \hat{\beta} \bar{F}_t)'$ , as well. Theorem 2 states the behavior of the GLS risk premia estimator (14) for small or zero values of the  $\beta$ 's.

**Theorem 2.** *When Assumptions 1 and 2 hold, the behavior of the GLS risk premia estimator (14) in large samples in case of zero or weak  $\beta$ 's is characterized by:*

1. *When  $\beta = 0$  and*



(a)  $E(\bar{R}) = \iota_n \lambda_1 :$

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \lambda_1) \\ \hat{\lambda}_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \end{pmatrix} \psi_R, \quad (15)$$

so  $\hat{\lambda}_F$  converges to a multi-variate  $t$  distributed random variable with location 0, scale parameter  $Q_{11.F} Q_{FF}^{-1}$  and  $n-k$  degrees of freedom and the mean of the limit behavior of  $\hat{\lambda}_F$  equals zero when  $n > k$ .

(b)  $E(\bar{R}) = c :$

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \frac{1}{a} b' \Omega^{-\frac{1}{2}} c) \\ \hat{\lambda}_F - \sqrt{T}(\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} c \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \end{pmatrix} \Omega^{-\frac{1}{2}} \psi_R \quad (16)$$

where  $\Psi_\beta = \text{vecinv}_k(\psi_\beta)$ ,  $a = \iota'_n \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} \iota_n$ ,  $b = M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} \iota_n$ ,  $Q_{11.F} = Q_{11} - Q_{1F} Q_{FF}^{-1} Q_{F1}$ .

2. When  $\beta$  has a weak value such that  $\beta = \frac{1}{\sqrt{T}} B$  and

(a)  $E(\bar{R}) = \iota_n \lambda_1 + \frac{1}{\sqrt{T}} B \lambda_F :$

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \lambda_1) \\ \hat{\lambda}_F - \lambda_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ [(B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta)]^{-1} (B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \\ \Omega^{-\frac{1}{2}} (\psi_{R,\beta} + \Psi_\beta (\rho - \lambda_F)) \end{pmatrix} \quad (17)$$

and the order of finite moments of the limit behavior of  $\hat{\lambda}_F - \lambda_F$  is equal to  $n - k$ .

(b)  $E(\bar{R}) = c :$

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \frac{1}{a} b' \Omega^{-\frac{1}{2}} c) \\ \hat{\lambda}_F - \sqrt{T}[(B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta)]^{-1} (B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} c \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ [(B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta)]^{-1} (B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \end{pmatrix} \Omega^{-\frac{1}{2}} \psi_R, \quad (18)$$

where  $\Psi_\beta = \text{vecinv}_k(\psi_\beta)$ ,  $a = \iota'_n \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} (B + \Psi_\beta)} \Omega^{-\frac{1}{2}} \iota_n$ ,  $b = M_{\Omega^{-\frac{1}{2}} (B + \Psi_\beta)} \Omega^{-\frac{1}{2}} \iota_n$ .

**Proof.** see Appendix D. ■

Theorems 1 and 2 reveal that the FM and GLS risk premia estimators do not have normal limiting distributions in case of small and/or many  $\beta$ 's. Test statistics that are based on the asymptotic normality of these risk premia estimators, like, for example, the Wald  $t$  statistic, therefore also do not converge to a normal or  $\chi^2$  limiting distributions in these cases. Hence, statistical inference that is based on such statistics is unreliable in case of small and/or many  $\beta$ 's. This shows the need for statistics that remain trustworthy in such cases.

### 3 Tests of the risk premium

We propose a number of alternative statistics to test a hypothesis on the risk premia, like, for example,  $H_0 : \lambda_F = \lambda_{F,0}$ , whose limiting distributions are not affected by the value or number of  $\beta$ 's. Since our interest is on the risk premia  $\lambda_F$ , we remove  $\lambda_1$  from the model by removing the return on the  $n$ -th asset and taking all other asset returns in deviation from the return on the  $n$ -th asset.<sup>2</sup> The moment conditions then become

$$\begin{aligned} E(\mathcal{R}_t) &= \mathcal{B}\lambda_F \\ \text{cov}(\mathcal{R}_t, F_t) &= \mathcal{B}\text{var}(F_t) \\ E(F_t) &= \mu_F, \end{aligned} \tag{19}$$

with  $\mathcal{R}_t = R_{1t} - \iota_{n-1}R_{nt}$  and  $\mathcal{B} = \beta_1 - \iota_{n-1}\beta_n$ , for  $R_t = (R'_{1t} \vdots R'_{nt})'$ ,  $\beta = (\beta'_1 \vdots \beta'_n)'$ ;  $R_{1t} : (n-1) \times 1$ ,  $R_{nt} : 1 \times 1$ ,  $\beta_1 : (n-1) \times k$ ,  $\beta_n : 1 \times k$ . Under  $H_0 : \lambda_F = \lambda_{F,0}$ , a least squares estimator for  $\mathcal{B}$  is given by

$$\tilde{\mathcal{B}} = \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_{F,0}) \left[ \sum_{j=1}^T (\bar{F}_j + \lambda_{F,0})(\bar{F}_j + \lambda_{F,0})' \right]^{-1}. \tag{20}$$

We use Assumption 1 to determine the joint behavior of the average returns and the least squares estimator  $\tilde{\mathcal{B}}$  under  $H_0 : \lambda_F = \lambda_{F,0}$  in large samples.

**Lemma 2.** *Under  $H_0 : \lambda_F = \lambda_{F,0}$  and when Assumption 1 holds,*

$$\sqrt{T} \begin{pmatrix} \bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0} \\ \text{vec}(\tilde{\mathcal{B}} - \mathcal{B}) \\ \bar{F} - \mu_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_{\mathcal{R}} \\ \psi_{\mathcal{B}} \\ \psi_F \end{pmatrix},$$

where  $\psi_{\mathcal{R}}$ ,  $\psi_{\mathcal{B}}$  and  $\psi_F$  are independent  $(n-1) \times 1$ ,  $(n-1)k \times 1$  and  $k \times 1$  normal distributed random vectors with mean zero and covariance matrices  $(1 - \lambda'_{F,0}\mathcal{Q}(\lambda_F)_{FF}\lambda_{F,0}) \otimes \Sigma$ ,  $\mathcal{Q}(\lambda_F)_{FF} \otimes \Sigma$  and  $V_{FF}$ , with  $\mathcal{Q}(\lambda_F)_{FF} = \left[ \lim_{T \rightarrow \infty} E \left( \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F)(\bar{F}_t + \lambda_F)' \right) \right]^{-1}$  and  $\Sigma = \Omega_{11} - \iota_{n-1}\omega_{1n} - \omega_{n1}\iota'_{n-1} + \iota_{n-1}\omega_{nn}\iota'_{n-1}$  for  $\Omega = \begin{pmatrix} \Omega_{11} & \omega_{1n} \\ \omega_{n1} & \omega_{nn} \end{pmatrix}$ ,  $\Omega_{11} : (n-1) \times (n-1)$ ,  $\omega_{1n} = \omega'_{n1} : 1 \times (n-1)$  and  $\omega_{nn} : 1 \times 1$ .

**Proof.** see Appendix E. ■

The Wald statistic that is based on the FM risk premia estimator to test  $H_0 : \lambda_F = \lambda_{F,0}$  reads

$$\text{FM-W}(\lambda_{F,0}) = (\hat{\lambda}_F - \lambda_{F,0})' \text{var}(\hat{\lambda}_F)^{-1} (\hat{\lambda}_F - \lambda_{F,0}). \tag{21}$$

For size-able full rank values of  $\beta$  and under  $H_0 : \lambda_F = \lambda_{F,0}$ ,  $\text{FM-W}(\lambda_{F,0})$  converges in distribution to a  $\chi^2(k)$  distributed random variable when the sample size gets large. Theorem 1 shows that  $\hat{\lambda}_F$  converges to a non-normal distributed random variable that is centered around zero for weak and zero values of the  $\beta$ 's. The Wald statistic  $\text{FM-W}(\lambda_{F,0})$  (21) does therefore not converge to a  $\chi^2(k)$  distributed random variable for such values of the  $\beta$ 's. It implies that the Wald statistic is unreliable for conducting inference in these cases. To indicate a solution to this problem, we note that the Wald statistic that

<sup>2</sup>The results are invariant with respect to the asset return that is dropped and with respect to which all assets returns are taken in deviation from.

tests  $H_0$  can be specified as<sup>3</sup>

$$\text{FM-W}(\lambda_{F,0}) = T(\bar{R} - \iota_N \hat{\lambda}_1 - \hat{\beta} \lambda_{F,0})' (\iota_n \dot{\beta}) \left[ (\iota_n \dot{\beta})' \hat{\Theta} (\iota_n \dot{\beta}) \right]^{-1} (\iota_n \dot{\beta})' (\bar{R} - \iota_N \hat{\lambda}_1 - \hat{\beta} \lambda_{F,0}), \quad (22)$$

with  $\hat{\Theta} = \hat{\Omega}(1 + \hat{\lambda}'_F (\frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t)^{-1} \hat{\lambda}_F)$ , see Shanken (1992). Lemma 1 shows that  $\bar{R} - \iota_N \hat{\lambda}_1 - \beta \lambda_{F,0}$  and  $\hat{\beta}$  are independently distributed in large samples. Because of this dependence,  $\bar{R} - \iota_N \hat{\lambda}_1 - \hat{\beta} \lambda_{F,0} = \bar{R} - \iota_N \lambda_1 - \beta \lambda_{F,0} - \iota_n (\hat{\lambda}_1 - \lambda_1) - (\hat{\beta} - \beta) \lambda_{F,0}$ , which is part of FM-W( $\lambda_{F,0}$ ) (22), is, however, not independent of  $\hat{\beta}$  in large samples. This is the origin of the convergence issues with the Wald statistic when the sample size gets large and the  $\beta$ 's are relatively small.

Lemma 2 shows that  $\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}$  and  $\tilde{\mathcal{B}}$  are independently distributed when the sample size gets large. Statistics that are based on  $\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}$  and/or  $\tilde{\mathcal{B}}$  therefore possess better convergence properties than the Wald-statistic. We propose some of these statistics which are based on the GMM and instrumental variable statistics that are proposed in Kleibergen (2002, 2005a) and Moreira (2003). Since the linear factor model is not accommodated by the GMM framework from Kleibergen (2005a), the statistics do not directly result from the GMM and instrumental variable statistics.

**Fama-MacBeth LM statistic** The independence of  $\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}$  and  $\tilde{\mathcal{B}}$  in large samples implies that we can base a statistic with convenient properties on their product.

**Theorem 3.** *Under  $H_0 : \lambda_F = \lambda_{F,0}$  and when Assumption 1 holds,*

$$\text{FM-LM}(\lambda_{F,0}) = \frac{T}{1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F) \tilde{\mathcal{B}}^{-1} \lambda_{F,0}} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\mathcal{B}} (\tilde{\mathcal{B}}' \tilde{\Sigma} \tilde{\mathcal{B}})^{-1} \tilde{\mathcal{B}}' (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}), \quad (23)$$

where  $\hat{\mathcal{Q}}(\lambda_F) = \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F)'$  and  $\tilde{\Sigma} = \frac{1}{T-k} \sum_{t=1}^T (\bar{\mathcal{R}}_t - \tilde{\mathcal{B}} (\bar{F}_t + \lambda_{F,0})) (\bar{\mathcal{R}}_t - \tilde{\mathcal{B}} (\bar{F}_t + \lambda_{F,0}))'$ , converges to a  $\chi^2(k)$  distributed random variable when the sample size gets large for all values of the  $\beta$ 's.

**Proof.** Conditional on  $\tilde{\mathcal{B}} : \sqrt{T} \tilde{\mathcal{B}}' (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}) \xrightarrow{d} \tilde{\mathcal{B}}' \psi_{\mathcal{R}}$  so given  $\tilde{\mathcal{B}} : T (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\mathcal{B}} (\tilde{\mathcal{B}}' \tilde{\Sigma} \tilde{\mathcal{B}})^{-1} \tilde{\mathcal{B}}' (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}) \xrightarrow{d} \psi'_{\mathcal{R}} \tilde{\mathcal{B}} (\tilde{\mathcal{B}}' \tilde{\Sigma} \tilde{\mathcal{B}})^{-1} \tilde{\mathcal{B}}' \psi_{\mathcal{R}}$  which implies the unconditional result for FM-LM( $\lambda_{F,0}$ ) since  $\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}$  and  $\tilde{\mathcal{B}}$  are independent in large samples as shown in Lemma 2 and  $\hat{\mathcal{Q}}(\lambda_F) \xrightarrow{p} \mathcal{Q}(\lambda_F)$ ,  $\tilde{\Sigma} \xrightarrow{p} \Sigma$ . ■

The FM-Lagrange multiplier (LM) statistic is based on the FM-Wald statistic in (22) which explains why we refer to it as a Fama-MacBeth LM statistic. Compared with the FM-Wald statistic (22), the primary differences are the estimators of the  $\beta$ 's,  $\tilde{\mathcal{B}}$  instead of  $\hat{\beta}$ , and that  $\hat{\lambda}_1$  is no longer present since it has been removed by taking the returns in deviation from the returns on the  $n$ -th asset. Identical to the FM risk premia estimator, FM-LM( $\lambda_{F,0}$ ) is not invariant to transformations of the asset returns so it differs in this respect from the standard LM statistics which are invariant to transformations.

**GLS-LM statistic** The GLS risk premia estimator (14) is invariant to transformations of the asset returns so we obtain an invariant LM statistic by incorporating the inverse of the covariance matrix in (23).

---

<sup>3</sup>To obtain this expression of the Wald statistic, we note that  $\text{var}(\hat{\lambda}_1, \hat{\lambda}_F) = \left[ (\iota_n \dot{\beta})' (\iota_n \dot{\beta}) \right]^{-1} (\iota_n \dot{\beta})' \hat{\Theta} (\iota_n \dot{\beta}) \dot{\beta} \left[ (\iota_n \dot{\beta})' (\iota_n \dot{\beta}) \right]^{-1}$  and  $\left[ (\iota_n \dot{\beta})' (\iota_n \dot{\beta}) \right] \left( \left( \frac{\hat{\lambda}_1}{\hat{\lambda}_F} \right) - \left( \frac{\lambda_1}{\lambda_{F,0}} \right) \right) = (\iota_n \dot{\beta})' [\bar{R} - \iota_N \hat{\lambda}_1 - \hat{\beta} \lambda_{F,0}]$ .

**Theorem 4.** Under  $H_0 : \lambda_F = \lambda_{F,0}$  and when Assumption 1 holds,

$$\text{GLS-LM}(\lambda_{F,0}) = \frac{T}{1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0}} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} \tilde{\mathcal{B}} (\tilde{\mathcal{B}}' \tilde{\Sigma}^{-1} \tilde{\mathcal{B}})^{-1} \tilde{\mathcal{B}}' \tilde{\Sigma}^{-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}), \quad (24)$$

converges to a  $\chi^2(k)$  distributed random variable when the sample size gets large for all values of the  $\beta$ 's.

**Proof.** Identical to the proof of Theorem 3. ■

Unlike the FM-LM statistic (23), the GLS-LM statistic (24) is invariant to transformations of the asset returns. The GLS-LM statistic (24) does, however, depend on the inverse of the covariance matrix  $\tilde{\Sigma}$  which can be of a large dimension and therefore difficult to obtain. For example, this is a  $99 \times 99$  matrix in Jagannathan and Wang (1996).

**Factor AR statistic** When the disturbances in the linear factor model (2) are independently normally distributed with mean zero and covariance matrix  $\Omega$ , we can construct the likelihood function of  $\lambda_1$ ,  $\lambda_F$ ,  $\beta$  and  $\Omega$ , see *e.g.* Gibbons (1982). In an identical manner, we can construct the likelihood function for a linear factor model in which the expected asset returns are not restricted to be equal to  $\iota_n \lambda_1 + \beta \lambda_F$ . After concentrating with respect to  $\lambda_1$  and  $\beta$ , the difference between the logarithms of the likelihoods of the unrestricted linear factor model and the restricted factor model under  $H_0 : \lambda_F = \lambda_{F,0}$  is proportional to<sup>4</sup>:

$$\text{FAR}(\lambda_{F,0}) = \frac{T}{1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0}} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}). \quad (25)$$

We refer to this statistic as the factor AR statistic (FAR) since it is similar to the Anderson-Rubin statistic, see Anderson and Rubin (1949), in the instrumental variables regression model. The FAR statistic is proportional to the square of the Hansen-Jagannathan distance evaluated at  $\lambda_{F,0}$  when we use  $\tilde{\Sigma}$  as the covariance matrix estimator, see Hansen and Jagannathan (1997).

**Theorem 5.** Under  $H_0 : \lambda_F = \lambda_{F,0}$  and when Assumption 1 holds, the Factor AR statistic (25) converges to a  $\chi^2(n-1)$  distributed random variable when the sample size gets large for all values of the  $\beta$ 's.

**Proof.** results directly from the large sample behavior of  $\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}$  stated in Lemma 2. ■

Because of the relationship between the FAR statistic (25) and the likelihood function, minimizing the FAR statistic over  $\lambda_{F,0}$  is identical to maximizing the likelihood. Hence, the maximum likelihood estimator (MLE) for  $\lambda_{F,0}$  is obtained by minimizing  $\text{FAR}(\lambda_{F,0})$  over  $\lambda_{F,0}$ .

**Misspecification factor pricing tests** The GLS-LM statistic (24) is a quadratic form of the derivative of  $\text{FAR}(\lambda_{F,0})$  since<sup>5</sup>

$$\frac{\partial}{\partial \lambda_{F,0}} \text{FAR}(\lambda_{F,0}) = c (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} \tilde{\mathcal{B}}, \quad (26)$$

with  $c = \frac{-2T}{1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0}} \left[ 1 + \frac{1}{T-k-1} \text{FAR}(\lambda_{F,0}) \right]$ . The GLS-LM statistic is therefore equal to zero at the MLE of  $\lambda_F$  from Gibbons (1982) which also minimizes  $\text{FAR}(\lambda_{F,0})$ . Besides the MLE, the GLS-LM statistic is also equal to zero at other values of  $\lambda_{F,0}$  which set the derivative of the likelihood to zero, like, for example, local maxima and inflexion points. This hampers inference using the GLS-LM statistic.

<sup>4</sup>We refer to Appendix G for the construction of  $\text{FAR}(\lambda_{F,0})$ .

<sup>5</sup>see Appendix F for a proof.

To overcome the difficulty with the GLS-LM statistic, it is convenient to conduct a pre-test that verifies the validity of the moment conditions  $E(\mathcal{R}_t) = \mathcal{B}\lambda_F$  at  $\lambda_F = \lambda_{F,0}$  which are part of the moment equations (19). A statistic that is well suited for this purpose is the factor pricing statistic

$$\text{JGLS}(\lambda_{F,0}) = \text{FAR}(\lambda_{F,0}) - \text{GLS-LM}(\lambda_{F,0}). \quad (27)$$

**Theorem 6.** *Under  $H_0 : \lambda_F = \lambda_{F,0}$  and when Assumption 1 holds, the factor pricing JGLS( $\lambda_{F,0}$ ) statistic (27) converges to a  $\chi^2(n - k - 1)$  distributed random variable that is independently distributed of the  $\chi^2(k)$  distributed random variable where GLS-LM( $\lambda_{F,0}$ ) (24) converges to when the sample size gets large for all possible values of the  $\beta$ 's.*

**Proof.** results from the large sample behavior of  $\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0}$  documented in Lemma 2. The independence from GLS-LM( $\lambda_{F,0}$ ) results since JGLS( $\lambda_{F,0}$ ) projects  $\tilde{\Sigma}^{-\frac{1}{2}}(\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0})$  on the space orthogonal to  $\tilde{\Sigma}^{-\frac{1}{2}}\tilde{\mathcal{B}}$  where GLS-LM( $\lambda_{F,0}$ ) projects it onto. ■

The factor pricing statistic tests  $E(\mathcal{R}_t) = \mathcal{B}\lambda_F$  for a pre-specified value of the risk premia  $\lambda_F$ . The FM-LM statistic can be used as well for this purpose

$$\text{JFM}(\lambda_{F,0}) = \text{FAR}(\lambda_{F,0}) - \text{FM-LM}(\lambda_{F,0}). \quad (28)$$

**Theorem 7.** *Under  $H_0 : \lambda_F = \lambda_{F,0}$  and when Assumption 1 holds, the factor pricing JFM( $\lambda_{F,0}$ ) statistic (28) converges to a  $\chi^2(n - k - 1)$  distributed random variable that is independently distributed of the  $\chi^2(k)$  distributed random variable where FM-LM( $\lambda_{F,0}$ ) (23) converges to when the sample size gets large for all possible values of the  $\beta$ 's.*

**Proof.** results from the large sample behavior of  $\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0}$  documented in Lemma 2. The independence from FM-LM( $\lambda_{F,0}$ ) results since JFM( $\lambda_{F,0}$ ) projects  $\tilde{\Sigma}^{-\frac{1}{2}}(\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0})$  on the space orthogonal to  $\tilde{\Sigma}^{\frac{1}{2}}\tilde{\mathcal{B}}$  where FM-LM( $\lambda_{F,0}$ ) projects it onto. ■

The factor pricing statistics JGLS( $\lambda_{F,0}$ ) and JFM( $\lambda_{F,0}$ ) resolve the inferential problems with the GLS-LM and FM-LM statistics when we use them as pre-tests before we apply the GLS-LM and FM-LM statistics. A test which tests the adequacy of the moment conditions and  $H_0 : \lambda_F = \lambda_{F,0}$  with  $(1 - \alpha) \times 100\%$  significance then results by applying the JGLS( $\lambda_{F,0}$ ) or JFM( $\lambda_{F,0}$ ) statistics with  $(1 - \alpha_J) \times 100\%$  significance and in case we do not reject, we apply the GLS-LM( $\lambda_{F,0}$ ) or FM-LM( $\lambda_{F,0}$ ) statistics with  $(1 - \alpha_{LM}) \times 100\%$  significance, where  $(1 - \alpha) = (1 - \alpha_J)(1 - \alpha_{LM})$  so  $\alpha = \alpha_J + \alpha_{LM} + \alpha_J\alpha_{LM} \approx \alpha_J + \alpha_{LM}$ , see Kleibergen (2005a,b).

**Conditional likelihood ratio statistic** The likelihood ratio statistic to test  $H_0 : \lambda_F = \lambda_{F,0}$  against  $H_1 : \lambda_F \neq \lambda_{F,0}$  equals twice the difference between the logarithms of the concentrated likelihoods under  $H_1$  and  $H_0$  :

$$\text{LR}(\lambda_{F,0}) = 2 [\max_{\lambda_F} \ln \mathcal{L}_{\text{res}}(\lambda_F) - \ln \mathcal{L}_{\text{res}}(\lambda_{F,0})]. \quad (29)$$

When  $k = 1$ , the likelihood ratio statistic can be specified as<sup>6</sup>

$$\text{CLR}(\lambda_{F,0}) = \frac{1}{2} \left[ \text{FAR}(\lambda_{F,0}) - r(\lambda_{F,0}) + \sqrt{(\text{FAR}(\lambda_{F,0}) + r(\lambda_{F,0}))^2 - 4r(\lambda_{F,0})\text{JFAR}(\lambda_{F,0})} \right], \quad (30)$$

with  $r(\lambda_{F,0}) = \hat{\mathcal{Q}}(\lambda_F)_{FF} \tilde{\mathcal{B}}' \tilde{\Sigma}^{-1} \tilde{\mathcal{B}}$ , which corresponds with the conditional likelihood (CLR) ratio statistic of Moreira (2003) for the linear instrumental variables regression model.

<sup>6</sup>see Appendix G for a derivation of the expression of CLR( $\lambda_{F,0}$ ).

**Theorem 8.** Under  $H_0 : \lambda_F = \lambda_{F,0}$ ,  $k = 1$ , Assumption 1 and given  $r(\lambda_{F,0})$ , the conditional likelihood ratio statistic  $CLR(\lambda_{F,0})$  (30) converges to the random variable:

$$\frac{1}{2} \left[ \varphi_k + \varphi_{n-k-1} - r(\lambda_{F,0}) + \sqrt{(\varphi_k + \varphi_{n-k-1} + r(\lambda_{F,0}))^2 - 4r(\lambda_{F,0})\varphi_{n-k-1}} \right], \quad (31)$$

where  $\varphi_k$  and  $\varphi_{n-k-1}$  are independent  $\chi^2(k)$  and  $\chi^2(n-k-1)$  distributed random variables, when the sample size gets large.

**Proof.** results from the large sample behavior of  $\bar{\mathcal{R}} - \tilde{\mathcal{B}}_{\lambda_{F,0}}$  documented in Lemma 2 and that  $FAR(\lambda_{F,0}) \xrightarrow{d} \varphi_k + \varphi_{n-k-1}$ ,  $JFAR(\lambda_{F,0}) \xrightarrow{d} \varphi_{n-k-1}$ . ■

The expression of the large sample distribution of the CLR statistic in Theorem 8 only applies when  $k = 1$ . It provides a tight upperbound on the large sample distribution of the CLR statistic when  $r(\lambda_{F,0})$  is a statistic that tests for the rank of  $\mathcal{B}$  using  $\tilde{\mathcal{B}}$  when  $k$  exceeds one, see Kleibergen (2005b). The conditional distribution of the CLR statistic (31) is straightforward to simulate from and results from simulating independent  $\chi^2(k)$  and  $\chi^2(n-k-1)$  distributed random variables which are alongside  $r(\lambda_{F,0})$  used to compute the expression in (31). Hence, for every value of  $r(\lambda_{F,0})$ , we use another critical value for  $CLR(\lambda_{F,0})$ . When  $r(\lambda_{F,0})$  is large,  $CLR(\lambda_{F,0})$  corresponds with GLS-LM( $\lambda_{F,0}$ ) while it corresponds with  $FAR(\lambda_{F,0})$  for small values of  $r(\lambda_{F,0})$ .

**Kan and Zhang Misspecification** The factor pricing JGLS and JFM statistics, which test the hypothesis of factor pricing at  $H_0 : \lambda_F = \lambda_{F,0}$ , and the FAR statistic, which tests the joint hypothesis of factor pricing and  $H_0$ , have power against the misspecification analysed by Kan and Zhang (1999). The combinations of the JGLS and GLS-LM statistics and JFM and FM-LM statistics, that were proposed to overcome the spurious power declines of the GLS-LM and FM-LM statistics, therefore also provide power against misspecification of the asset returns.

**MLE-Wald statistic** Besides the FM and GLS Wald statistics, we also compare our novel statistics with the less often used Wald ( $t$ ) statistic that is based on the MLE of Gibbons (1982). Proposition 1 shows that the MLE results from the eigenvectors that belong to the largest roots of a characteristic polynomial.

**Proposition 1.** Under independent normal disturbances with fixed covariance matrix  $\Omega$ , the MLE of  $\lambda_F$  in the linear factor model (3) results from the eigenvectors that belong to the  $k$  largest roots of the characteristic polynomial:

$$\left| \theta \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] - \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \mathcal{R}_t' \right] \Sigma^{-1} \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] \right| = 0, \quad (32)$$

with  $\Sigma = \frac{1}{T} \sum_{t=1}^T R_t R_t'$  :

$$\tilde{\lambda}_F = W_2^{-1} w_1', \quad (33)$$

and where  $w_1 : 1 \times k$ ,  $W_2 : k \times k$  and  $\begin{pmatrix} w_1 \\ W_2 \end{pmatrix}$  contains the  $k$  eigenvectors that belong to the  $k$  largest roots of the characteristic polynomial (32),  $\begin{pmatrix} w_1 \\ W_2 \end{pmatrix} = (q_1 \dots q_k)$ , with  $q_i$ ,  $i = 1, \dots, k$ , the  $(k+1) \times 1$  eigenvectors that belong to the  $k$  largest (of the  $k+1$ ) roots of the characteristic polynomial (32).

**Proof.** The proof is part of the construction of the conditional likelihood ratio statistic in Appendix G. It is straightforward to verify that the derivative (26) is zero at the maximum likelihood estimator when  $\tilde{\mathcal{B}}$  is constructed using  $\tilde{\lambda}_F$ . ■

The MLE can as well be obtained through numerical optimization. The derivative of the likelihood (26) is, however, equal to zero at all  $k+1$  combinations of the eigenvectors that can be used to

construct  $w_1 W_2^{-1}$  (only one of which gives the MLE). Thus the likelihood function has multiple modes and numerical optimization does not necessarily provide the MLE since it can as well provide another value of  $\lambda_F$  where the derivative of the likelihood is zero. Hence, it is preferable (and much more convenient) to compute the MLE using Proposition 1 compared to numerical optimization.

The covariance matrix of the MLE equals<sup>7</sup>

$$\text{var}(\tilde{\lambda}_F) = \frac{1}{T} (1 + \tilde{\lambda}'_F \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} \tilde{\lambda}_F) \left( \tilde{\mathbf{B}}' \tilde{\Sigma}^{-1} \tilde{\mathbf{B}} \right)^{-1}, \quad (34)$$

where  $\tilde{\Sigma}$  uses the MLEs for  $\tilde{\lambda}_F$  and  $\tilde{\mathbf{B}}$ , and has led to the errors in variables correction of the covariance matrix of the FM and GLS estimators proposed by Shanken (1992). For the Wald statistic based on the MLE, we use the covariance matrix estimator (34).

## 4 Power and Size Comparison

We analyze the size and power of the different statistics in a simulation experiment that is based on Jagannathan and Wang (1996). We therefore generate assets returns from the one factor asset pricing model

$$R_t = \iota_n \lambda_1 + \beta (\bar{F}_t + \lambda_F) + \varepsilon_t, \quad t = 1, \dots, T, \quad (35)$$

with  $R_t$  the  $n \times 1$  vector of (excess) asset returns,  $\beta$  the  $n \times 1$  vector of  $\beta$ 's,  $\lambda_1$  the zero- $\beta$  return,  $\lambda_F$  the risk premium and the disturbances  $\varepsilon_t$  are independently normal distributed with mean zero and covariance matrix  $\Omega$ . We use parameter settings that are obtained from the data used by Jagannathan and Wang (1996) so  $n = 100$ ,  $T = 330$ ,  $\bar{F}_t$  is the demeaned return on the value weighted portfolio,  $\lambda_1$  equals the FM estimate  $\hat{\lambda}_1$  and  $\Omega$  equals the least squares covariance matrix estimate.

We generate asset returns from the linear factor model (35) for three different values of  $\beta$ , 1.  $\beta = \hat{\beta}_{\text{VW}}$ , 2.  $\beta = 0.25 \hat{\beta}_{\text{VW}}$  and 3.  $\beta = 0.1 \hat{\beta}_{\text{VW}}$ , with  $\hat{\beta}_{\text{VW}}$  the least squares estimator for the  $\beta$ 's of the asset returns with respect to the value weighted portfolio. We use a range of values of  $\lambda_F$  and the different statistics that we discussed previously to test  $H_0 : \lambda_F = 3$  with 95% significance.

The power curves in Panel 2 reveal the issues involved with the different statistics. The lefthandside of Panel 2, which consists of Figures 2.1, 2.3 and 2.5, displays the power curves of the FM and GLS based Wald and LM statistics and the MLE-Wald (MLE-W) statistic. The righthandside of Panel 2, which consists of Figures 2.2, 2.4 and 2.6, displays the power curves of the CLR, FAR and combinations of the JGLS and GLS-LM statistics and the JFM and FM-LM statistics. These combinations are such that a 99% critical value is applied to JGLS and JFM and a 96% critical value to GLS-LM and FM-LM so the overall size of the tests equals 5%.

The power curves of the FM and GLS Wald statistics in Figure 2.1 for which  $\beta = \hat{\beta}_{\text{VW}}$  show that the rejection frequency of the FM Wald statistic is approximately 5% at  $\lambda_F = 3$  while the size of the MLE and GLS Wald statistics differ considerably from 5%. Hence, these statistics are size distorted. Figures 2.3 and 2.5, for which  $\beta$  equals  $0.25 \hat{\beta}_{\text{VW}}$  and  $0.1 \hat{\beta}_{\text{VW}}$  resp., show that all Wald statistics are size distorted for smaller values of  $\beta$  while the FM and GLS-LM statistics are not size distorted. These Figures also demonstrate a spurious decline of power of the FM and GLS-LM statistics at values of  $\lambda_F$  which are considerable different from the hypothesized value of three. These power declines result because the FM and GLS LM statistics are proportional to quadratic forms of derivatives of objective functions with respect to  $\lambda_F$ . Hence, the FM-LM and GLS-LM statistics are equal to zero at minima, maxima and inflexion points of these statistics which explains the spurious power declines of the FM and GLS-LM statistics.

---

<sup>7</sup>The covariance matrix is constructed in Appendix I.

Panel 2: Power curves of statistics that test  $H_0 : \lambda_F = 3$  with 95% significance

$\beta = \hat{\beta}_{VW} :$

Figure 2.1

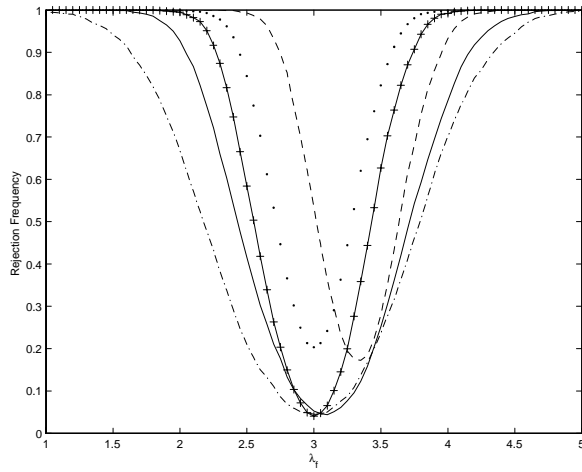
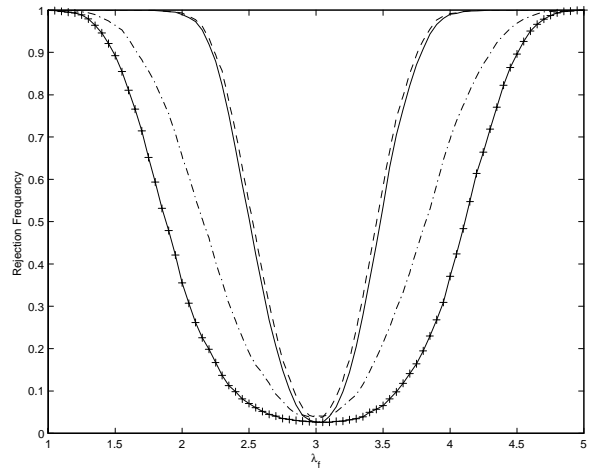


Figure 2.2



$\beta = 0.25\hat{\beta}_{VW} :$

Figure 2.3

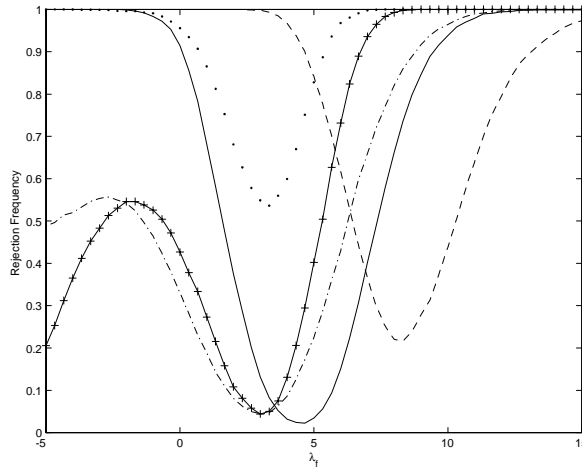
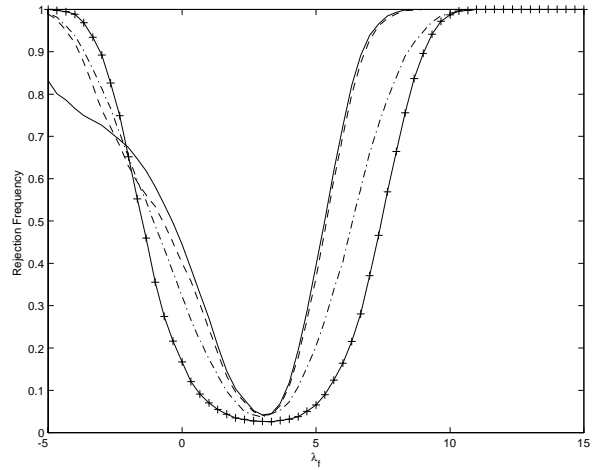


Figure 2.4



$\beta = 0.1\hat{\beta}_{VW} :$

Figure 2.5

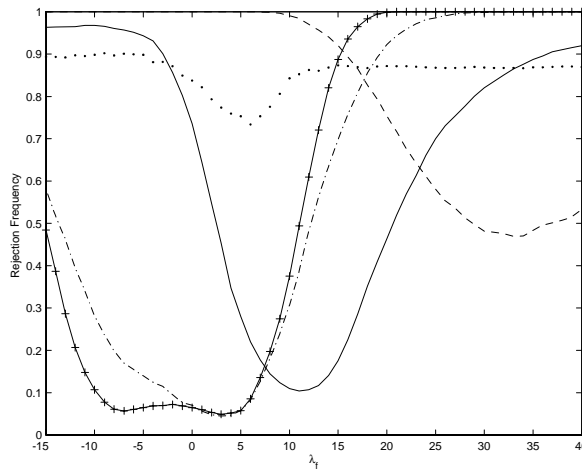
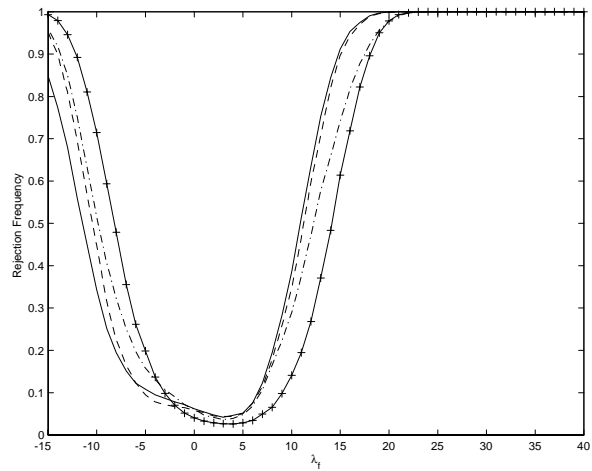


Figure 2.6



Figures 2.1, 2.3, 2.5: Power curves of FM-W (solid), GLS-W (dashed), FM-LM (dashed-dotted), MLE-W (points) and GLS-LM (solid with plusses) statistics.

Figures 2.2, 2.4, 2.6: Power curves of CLR (solid), combination of JGLS and GLS-LM (dashed), combination of JFM and FM-LM (dashed-dotted) and FAR (solid with plusses) statistics.



It is remarkable that the GLS and MLE Wald statistics are already size distorted when  $\beta$  equals  $\hat{\beta}_{VW}$  in Figure 2.1. This size distortion results from the inversion of the  $100 \times 100$  ( $99 \times 99$ ) covariance matrices  $\hat{\Omega}$  and  $\tilde{\Sigma}$ . Appendix H explains why this size distortion does not occur for the other statistics that also contain inverses of  $99 \times 99$  covariance matrices.

It is interesting to note that although the MLE Wald statistic is size-distorted that the minimum of the power curve remains approximately at the hypothesized value of three while it shifts upwards for the GLS and FM Wald statistics when the value of  $\beta$  gets smaller. This results from the downward bias of the FM and GLS risk premia estimators as stated in Theorem 1. Because of the downward biases of these estimators, they are on average smaller than the true  $\lambda_F$  so the power curves attain their minima at values of  $\lambda_F$  which are larger than the hypothesized one. The power curve of the MLE Wald statistic has its minimum approximately at the hypothesized value which indicates that the bias of the MLE is smaller than that of the FM and GLS estimators. Another interesting feature is that this bias does not depend on the value of  $\beta$  while, as shown by Theorem 1 and the power curves of the FM and GLS Wald statistics, the bias of the FM and GLS Wald statistics increases when  $\beta$  gets smaller. For GMM, it is known that the estimators that are comparable to the MLE in GMM have smaller biases than the estimators that are comparable to the FM and GLS estimators, see *e.g.* Newey and Smith (2004) so similar results hold in linear factor models as well.

Figures 2.2, 2.4 and 2.6 only contain power curves of statistics whose large sample distributions are unaffected by small or zero values of the  $\beta$ 's. These Figures therefore reveal no size distortion for any of the depicted statistics. Figures 2.2-2.6 show that the power of the FAR statistic is in general below the power of the other (combinations of) statistics but the FAR statistic has similar power than the other statistics when the  $\beta$ 's are very small. The smaller power of the FAR statistic in case of reasonably large  $\beta$ 's results from the larger degree of freedom parameter of its large sample distribution compared to the other statistics.

Figures 2.2-2.6 show that the combinations of the JGLS and GLS-LM statistics and JFM and FM-LM statistics overcome the spurious power decline of the GLS-LM and FM-LM statistics. These combined statistics therefore perform appropriately and also provide power against misspecification, *i.e.* no factor pricing. The combination of the JGLS and GLS-LM statistics and the CLR statistics have comparable power curves and outperform the combination of the JFM and FM-LM statistics. We note that the CLR statistic has no power against misspecification.

The power curves in Panel 2 clearly show the appeal of the statistics proposed in the previous section. Compared to the FM-Wald statistic, the optimal configuration of the statistics, which are the combined JGLS and GLS-LM statistics and the CLR statistic, do not indicate any sacrifice of power to improve the size. For example, in Figure 2.1, these statistics already dominate the FM-Wald statistic in terms of power. This continues to hold in the other Figures where the FM-Wald statistics eventually become enormously size distorted for very small values of the  $\beta$ 's.

## 5 Tests on some of the risk premia

The statistics that we proposed so far test hypotheses that are specified on all risk premia,  $H_0 : \lambda_F = \lambda_{F,0}$ . Instead of testing hypotheses specified on all parameters, we often want to test hypotheses that are specified on sub-sets of the parameters, for example,  $H_0^* : \theta_F = \theta_{F,0}$ , where  $\lambda_F = (\nu_F' : \theta_F')$ , with  $\nu_F : k_\nu \times 1$ ;  $\theta_{F,0}, \theta_F : k_\theta \times 1$  and  $k = k_\nu + k_\theta$ . Kleibergen (2006) shows that the limiting distributions of the sub-set statistics in the linear instrumental variables regression model are bounded by limiting distributions identical to those in Theorems 3-8 when the partialled out parameters ( $\nu_F$ ) are estimated by maximum likelihood. This result extends to the linear factor model. The maximum likelihood estimator for  $\nu_F$  given  $H_0^* : \theta_F = \theta_{F,0}$  results in a straightforward manner from the characteristic

polynomial in Proposition 1 when we replace  $\mathcal{R}_t$  and  $(\frac{1}{\bar{F}_t})$  by the residuals of time series regression of  $\mathcal{R}_t$  and  $(\frac{1}{\bar{F}_{t,1}})$  on  $\bar{F}_{t,2} + \theta_{F,0}$ , with  $\bar{F}_t = (\bar{F}'_{t,1} : \bar{F}'_{t,2})'$ ,  $\bar{F}_{t,1} : k_\nu \times 1$ ,  $\bar{F}_{t,2} : k_\theta \times 1$ .

**Theorem 9.** Under  $H_0^* : \theta_F = \theta_{F,0}$ , when Assumption 1 holds and  $\tilde{\nu}_F(\theta_{F,0})$  is the maximum likelihood estimator of  $\nu_F$ , the large sample distribution(s) of

1.  $FM-LM((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$  and  $GLS-LM((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$  are bounded from above by  $\chi^2(k_\theta)$  distributions.
2.  $FAR((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$  is bounded from above by a  $\chi^2(n - k_\nu - 1)$  distribution.
3.  $JFM((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$  and  $JGLS((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$  are bounded from above by  $\chi^2(n - k - 1)$  distributions which are independent from the large sample distributions of  $FM-LM((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$  and  $GLS-LM((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$ .
4.  $CLR((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$  given  $r((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$  is bounded from above by

$$\frac{1}{2} \left[ \varphi_{k_\theta} + \varphi_{n-k-1} - r((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})') + \sqrt{(\varphi_{k_\theta} + \varphi_{n-k-1} + r((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})'))^2 - 4r((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')\varphi_{n-k-1}} \right],$$

where  $\varphi_{k_\theta}$  and  $\varphi_{n-k-1}$  are independent  $\chi^2(k_\theta)$  and  $\chi^2(n - k - 1)$  distributed random variables and  $r((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$  is the smallest eigenvalue of  $\hat{Q}((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})'_{FF})^{\frac{1}{2}'} \tilde{B}' \tilde{\Sigma}^{-1} \tilde{B} \hat{Q}((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})'_{FF})^{\frac{1}{2}}$  and  $\tilde{B}$ ,  $\tilde{\Sigma}$  are computed using  $(\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})'$ .

for all possible values of  $B_\theta$ .

**Proof.** see Kleibergen (2006). ■

Theorem 9 shows that the statistics that we constructed previously can be used to conduct tests on some of the risk premia and become conservative when we do so. This results holds for all possible values of the  $\beta$ 's.

## 6 Confidence sets of the risk premia from Jagannathan and Wang (1996)

When we specify a range of values of  $\lambda_{F,0}$ , or  $\theta_{F,0}$ , we can use the previously discussed statistics to constructs confidence sets for the risk premia. The  $(1 - \alpha) \times 100\%$  asymptotic confidence set contains all values of  $\lambda_{F,0}$ , or  $\theta_{F,0}$ , for which the value of the statistic that is used to test  $H_0 : \lambda_F = \lambda_{F,0}$ , or  $H_0^* : \theta_F = \theta_{F,0}$ , is below its  $(1 - \alpha) \times 100\%$  asymptotic critical value.

We construct the asymptotic confidence sets for the risk premia for the asset returns from Jagannathan and Wang (1996). Jagannathan and Wang (1996) use the factor model,

$$R_t = \iota_n \lambda_1 + \text{size} \times \lambda_s + \beta(\bar{F}_t + \lambda_F) + \varepsilon_t, \quad t = 1, \dots, T, \quad (36)$$

	FM	FM-JW	GLS	MLE
$\lambda_{vw}$	-0.25 (0.25)	-0.40 (0.46)	-0.17 (0.17)	-0.36 (0.30)
$\lambda_{premi}$	0.18 (0.066)	0.20 (0.073)	0.11 (0.044)	0.47 (0.10)
$\lambda_{labor}$	0.09 (0.047)	0.10 (0.053)	0.097 (0.033)	0.41 (0.079)

Table 1: Risk premia estimates for the Jagannathan-Wang (1996) data (Standard errors between brackets).

to describe the return on hundred size and  $\beta$  sorted portfolios that are collected into the vector of asset returns  $R_t$ , so  $n = 100$ , for three hundred and thirty monthly observations, so  $T = 330$ . The  $n \times 1$  vector size reflects the relative size (market capitalization) of the different portfolios and is constant over time. The  $k \times 1$  vector of (demeaned) factors  $\bar{F}_t$  contains three different factors (so  $k = 3$ ): the return on a value weighted portfolio, the yield premium between low and high grade corporate bonds and the growth of per capita labor income. Since our interest is on the risk premia  $\lambda_F$ , we remove  $\lambda_1$  and  $\lambda_s$  from the factor model (36) in an analogous manner as by taking the returns in deviation from the  $n$ -th asset return as we used previously to remove  $\lambda_1$ .<sup>8</sup> The removal of the size factor implies that the value of  $n$  in the degrees of freedom parameters of the (bounding) limiting distributions in Theorem 9 reduces to  $n - 1$  (and  $n - 1$  therefore to  $n - 2$ ).

Table 1 contains estimates of the risk premium on the value weighted portfolio ( $\lambda_{vw}$ ), the yield premium ( $\lambda_{premi}$ ) and labor income growth ( $\lambda_{labor}$ ). The estimators that we use are the FM estimator (FM), the GLS estimator (GLS), the FM estimator as used by Jagannathan and Wang (FM-JW)<sup>9</sup> and the MLE. The estimates in Table 1 show a size-able dispersion between the risk premia estimates on the yield premium and labor income growth that result from the FM and GLS procedures compared to the MLE. This results from the downward bias of the FM and GLS estimators when the number of assets is large and the  $\beta$ 's are relatively small. The  $\beta$ 's of the yield premium and labor income growth are rather small while the  $\beta$ 's of the value-weighted return portfolio are sizeable. This explains the similarity of the risk premia estimates for the value-weighted return portfolio and the differences for the risk premia on the yield premium and labor income growth.

Table 1 reports the standard errors of the estimates of the risk premia. We use these standard errors alongside with the statistics discussed previously to construct 95% asymptotic confidence sets for the different risk premia. Panel 3 contains the  $p$ -value plots of the tests on the risk premia using the different statistics. Figures 3.1 and 3.2 in Panel 3 show the  $p$ -value plots of tests of the hypothesis  $H_0 : \lambda_{vw} = \lambda_{vw,0}$ , where  $\lambda_{vw}$  is the risk premium on the value weighted portfolio, for a range of values of  $\lambda_{vw,0}$ . Figures 3.3 and 3.4 contain the  $p$ -value plots of tests of  $H_0 : \lambda_{premi} = \lambda_{premi,0}$ , where  $\lambda_{premi}$  is the risk premium on the yield premium, for a range of values of  $\lambda_{premi,0}$  and Figures 3.5 and 3.6 contain the  $p$ -value plots of tests of  $H_0 : \lambda_{labor} = \lambda_{labor,0}$ , where  $\lambda_{labor}$  is the risk premium on the labor income growth, for a range of values of  $\lambda_{labor,0}$ . The tests of the hypotheses on the risk premia are all conducted in the linear factor model (36) and test only one element of the vector of risk premia. The MLE is used to estimate the other risk premia in case of the GLS-LM, JGLS, FAR and CLR statistics

<sup>8</sup>The size and constant term  $\iota_n$  are removed from the factor model (36) by projecting  $R_t$  and  $\beta$  on a  $n - 2$  dimensional space that is orthogonal to  $(\iota_n : \text{size})$ . Hence, we use  $\mathcal{R}_t = AR_t$  and  $\mathcal{B} = A\beta$ , with  $A : (n - 2) \times n$ ,  $A(\iota_n : \text{size}) = 0$ . We note that this specification is not unique (but the resulting statistics are) and  $A$  equals, for example,  $(I_{n-1} : -\iota_{n-1})$  in the case when size is not present. After this transformation, the moment conditions are identical to (19).

<sup>9</sup>Jagannathan and Wang (1996) estimate the risk premia using the FM estimator with a  $\beta$ -vector where the  $\beta$ 's of the different factors result from three separate time-series regressions of the excess returns on a constant term and the respective factor.

and the FM estimator is used for the FM-LM statistic. The  $p$ -values for these statistics result from the bounding limiting distributions in Theorem 9 and are thus conservative. The  $p$ -values for the CLR statistic are obtained using its bounding conditional limiting distribution given  $r((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$  which is computed for a large range of values of  $r((\tilde{\nu}_F(\theta_{F,0})' : \theta'_{F,0})')$ . All Figures in Panel 3 contain a dotted line at 95% whose intersection with the  $p$ -value plots reveals the 95% confidence set.

Figures 3.1 and 3.2 show the  $p$ -values for tests on the risk premium of the value weighted return. Figure 3.1 (and Figures 3.3 and 3.5) contain the  $p$ -value plots that result from the FM-W, GLS-W, FM-LM, MLE-W and GLS-LM statistics. Most of the  $p$ -value plots in Figure 3.1 are rather similar which results since the  $\beta$ 's of the value weighted return are sizeable. The main difference between the  $p$ -value plots of the FM, GLS, MLE Wald statistics and the GLS-LM statistic is therefore that the latter leads to a slightly larger 95% confidence set. The  $p$ -value plot of the FM-LM statistic is rather strange and given its inferior behavior in the power study compared to the size-correct GLS-LM and CLR statistics, we do not discuss this statistic any further. All statistics do not reject the hypothesis of a zero risk premium at the 95% significance level which is as expected since the risk premium on portfolio returns, like, for example, the value weighted index, is typically assumed to be equal to zero.

Figure 3.2 (and Figures 3.4, 3.6) only contains  $p$ -value plots of statistics that remain size-correct in case of small and/or many  $\beta$ 's, *i.e.* the GLS-LM, CLR, JGLS and FAR statistics. The  $p$ -value plots of the JGLS and FAR statistic are all rather low which is due to the large degree of freedom parameter of their (bounding)  $\chi^2$  limiting distributions, *i.e.* 95 and 96. These  $p$ -value plots show that the factor pricing hypothesis is not rejected. The  $p$ -value plots of the GLS-LM and CLR statistic are very similar which results since the conditioning statistic which is used in the CLR statistic is rather large. None of the statistics rejects the hypothesis that the risk premium on the value-weighted return is zero.

Figures 3.3 and 3.4 show the  $p$ -value plots for tests of the risk premium on the yield premium. In line with the risk premia estimates in Table 1, the  $p$ -value plots in Figure 3.3 differ considerably. Because of the downward bias of the FM and GLS estimators, which results since the  $\beta$ 's on the yield premium are rather small and probably only non-zero for a fraction of the number of portfolios, the 95% confidence set for the yield risk premium lies at rather small values that lie just above zero when we use the FM and GLS estimators. The GLS-LM, MLE-W, CLR and FAR statistics all lead to 95% confidence sets that are centered around the MLE. All these statistics therefore lead to 95% confidence sets for the yield premium that contain much larger values of the yield risk premium than the 95% confidence sets that result from the FM and GLS Wald ( $t$ ) statistics. The 95% confidence sets that result from the GLS-LM and CLR statistics in Figure 3.4 are rather similar and are much larger than the 95% confidence set that results from the MLE-W statistic. The power and size analysis in Panel 2 shows that the GLS-LM and CLR statistics are much more trustworthy than the MLE-W statistics and also than the FM and GLS Wald statistics. Hence, the 95% confidence sets that result from the GLS-LM and CLR statistics are more reliable than those that result from the FM, GLS and MLE Wald statistics. The JGLS statistic in Figure 3.4 shows that the factors are the only priced elements and factor pricing is thus not rejected.

Figures 3.5 and 3.6 show the  $p$ -values for tests of the risk premium on labor income growth. Because the  $\beta$ 's on labor income growth are rather small and are probably only non-zero for a fraction of the large number of portfolios, there is again a considerable downward bias in the FM and GLS risk premium estimators compared to the MLE as shown in Table 1. The  $p$ -value plots that result from the FM and GLS Wald statistics and the GLS-LM and MLE-W statistics therefore differ considerably. The 95% confidence set that results from the GLS-LM statistic is much larger than the one that results from the FM Wald statistic and contains much larger values of the risk premium. It is striking to see that the intersection of both confidence sets is even empty. The  $p$ -value plot of the CLR statistic in Figure 3.6 is similar to that of the GLS-LM statistic. The JGLS and FAR statistics in Figure 3.6 show that no mispricing is occurring and that factor pricing is not rejected.

Panel 3: One minus  $p$ -value plots for  $\lambda_{vw}$ ,  $\lambda_{\text{prem}}$  and  $\lambda_{\text{labor}}$ .

$\lambda_{vw}$  :

Figure 3.1

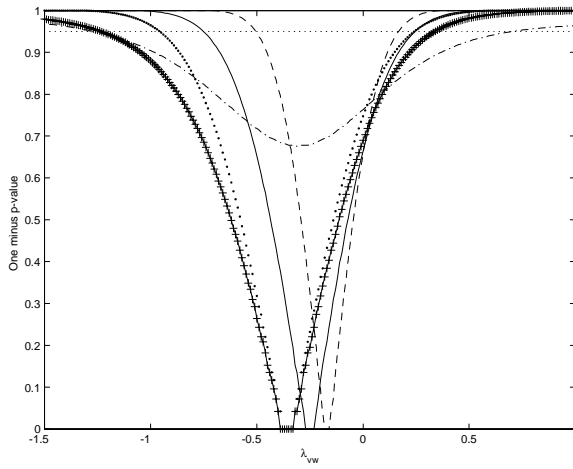
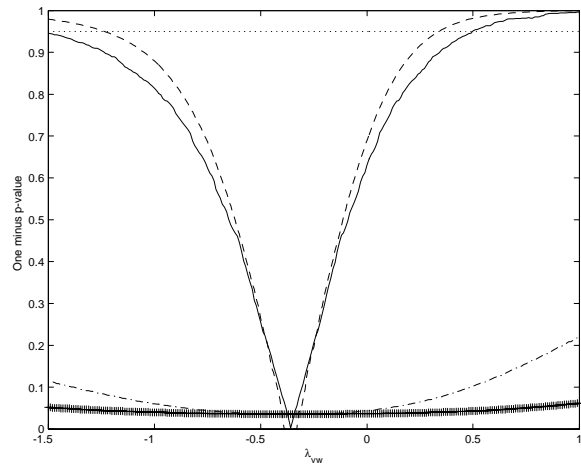


Figure 3.2



$\lambda_{\text{prem}}$  :

Figure 3.3

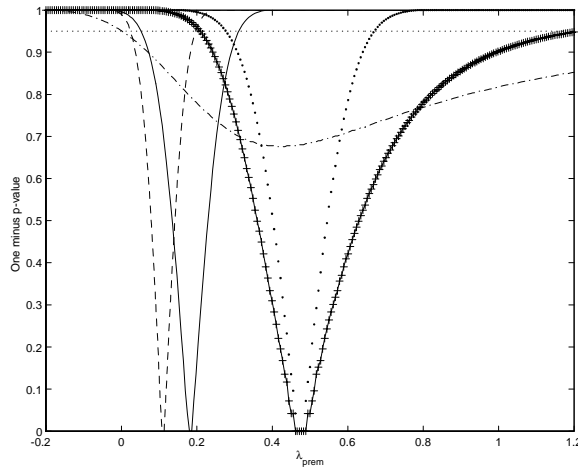
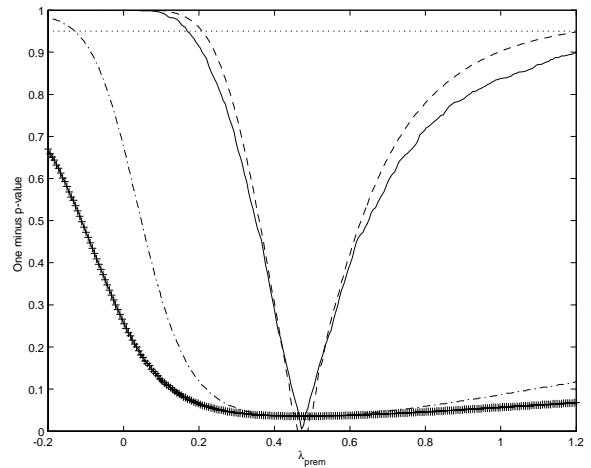


Figure 3.4



$\lambda_{\text{labor}}$  :

Figure 3.5

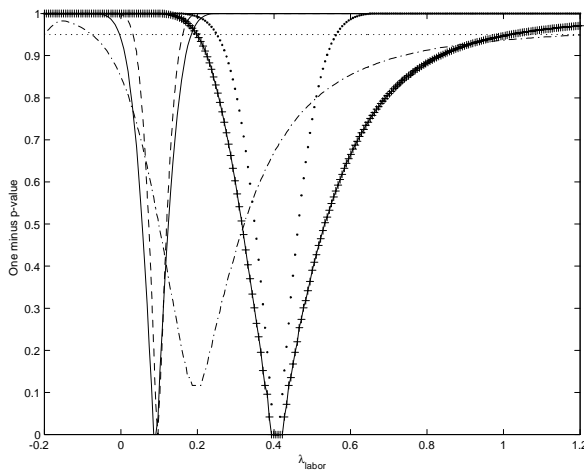


Figure 3.6

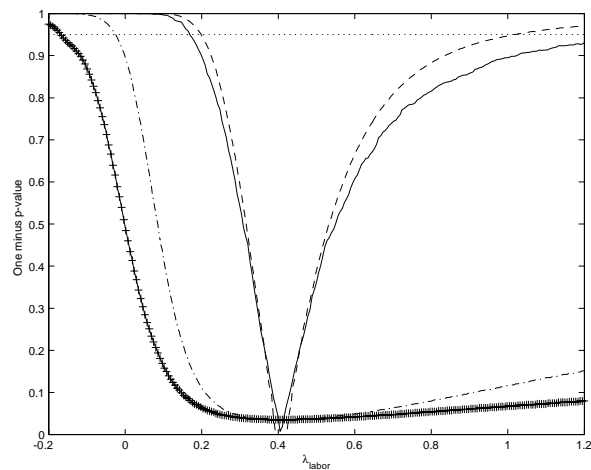


Figure 3.1-3.5: One minus  $p$ -value plot for FM-W (solid), GLS-W (dashed), FM-LM (dashed-dotted) MLE-W (points) and GLS-LM (solid with plusses) statistics.

Figure 3.2-3.6: One minus  $p$ -value plot for CLR (solid), JGLS (solid with plusses), FAR (dashed-dotted) and GLS-LM (dashed) statistics.

Panel 3 shows the importance of using the size-correct statistics that we introduced previously. Because of the large number of portfolios in the Jagannathan-Wang data, the bias in the FM and GLS risk premia estimator is considerable when the  $\beta$ 's of the associated factors are small and/or only non-zero for a fraction of the portfolios. The Wald statistic that is based on the FM risk premia estimator is therefore size distorted and unreliable. The statistics that we introduced remain size-correct and lead to inference that is centered around the MLE. These statistics therefore do a better job in reflecting the risk premium on the factors whose  $\beta$ 's are relatively small and/or only matter for a fraction of the assets. This is nicely shown for the Jagannathan-Wang data where the confidence sets that result from the FM Wald statistic show that the risk premia are almost negligible while the confidence sets that result from the CLR and GLS-LM statistics indicate sizeable and thus important risk premia. This shows that the usage of these improved inferential procedures can lead to more interesting conclusions.

## 7 Conclusions

We reveal the inadequacy of standard inferential procedures that are based on the FM and GLS risk premia estimators when the  $\beta$ 's are small and/or the number of assets is large. We propose a number of alternative statistics whose large sample distributions remain trustworthy in such cases and lead to inference that is centered around the maximum likelihood estimator of Gibbons (1982). We provide a simple manner to compute the maximum likelihood estimator. We apply these statistics to the size and  $\beta$  sorted portfolio return series from Jagannathan and Wang (1996). The confidence sets that result from the FM and GLS Wald statistics show that the risk premia on the yield premium and labor income growth are rather small and perhaps negligible. The confidence sets that result from the new statistics indicate a more interesting conclusion as they indicate much larger values of the risk premia which thus show that these risk premia are economically relevant.

In many cases the return vector that is used in the linear factor model consists of the return on  $\beta$  and size sorted portfolios, see *e.g.* Fama and French (1992,1993,1996) and Jagannathan and Wang (1996,1998). This sorting introduces a further element of correlation between the average return vector that is used to test the risk premia and the estimator of  $\beta$ . We showed that the size distortions of the FM and GLS Wald/ $t$  statistics originate from this kind of correlation. Our proposed statistics use  $\beta$  estimators that are uncorrelated with the average return vector that is used to test the risk premia and are therefore not size distorted in case of small and/or many  $\beta$ 's. When the returns result from some kind of selection procedure, the  $\beta$  estimators and average returns that are used in the proposed statistics remain uncorrelated. Hence, these statistics are size correct when the returns result from a selection procedure.

The linear factor models analysed in this paper assumes a constant covariance matrix of the asset returns over time. We plan to relax this assumption in future work.

## Appendix

**A. Assumption 1.** The central limit theorem for the specification of the linear factor model (2) implied by Assumption 1 reads

$$\begin{aligned}
& \frac{1}{\sqrt{T}} \sum_{t=1}^T \left( \begin{array}{c} 1 \\ F_t \\ F_t - \mu_f \end{array} \otimes u_t \right) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left( \begin{array}{c} 1 \\ F_t \\ F_t - \mu_f \end{array} \otimes (R_t - \iota_n \lambda_1 - \beta(F_t - \mu_F + \lambda_F)) \right) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left( \begin{array}{c} 1 \\ F_t \\ F_t - \mu_f \end{array} \otimes (R_t - \iota_n \lambda_1 - \beta(F_t - \bar{F} + \bar{F} - \mu_F + \lambda_F)) \right) \\
&= \frac{1}{\sqrt{T}} \left( \sum_{t=1}^T \begin{array}{c} 1 \\ F_t \end{array} \otimes (R_t - \iota_n \lambda_1 - \beta(\bar{F}_t + \lambda_F)) - \sum_{t=1}^T \begin{array}{c} 1 \\ F_t \end{array} \otimes \beta(\bar{F} - \mu_F) \right) \\
&= \left( \begin{array}{c} \frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{array}{c} 1 \\ F_t \end{array} \otimes (R_t - \iota_n \lambda_1 - \beta(\bar{F}_t + \lambda_F)) - \sqrt{T} \begin{array}{c} 1 \\ \bar{F} \end{array} \otimes \beta(\bar{F} - \mu_F) \\ \sqrt{T}(\bar{F} - \mu_f) \end{array} \right) \\
&\xrightarrow{d} \begin{pmatrix} \varphi_R - \beta\varphi_F \\ \varphi_\beta - (\mu_F \otimes \beta)\varphi_F \\ \varphi_F \end{pmatrix}.
\end{aligned}$$

since  $\bar{F} \xrightarrow{p} \mu_F$  and which results in convergence to a normal distributed random vector with mean zero and covariance matrix

$$\left( \begin{array}{cc} \begin{pmatrix} 1 & \mu'_F \\ \mu_F & \mu_F \mu'_F \end{pmatrix} \otimes (\Omega + \beta V_{FF} \beta') + \begin{pmatrix} 0 & 0 \\ 0 & V_{FF} \end{pmatrix} \otimes \Omega & \begin{array}{c} -\beta V_{FF} \\ -(\mu_F \otimes \beta) V_{FF} \\ V_{FF} \end{array} \\ \begin{array}{cc} -V_{FF} \beta' & -V_{FF} (\mu_F \otimes \beta)' \end{array} & \end{array} \right).$$

**B. Proof of Lemma 1.** Assumption 1 states that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left( \begin{array}{c} 1 \\ F_t \\ F_t - \mu_F \end{array} \otimes (R_t - \iota_n \lambda_1 - \beta(\bar{F}_t + \lambda_F)) \right) \xrightarrow{d} \begin{pmatrix} \varphi_R \\ \varphi_\beta \\ \varphi_F \end{pmatrix},$$

with  $(\varphi'_R \vdots \varphi'_\beta \vdots \varphi'_F)' \sim N(0, V)$ . We pre-multiply it by

$$\left( \begin{array}{cc} 1 & 0 \\ -(\frac{1}{T} \sum_{j=1}^T \bar{F}_j \bar{F}'_j)^{-1} \bar{F} & (\frac{1}{T} \sum_{j=1}^T \bar{F}_j \bar{F}'_j)^{-1} \\ 0 & 0 \end{array} \right) \otimes \begin{pmatrix} I_n & 0 \\ 0 & I_k \end{pmatrix}$$

to obtain, since  $\sum_{t=1}^T \bar{F}_t = 0$ ,  $\hat{\beta} = \sum_{t=1}^T R_t \bar{F}'_t (\sum_{j=1}^T \bar{F}_j \bar{F}'_j)^{-1}$ ,

$$\begin{aligned}
& \sum_{t=1}^T [(\iota_n \lambda_1 + \beta(\bar{F}_t + \lambda_F)) F'_t - (\iota_n \lambda_1 + \beta(\bar{F}_t + \lambda_F)) \bar{F}'_t] (\sum_{j=1}^T \bar{F}_j \bar{F}'_j)^{-1} = \\
& \sum_{t=1}^T [\beta \bar{F}_t (F_t - \bar{F})'] (\sum_{j=1}^T \bar{F}_j \bar{F}'_j)^{-1} = \beta,
\end{aligned}$$

and

$$\lim_{T \rightarrow \infty} \begin{pmatrix} 1 & 0 \\ -(\frac{1}{T} \sum_{j=1}^T \bar{F}_j \bar{F}_j')^{-1} \bar{F} & (\frac{1}{T} \sum_{j=1}^T \bar{F}_j \bar{F}_j')^{-1} \end{pmatrix} Q \begin{pmatrix} 1 & 0 \\ -(\frac{1}{T} \sum_{j=1}^T \bar{F}_j \bar{F}_j')^{-1} \bar{F} & (\frac{1}{T} \sum_{j=1}^T \bar{F}_j \bar{F}_j')^{-1} \end{pmatrix}' = \begin{pmatrix} Q_{11} & 0 \\ 0 & Q_{FF.1}^{-1} \end{pmatrix}$$

with  $Q_{FF.1} = \lim_{T \rightarrow \infty} E(\frac{1}{T} \sum_{j=1}^T \bar{F}_j \bar{F}_j')$ , that

$$\sqrt{T} \begin{pmatrix} \bar{R} - \iota_n \lambda_1 - \beta \lambda_F \\ \text{vec}(\hat{\beta} - \beta) \\ \bar{F} - \mu_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_R \\ \psi_\beta \\ \psi_F \end{pmatrix},$$

where  $\psi_R$ ,  $\psi_\beta$  and  $\psi_F$  are independent  $n \times 1$ ,  $nk \times 1$  and  $k \times 1$  normal distributed random vectors with mean 0 and covariance matrices,  $\Omega$ ,  $Q_{FF.1}^{-1} \otimes \Omega$  and  $V_{FF}$ .

**C. Proof of Theorem 1.** To analyze the behavior of the FM estimator, we rephrase Lemma 1 as

$$\begin{pmatrix} \bar{R} \\ \text{vec}(\hat{\beta}) \\ \bar{F} \end{pmatrix} \underset{a}{=} \begin{pmatrix} \iota_n \lambda_1 + \beta \lambda_k \\ \text{vec}(\beta) \\ \mu_F \end{pmatrix} + \frac{1}{\sqrt{T}} \begin{pmatrix} \psi_R \\ \psi_\beta \\ \psi_F \end{pmatrix} + o_p(\frac{1}{\sqrt{T}}),$$

where  $\underset{a}{=}$  implies equality in large samples and  $o_p(\frac{1}{\sqrt{T}})$  indicates that all other elements are of a lower order than  $\frac{1}{\sqrt{T}}$ .

1.  $\beta = 0$ . We use that since  $(\iota_n : \frac{1}{\sqrt{T}} \Psi_\beta)' (\iota_n : \frac{1}{\sqrt{T}} \Psi_\beta) = \begin{pmatrix} \iota_n' \iota_n & \frac{1}{\sqrt{T}} \iota_n' \Psi_\beta \\ \frac{1}{\sqrt{T}} \Psi_\beta' \iota_n & \frac{1}{T} \Psi_\beta' \Psi_\beta \end{pmatrix}$ ,

$$\begin{aligned} & \left[ (\iota_n : \frac{1}{\sqrt{T}} \Psi_\beta)' (\iota_n : \frac{1}{\sqrt{T}} \Psi_\beta) \right]^{-1} = \\ & \begin{pmatrix} a^{-1} & -a^{-1} \iota_n' \Psi_\beta (\Psi_\beta' \Psi_\beta)^{-1} \\ -\sqrt{T} (\Psi_\beta' M_{i_n} \Psi_\beta)^{-1} \frac{\Psi_\beta' \iota_n}{n} & T (\Psi_\beta' M_{i_n} \Psi_\beta)^{-1} \end{pmatrix}. \end{aligned}$$

with  $a = \iota_n' M_{\Psi_\beta} \iota_n$ .

1a.  $E(\bar{R}) = \iota_n \lambda_1$  and

$$\begin{aligned} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} &= \left[ (\iota_n : \hat{\beta})' (\iota_n : \hat{\beta}) \right]^{-1} (\iota_n : \hat{\beta})' \bar{R} \\ &\underset{a}{=} \left[ (\iota_n : \frac{1}{\sqrt{T}} \Psi_\beta)' (\iota_n : \frac{1}{\sqrt{T}} \Psi_\beta) \right]^{-1} (\iota_n : \frac{1}{\sqrt{T}} \Psi_\beta)' (\iota_n \lambda_1 + \frac{1}{\sqrt{T}} \psi_R) \\ &\underset{a}{=} \begin{pmatrix} a^{-1} & -a^{-1} \iota_n' \Psi_\beta (\Psi_\beta' \Psi_\beta)^{-1} \\ -\sqrt{T} (\Psi_\beta' M_{i_n} \Psi_\beta)^{-1} \frac{\Psi_\beta' \iota_n}{n} & T (\Psi_\beta' M_{i_n} \Psi_\beta)^{-1} \end{pmatrix} \\ &\begin{pmatrix} \iota_n' \iota_n \lambda_1 + \frac{1}{\sqrt{T}} \iota_n' \psi_R \\ \frac{1}{\sqrt{T}} \Psi_\beta' \iota_n \lambda_1 + \frac{1}{T} \Psi_\beta' \psi_R \end{pmatrix} \\ &\underset{a}{=} \begin{pmatrix} a^{-1} (a \lambda_1 + \frac{1}{\sqrt{T}} b' \psi_R) \\ (\Psi_\beta' M_{i_n} \Psi_\beta)^{-1} \Psi_\beta' M_{i_n} \psi_R \end{pmatrix} \\ &\underset{a}{=} \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ (\Psi_\beta' M_{i_n} \Psi_\beta)^{-1} \Psi_\beta' M_{i_n} \end{pmatrix} \psi_R \end{aligned}$$



where  $\Psi_\beta = \text{vecinv}_k(\psi_\beta)$  and  $b = M_{\Psi_\beta} \iota_n$ , so, since  $\Psi_\beta$  and  $\psi_R$  are independent,

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \lambda_1) \\ \hat{\lambda}_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ (\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \Psi'_\beta M_{i_n} \end{pmatrix} \psi_R.$$

Since  $\psi_R$  is a normal distributed random vector that is independent of  $\Psi_\beta$ , the moments of limiting distribution of  $\hat{\lambda}_F$  are determined by the moments of  $(\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \Psi'_\beta M_{i_n}$ . When  $\Omega$  is an identity matrix,  $\Psi'_\beta M_{i_n} \Psi_\beta$  is a Wishart distributed random matrix with  $n - 1$  degrees of freedom and scale matrix  $Q_{FF}$ , see *e.g.* Muirhead (1982). Hence,  $(\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \Psi'_\beta M_{i_n}$  is the square root of the inverse of the Wishart distributed random matrix. To proof the order of finite moments of this matrix, we use an alternative definition of a multi-variate  $t$  distributed random variable.

A multi-variate  $t$  distributed random  $k \times 1$  vector  $x$  with  $\nu$  degrees,  $k \times k$  dimensional scale matrix  $\Sigma^{-1}$  and  $k \times 1$  scale vector  $\alpha$  can be defined by

$$x = \alpha + W^{-\frac{1}{2}} z,$$

where  $z \sim N(0, I)$  and independent of the  $k \times k$  Wishart distributed random matrix  $W$  with  $\nu + k - 1$  degrees of freedom and scale matrix  $\Sigma$ . Since the moments of the multi-variate  $t$  exist up to order  $\nu$ , the moments of the square root of the inverted-Wishart thus exist up to order  $\nu + k - 1$ . This definition of a multi-variate  $t$  distributed random matrix results from the definition of a matrix-variate  $t$  distributed random matrix, see *e.g.* Zellner (1971) and Drèze and Richard (1983).

In our case  $W = \Psi'_\beta M_{i_n} \Psi_\beta$  which is a  $k \times k$  Wishart distributed matrix with  $n - 1$  degrees of freedom. The moments of  $(\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \Psi'_\beta M_{i_n}$  exist therefore up to order  $n - 1 - k + 1 = n - k$  which holds as well for the moments of  $\hat{\lambda}_F$ . The difference of the covariance matrix of  $\Omega$  from the identity matrix does not change the order of finite moments.

**1b.** When  $\beta$  is equal to zero and  $E(\bar{R}) = c$ :

$$\begin{aligned} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} &= \begin{bmatrix} (\iota_n : \hat{\beta})' (\iota_n : \hat{\beta}) \end{bmatrix}^{-1} (\iota_n : \hat{\beta})' \bar{R} \\ &= \frac{1}{a} \begin{bmatrix} (\iota_n : \frac{1}{\sqrt{T}} \Psi_\beta)' (\iota_n : \frac{1}{\sqrt{T}} \Psi_\beta) \end{bmatrix}^{-1} (\iota_n : \frac{1}{\sqrt{T}} \Psi_\beta)' (c + \frac{1}{\sqrt{T}} \psi_R) \\ &= \begin{pmatrix} a^{-1} & -a^{-1} \iota_n' \Psi_\beta (\Psi'_\beta \Psi_\beta)^{-1} \\ -\sqrt{T} (\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \frac{\Psi'_\beta \iota_n}{n} & T (\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \end{pmatrix} \\ &\quad \begin{pmatrix} \iota_n' c + \frac{1}{\sqrt{T}} \iota_n' \psi_R \\ \frac{1}{\sqrt{T}} \Psi'_\beta c + \frac{1}{T} \Psi'_\beta \psi_R \end{pmatrix} \\ &= \begin{pmatrix} a^{-1} b' c + \frac{1}{\sqrt{T}} a^{-1} b' \psi_R \\ \sqrt{T} (\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \Psi_\beta^{-1} \Psi'_\beta M_{i_n} c + (\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \Psi_\beta^{-1} \Psi'_\beta M_{i_n} \psi_R \end{pmatrix}. \\ &= \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ \sqrt{T} (\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \Psi_\beta^{-1} \Psi'_\beta M_{i_n} c \end{pmatrix} c + \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ (\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \Psi_\beta^{-1} \Psi'_\beta M_{i_n} \end{pmatrix} \psi_R \end{aligned}$$

and

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \frac{1}{a} b' c) \\ \hat{\lambda}_F - \sqrt{T} (\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \Psi_\beta^{-1} \Psi'_\beta M_{i_n} c \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ (\Psi'_\beta M_{i_n} \Psi_\beta)^{-1} \Psi_\beta^{-1} \Psi'_\beta M_{i_n} \end{pmatrix} \psi_R$$

so  $\hat{\lambda}_F$  diverges when  $T$  gets large.

**2a.**  $\beta = \frac{1}{\sqrt{T}}B$ , with  $B$  a  $n \times k$  matrix of full rank and  $E(\bar{R}) = \iota_n \lambda_1 + \frac{1}{\sqrt{T}}B\lambda_F$  so

$$\begin{aligned}
\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} &= \frac{1}{a} \left[ (\iota_n : \hat{\beta})' (\iota_n : \hat{\beta}) \right]^{-1} (\iota_n : \hat{\beta})' \bar{R} \\
&= \frac{1}{a} \left[ (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta))' (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta)) \right]^{-1} (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta))' (\iota_n \lambda_1 + \frac{1}{\sqrt{T}}(B\lambda_F + \psi_R)) \\
&= \frac{1}{a} \begin{pmatrix} a^{-1} & -a^{-1}\sqrt{T}\iota_n'(B + \Psi_\beta)((B + \Psi_\beta)'(B + \Psi_\beta))^{-1} \\ -\sqrt{T}[(B + \Psi_\beta)'M_{\iota_n}(B + \Psi_\beta)]^{-1}\frac{(B + \Psi_\beta)'\iota_n}{n} & [(B + \Psi_\beta)'M_{\iota_n}(B + \Psi_\beta)]^{-1} \end{pmatrix} \\
&\quad \begin{pmatrix} \iota_n'\iota_n\lambda_1 + \frac{1}{\sqrt{T}}\iota_n'(B\lambda_F + \psi_R) \\ \frac{1}{\sqrt{T}}(B + \Psi_\beta)'\iota_n\lambda_1 + \frac{1}{T}(B + \Psi_\beta)'(B\lambda_F + \psi_R) \end{pmatrix} \\
&= \frac{1}{a} \begin{pmatrix} a^{-1}(a\lambda_1 + \frac{1}{\sqrt{T}}a^{-1}b'(B\lambda_F + \psi_R)) \\ ((B + \Psi_\beta)'M_{\iota_n}(B + \Psi_\beta))^{-1}(B + \Psi_\beta)'M_{\iota_n}(B\lambda_F + \psi_R) \end{pmatrix} \\
&= \frac{1}{a} \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{T}}\frac{1}{a}b' \\ ((B + \Psi_\beta)'M_{\iota_n}(B + \Psi_\beta))^{-1}(B + \Psi_\beta)'M_{\iota_n} \end{pmatrix} (B\lambda_F + \psi_R)
\end{aligned}$$

with  $a = \iota_n' M_{(B + \Psi_\beta)} \iota_n$ ,  $b = M_{(B + \Psi_\beta)} \iota_n$ , so,

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \lambda_1) \\ \hat{\lambda}_F - \lambda_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}}\frac{1}{a}b' \\ [(B + \Psi_\beta)'(B + \Psi_\beta)]^{-1}(B + \Psi_\beta)'(I_n - \frac{1}{a}\iota_n b') \end{pmatrix} (\psi_{R,\beta} - \Psi_\beta \lambda_F).$$

In this case  $(B + \Psi_\beta)'(B + \Psi_\beta)$  has a non-central Wishart distribution when  $\Omega = I_n$  which has no implications for the order of the finite moments compared to the case when  $B = 0$ . Hence, the moments of  $\hat{\lambda}_F$  exist up to order  $n - k$ .

**2b.**  $\beta = \frac{1}{\sqrt{T}}B$ , with  $B$  a  $n \times k$  matrix of full rank and  $E(\bar{R}) = c$ :

$$\begin{aligned}
\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} &= \frac{1}{a} \left[ (\iota_n : \hat{\beta})' (\iota_n : \hat{\beta}) \right]^{-1} (\iota_n : \hat{\beta})' \bar{R} \\
&= \frac{1}{a} \left[ (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta))' (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta)) \right]^{-1} (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta))' (c + \frac{1}{\sqrt{T}}\psi_R) \\
&= \frac{1}{a} \begin{pmatrix} a^{-1} & -a^{-1}\sqrt{T}\iota_n'(B + \Psi_\beta)((B + \Psi_\beta)'(B + \Psi_\beta))^{-1} \\ -\sqrt{T}[(B + \Psi_\beta)'M_{\iota_n}(B + \Psi_\beta)]^{-1}\frac{(B + \Psi_\beta)'\iota_n}{n} & [(B + \Psi_\beta)'M_{\iota_n}(B + \Psi_\beta)]^{-1} \end{pmatrix} \\
&\quad \begin{pmatrix} \iota_n'c + \frac{1}{\sqrt{T}}\iota_n'\psi_R \\ \frac{1}{\sqrt{T}}(B + \Psi_\beta)'c + \frac{1}{T}(B + \Psi_\beta)'\psi_R \end{pmatrix} \\
&= \frac{1}{a} \begin{pmatrix} a^{-1}b'c + \frac{1}{\sqrt{T}}a^{-1}b'\psi_R \\ [(B + \Psi_\beta)'M_{\iota_n}(B + \Psi_\beta)]^{-1}(B + \Psi_\beta)'M_{\iota_n}[\sqrt{T}c + \psi_R] \end{pmatrix} \\
&= \frac{1}{a} \begin{pmatrix} \frac{1}{a}b' \\ \sqrt{T}[(B + \Psi_\beta)'M_{\iota_n}(B + \Psi_\beta)]^{-1}(B + \Psi_\beta)'M_{\iota_n} \end{pmatrix} c + \\
&\quad \begin{pmatrix} \frac{1}{\sqrt{T}}\frac{1}{a}b' \\ [(B + \Psi_\beta)'M_{\iota_n}(B + \Psi_\beta)]^{-1}(B + \Psi_\beta)'M_{\iota_n} \end{pmatrix} \psi_R
\end{aligned}$$

with  $a = \iota_n' M_{(B + \Psi_\beta)} \iota_n$ ,  $b = M_{(B + \Psi_\beta)} \iota_n$ , and

$$\begin{aligned}
&\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \lambda_1) \\ \hat{\lambda}_F - \sqrt{T}[(B + \Psi_\beta)'(B + \Psi_\beta)]^{-1}(B + \Psi_\beta)'(I_n - \frac{1}{a}\iota_n b')c \end{pmatrix} \\
&\xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}}\frac{1}{a}b' \\ [(B + \Psi_\beta)'(B + \Psi_\beta)]^{-1}(B + \Psi_\beta)'(I_n - \frac{1}{a}\iota_n b') \end{pmatrix} \psi_R.
\end{aligned}$$

**3.**  $\beta = \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}$ , with  $\beta_1 : n_1 \times 1$  and  $E(\bar{R}) = \iota_n \lambda_1 + B \lambda_F$ . The specification of the FM risk premia estimator is

$$\begin{aligned} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} &= \frac{1}{a} \left[ (\iota_n \vdots \hat{\beta})' (\iota_n \vdots \hat{\beta}) \right]^{-1} (\iota_n \vdots \hat{\beta})' \bar{R} \\ &= \frac{1}{a} \left[ (\iota_n \vdots ((\beta_1)_0) + \frac{1}{\sqrt{T}} \Psi_\beta)' (\iota_n \vdots ((\beta_1)_0) + \frac{1}{\sqrt{T}} \Psi_\beta) \right]^{-1} (\iota_n \vdots ((\beta_1)_0) + \frac{1}{\sqrt{T}} \Psi_\beta)' (\iota_n \lambda_1 + (B \lambda_F + \psi_R)) + o_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

We construct the limit of the two different elements of  $\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix}$ . We assume that  $n_1$  remains fixed and that the following limits hold when  $n$  and  $T$  become large:  $\frac{\Psi'_\beta \Psi_\beta}{n} \xrightarrow{p} W$ ,  $W = [\lim_{n \rightarrow \infty} \text{tr}(\frac{1}{n} \Omega)] V_{FF}^{-1}$ ,  $\frac{n}{T} \rightarrow c$ ,  $\frac{n_1}{T} \rightarrow 0$  such that  $\frac{1}{\sqrt{T}} \beta'_1 \Psi_{\beta_1} \xrightarrow{p} 0$ ,  $\frac{1}{\sqrt{n}} \beta'_1 \iota_{n_1} \rightarrow 0$  such that  $\frac{n_1}{\sqrt{n}}$

$$\begin{aligned} &\begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} (\iota_n \vdots ((\beta_1)_0) + \frac{1}{\sqrt{T}} \Psi_\beta)' (\iota_n \vdots ((\beta_1)_0) + \frac{1}{\sqrt{T}} \Psi_\beta) \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} = \\ &\begin{pmatrix} 1 & \frac{1}{\sqrt{n}} \iota'_{n_1} \beta_1 + \frac{1}{\sqrt{T}} (\frac{1}{\sqrt{n}} \iota'_{n_1} \Psi_\beta) \\ \frac{1}{\sqrt{n}} \beta'_1 \iota_{n_1} + \frac{1}{\sqrt{T}} (\frac{1}{\sqrt{n}} \Psi'_\beta \iota_n) & \beta'_1 \beta_1 + \frac{1}{\sqrt{T}} \beta'_1 \Psi_{\beta_1} + \frac{1}{\sqrt{T}} \Psi'_{\beta_1} \beta_1 + \frac{n}{T} \frac{\Psi'_\beta \Psi_\beta}{n} \end{pmatrix} = \\ &\begin{pmatrix} 1 & b'_1 \\ b_1 & B_1 \end{pmatrix} \end{aligned}$$

with  $b_1 = \frac{1}{\sqrt{n}} \beta'_1 \iota_{n_1} + \frac{1}{\sqrt{T}} (\frac{1}{\sqrt{n}} \Psi'_\beta \iota_n)$ ,  $B_1 = \beta'_1 \beta_1 + \frac{1}{\sqrt{T}} \beta'_1 \Psi_{\beta_1} + \frac{1}{\sqrt{T}} \Psi'_{\beta_1} \beta_1 + \frac{n}{T} \frac{\Psi'_\beta \Psi_\beta}{n}$ ,  $\Psi_{\beta_1} = (I_{n_1} \vdots 0) \Psi_\beta$ , and its inverse reads

$$\begin{pmatrix} (1 - b'_1 B_1^{-1} b_1)^{-1} & -(1 - b'_1 B_1^{-1} b_1)^{-1} b'_1 B_1^{-1} \\ -(B_1 - b_1 b'_1)^{-1} b_1 & (B_1 - b_1 b'_1)^{-1} \end{pmatrix}.$$

The large sample behavior of the second part of the FM estimator reads

$$\begin{aligned} &\begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} (\iota_n \vdots ((\beta_1)_0) + \frac{1}{\sqrt{T}} \Psi_\beta)' (\iota_n \lambda_1 + ((\beta_1)_0) \lambda_F + \frac{1}{\sqrt{T}} \psi_R) = \\ &\begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \iota'_n \iota_n \lambda_1 + \iota'_{n_1} \beta_1 \lambda_F + \frac{1}{\sqrt{T}} \iota'_n \psi_R \\ \beta'_1 \iota_{n_1} \lambda_1 + \beta'_1 \beta_1 \lambda_F + \frac{1}{\sqrt{T}} \beta'_1 \psi_{R_1} + \frac{1}{\sqrt{T}} \Psi'_\beta \iota_n \lambda_1 + \frac{1}{\sqrt{T}} \Psi'_{\beta_1} \beta_1 \lambda_F + \frac{1}{T} \Psi'_\beta \Psi_R \end{pmatrix} = \\ &\begin{pmatrix} \sqrt{n} \lambda_1 + \frac{1}{\sqrt{n}} \beta'_1 \iota_{n_1} \lambda_F + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{n}} \iota'_n \psi_R \\ \beta'_1 \iota_{n_1} \lambda_1 + n_1 B_1 \lambda_F + \frac{1}{\sqrt{T}} \beta'_1 \psi_{R_1} + \frac{1}{\sqrt{T}} \Psi'_\beta \iota_n \lambda_1 + \frac{1}{\sqrt{T}} \Psi'_{\beta_1} \beta_1 \lambda_F + \frac{1}{T} \Psi'_\beta \Psi_R \end{pmatrix} = \\ &\sqrt{n} \begin{pmatrix} 1 & \frac{1}{\sqrt{n}} \beta'_1 \iota_{n_1} + \frac{1}{\sqrt{nT}} \Psi'_\beta \iota_n \\ \frac{1}{\sqrt{n}} \beta'_1 \iota_{n_1} + \frac{1}{\sqrt{nT}} \Psi'_\beta \iota_n & \beta'_1 \beta_1 + \frac{n}{T} \frac{\Psi'_\beta \Psi_\beta}{n} + \frac{1}{\sqrt{T}} \Psi'_{\beta_1} \beta_1 + \frac{1}{\sqrt{T}} \beta'_1 \Psi_{\beta_1} \end{pmatrix} \lambda_1 + \begin{pmatrix} \frac{1}{\sqrt{n}} \beta'_1 \iota_{n_1} + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{n}} \iota'_n \Psi_\beta \\ \beta'_1 \beta_1 + \frac{n}{T} \frac{\Psi'_\beta \Psi_\beta}{n} + \frac{1}{\sqrt{T}} \Psi'_{\beta_1} \beta_1 + \frac{1}{\sqrt{T}} \beta'_1 \Psi_{\beta_1} \end{pmatrix} \lambda_F + \\ &+ \begin{pmatrix} 0 \\ \frac{n}{T} \frac{\Psi'_\beta \Psi_\beta}{n} \end{pmatrix} (\mu_F - \lambda_F) - \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{\sqrt{n}} \iota'_n \Psi_\beta \\ \frac{n}{T} \frac{\Psi'_\beta \Psi_\beta}{n} + \frac{1}{\sqrt{T}} \beta'_1 \Psi_{\beta_1} \end{pmatrix} \lambda_F + \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{\sqrt{n}} \iota'_n \psi_{R_1} \\ \frac{1}{\sqrt{T}} \beta'_1 \psi_{R_1} + \frac{1}{T} \Psi'_\beta \Psi_R \end{pmatrix} = \\ &\sqrt{n} \begin{pmatrix} 1 \\ b_1 \end{pmatrix} \lambda_1 + \begin{pmatrix} b_1 \\ B_1 \end{pmatrix} \lambda_F + \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{\sqrt{n}} \iota'_n \Psi_\beta \\ \frac{n}{T} \frac{\Psi'_\beta \Psi_\beta}{n} + \frac{1}{\sqrt{T}} \beta'_1 \Psi_{\beta_1} \end{pmatrix} (\mu_F - \lambda_F) + \\ &\begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{\sqrt{n}} \iota'_n \psi_{R_1} \\ \frac{1}{\sqrt{T}} \beta'_1 \psi_{R_1} + \frac{1}{T} \Psi'_\beta \Psi_R \end{pmatrix}, \end{aligned}$$

where  $\Psi_{R_1} = (I_{n_1} \ : \ 0)\Psi_R$ . The combined behavior for the estimator then results as

$$\begin{aligned}
\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} &= \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} \left[ \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} [\iota_n \ : \ ((\beta_1) + \frac{1}{\sqrt{T}}\Psi_\beta)]' [\iota_n \ : \ ((\beta_1) + \frac{1}{\sqrt{T}}\Psi_\beta)] \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} \right]^{-1} \\
&\begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} (\iota_n \ : \ ((\beta_1) + \frac{1}{\sqrt{T}}\Psi_\beta))' (\iota_n \lambda_1 + ((\beta_1) \lambda_F + \frac{1}{\sqrt{T}}\psi_R)) \\
&= \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (1 - b'_1 B_1^{-1} b_1)^{-1} & -(1 - b'_1 B_1^{-1} b_1)^{-1} b'_1 B_1^{-1} \\ -(B_1 - b_1 b'_1)^{-1} b_1 & (B_1 - b_1 b'_1)^{-1} \end{pmatrix} \\
&\left[ \sqrt{n} \begin{pmatrix} 1 \\ b_1 \end{pmatrix} \lambda_1 + \begin{pmatrix} b'_1 \\ B_1 \end{pmatrix} \lambda_F + \begin{pmatrix} 0 \\ \frac{n}{T} \frac{\Psi'_\beta \Psi_\beta}{n} \end{pmatrix} (\mu_F - \lambda_F) - \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{\sqrt{n}} \iota'_n \Psi_\beta \\ \frac{1}{\sqrt{T}} \beta'_1 \Psi_{\beta_1} \end{pmatrix} \lambda_F \right. \\
&\left. \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{\sqrt{n}} \iota'_n \psi_{R_1}^* \\ \frac{1}{\sqrt{T}} \beta'_1 \psi_{R_1}^* + \frac{1}{T} \Psi'_\beta \Psi_{R^*} \end{pmatrix} \right] \\
&= \begin{pmatrix} \lambda_1 \\ \lambda_F \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (1 - b'_1 B_1^{-1} b_1)^{-1} & -(1 - b'_1 B_1^{-1} b_1)^{-1} b'_1 B_1^{-1} \\ -(B_1 - b_1 b'_1)^{-1} b_1 & (B_1 - b_1 b'_1)^{-1} \end{pmatrix} \left[ \begin{pmatrix} 0 \\ \frac{n}{T} \frac{\Psi'_\beta \Psi_\beta}{n} \end{pmatrix} (\mu_F - \lambda_F) + \right. \\
&\left. - \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{\sqrt{n}} \iota'_n \Psi_\beta \\ \frac{1}{\sqrt{T}} \beta'_1 \Psi_{\beta_1} \end{pmatrix} \lambda_F \right] + \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (1 - b'_1 B_1^{-1} b_1)^{-1} & -b'_1 (B_1 - b_1 b'_1)^{-1} \\ -(B_1 - b_1 b'_1)^{-1} b_1 & (B_1 - b_1 b'_1)^{-1} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{\sqrt{n}} \iota'_n \psi_R^* \\ \frac{1}{\sqrt{T}} \beta'_1 \psi_{R_1}^* + \frac{1}{T} \Psi'_\beta \Psi_{R^*} \end{pmatrix}.
\end{aligned}$$

When we now use that  $\frac{n}{T} \frac{\Psi'_\beta \Psi_\beta}{n} \xrightarrow{p} W$ ,  $\frac{1}{\sqrt{T}} \beta'_1 \Psi_{\beta_1} \xrightarrow{p} 0$ ,  $b_1 \xrightarrow{p} 0$ ,  $B_1 \xrightarrow{p} \beta'_1 \beta_1 + W$ , we obtain that the bias of the FM risk premia estimator equals  $(\beta'_1 \beta_1 + W)^{-1} W (\mu_F - \lambda_F)$ .

#### D. Proof of Theorem 2.

1.  $\beta = 0$ . We use that since  $(\iota_n \ : \ \frac{1}{\sqrt{T}}\Psi_\beta)' \Omega^{-1} (\iota_n \ : \ \frac{1}{\sqrt{T}}\Psi_\beta) = \begin{pmatrix} \iota'_n \Omega^{-1} \iota_n & \frac{1}{\sqrt{T}} \iota'_n \Omega^{-1} \Psi_\beta \\ \frac{1}{\sqrt{T}} \Psi'_\beta \Omega^{-1} \iota_n & \frac{1}{T} \Psi'_\beta \Omega^{-1} \Psi_\beta \end{pmatrix}$ ,

$$\begin{aligned}
&\left[ (\iota_n \ : \ \frac{1}{\sqrt{T}}\Psi_\beta)' \Omega^{-1} (\iota_n \ : \ \frac{1}{\sqrt{T}}\Psi_\beta) \right]^{-1} = \\
&\begin{pmatrix} a^{-1} & -\sqrt{T} a^{-1} \iota'_n \Omega^{-1} \Psi_\beta (\Psi'_\beta \Omega^{-1} \Psi_\beta)^{-1} \\ -\sqrt{T} (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \frac{\Psi'_\beta \Omega^{-1} \iota_n}{\iota'_n \Omega^{-1} \iota_n} & T (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \end{pmatrix}.
\end{aligned}$$

with  $a = \iota'_n \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} \iota_n$ .

1a.  $E(\bar{R}) = \iota_n \lambda_1$  and since  $\hat{\Omega} \xrightarrow{p} \Omega$ ,

$$\begin{aligned}
\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} &= \left[ (\iota_n \ : \ \hat{\beta})' \hat{\Omega}^{-1} (\iota_n \ : \ \hat{\beta}) \right]^{-1} (\iota_n \ : \ \hat{\beta})' \hat{\Omega} \bar{R} \\
&\stackrel{a}{=} \left[ (\iota_n \ : \ \frac{1}{\sqrt{T}}\Psi_\beta)' \Omega^{-1} (\iota_n \ : \ \frac{1}{\sqrt{T}}\Psi_\beta) \right]^{-1} (\iota_n \ : \ \frac{1}{\sqrt{T}}\Psi_\beta)' \Omega^{-1} (\iota_n \lambda_1 + \frac{1}{\sqrt{T}}\psi_R) \\
&\stackrel{a}{=} \begin{pmatrix} a^{-1} & -\sqrt{T} a^{-1} \iota'_n \Omega^{-1} \Psi_\beta (\Psi'_\beta \Omega^{-1} \Psi_\beta)^{-1} \\ -\sqrt{T} (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \frac{\Psi'_\beta \Omega^{-1} \iota_n}{\iota'_n \Omega^{-1} \iota_n} & T (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \end{pmatrix} \\
&\begin{pmatrix} \iota'_n \Omega^{-1} \iota_n \lambda_1 + \frac{1}{\sqrt{T}} \iota'_n \Omega^{-1} \psi_R \\ \frac{1}{\sqrt{T}} \Psi'_\beta \Omega^{-1} \iota_n \lambda_1 + \frac{1}{T} \Psi'_\beta \Omega^{-1} \psi_R \end{pmatrix} \\
&\stackrel{a}{=} \begin{pmatrix} a^{-1} (a \lambda_1 + \frac{1}{\sqrt{T}} b' \Omega^{-\frac{1}{2}} \psi_R) \\ (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \psi_R \end{pmatrix} \\
&\stackrel{a}{=} \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \end{pmatrix} \Omega^{-\frac{1}{2}} \psi_R
\end{aligned}$$

where  $\Psi_\beta = \text{vecinv}_k(\psi_\beta)$  and  $b = M_{\Omega^{-\frac{1}{2}}\Psi_\beta} \Omega^{-\frac{1}{2}}\iota_n$ , so

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \lambda_1) \\ \hat{\lambda}_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}}\frac{1}{a}b' \\ (\Psi'_\beta\Omega^{-\frac{1}{2}}M_{\Omega^{-\frac{1}{2}}\Psi_\beta} \Omega^{-\frac{1}{2}}\Psi_\beta)^{-1}\Psi'_\beta\Omega^{-\frac{1}{2}}M_{\Omega^{-\frac{1}{2}}\Psi_\beta} \end{pmatrix} \Omega^{-\frac{1}{2}}\psi_R.$$

Since  $\psi_R$  is a normal distributed random vector that is independent of  $\Psi_\beta$ . The random matrix  $(\Psi'_\beta\Omega^{-\frac{1}{2}}M_{\Omega^{-\frac{1}{2}}\Psi_\beta} \Omega^{-\frac{1}{2}}\Psi_\beta)^{-1}\Psi'_\beta\Omega^{-\frac{1}{2}}M_{\Omega^{-\frac{1}{2}}\Psi_\beta}$  is the square root of the inverse of the Wishart distributed random matrix  $(\Omega^{-\frac{1}{2}}\Psi_\beta)'M_{\Omega^{-\frac{1}{2}}\Psi_\beta} \Omega^{-\frac{1}{2}}\Psi_\beta$  which has a Wishart distributed matrix with  $n - 1$  degrees of freedom and scale matrix  $Q_{FF}$ . Hence, it results from the proof of Theorem 1 that the limiting distribution of  $\hat{\lambda}_F$  is a multi-variate  $t$  distribution with  $n - k$  degrees of freedom, location 0 and scale matrix  $Q_{11.F}Q_{FF}^{-1}$  with  $Q_{11.F} = Q_{11} - Q_{1F}Q_{FF}^{-1}Q_{F1}$ .

**1b.** When  $\beta$  is equal to zero and  $E(\bar{R}) = c$ :

$$\begin{aligned} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} &= \left[ (\iota_n : \hat{\beta})' \hat{\Omega}^{-1} (\iota_n : \hat{\beta}) \right]^{-1} (\iota_n : \hat{\beta})' \hat{\Omega} \bar{R} \\ &= \frac{1}{a} \left[ (\iota_n : \frac{1}{\sqrt{T}}\Psi_\beta)' \Omega^{-1} (\iota_n : \frac{1}{\sqrt{T}}\Psi_\beta) \right]^{-1} (\iota_n : \frac{1}{\sqrt{T}}\Psi_\beta)' \Omega^{-1} (\iota_n \lambda_1 + \frac{1}{\sqrt{T}}\psi_R) \\ &= \frac{1}{a} \begin{pmatrix} a^{-1} & -\sqrt{T}a^{-1}\iota_n' \Omega^{-1} \Psi_\beta (\Psi'_\beta \Omega^{-1} \Psi_\beta)^{-1} \\ -\sqrt{T}(\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \frac{\Psi'_\beta \Omega^{-1} \iota_n}{\iota_n' \Omega^{-1} \iota_n} & T(\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \\ \left( \begin{array}{c} \iota_n' \Omega^{-1} c + \frac{1}{\sqrt{T}} \iota_n' \Omega^{-1} \psi_R \\ \frac{1}{\sqrt{T}} \Psi'_\beta \Omega^{-1} c + \frac{1}{T} \Psi'_\beta \Omega^{-1} \psi_R \end{array} \right) \end{pmatrix} \\ &= \frac{1}{a} \begin{pmatrix} a^{-1} \iota_n' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} c + \frac{1}{\sqrt{T}} b' \Omega^{-\frac{1}{2}} \psi_R \\ (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} (\sqrt{T}c + \psi_R) \end{pmatrix} \\ &= \frac{1}{a} \begin{pmatrix} a^{-1} b' \Omega^{-\frac{1}{2}} c \\ \sqrt{T} (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} c \end{pmatrix} + \\ &\quad \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ (\Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \Omega^{-\frac{1}{2}} \Psi_\beta)^{-1} \Psi'_\beta \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \Psi_\beta} \end{pmatrix} \Omega^{-\frac{1}{2}} \psi_R \end{aligned}$$

and

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - a^{-1}b'\Omega^{-\frac{1}{2}}c) \\ \hat{\lambda}_F - \sqrt{T}(\Psi'_\beta\Omega^{-\frac{1}{2}}M_{\Omega^{-\frac{1}{2}}\Psi_\beta} \Omega^{-\frac{1}{2}}\Psi_\beta)^{-1}\Psi'_\beta\Omega^{-\frac{1}{2}}M_{\Omega^{-\frac{1}{2}}\Psi_\beta} \Omega^{-\frac{1}{2}}c \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}}\frac{1}{a}b' \\ (\Psi'_\beta\Omega^{-\frac{1}{2}}M_{\Omega^{-\frac{1}{2}}\Psi_\beta} \Omega^{-\frac{1}{2}}\Psi_\beta)^{-1}\Psi'_\beta\Omega^{-\frac{1}{2}}M_{\Omega^{-\frac{1}{2}}\Psi_\beta} \end{pmatrix} \psi_R$$

and  $\hat{\lambda}_F$  diverges when  $T$  gets large.

**2a.**  $\beta = \frac{1}{\sqrt{T}}B$ , with  $B$  a  $n \times k$  matrix of full rank and  $E(\bar{R}) = \iota_n \lambda_1 + \frac{1}{\sqrt{T}}B$  so

$$\begin{aligned}
\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} &= \left[ (\iota_n : \hat{\beta})' \hat{\Omega}^{-1} (\iota_n : \hat{\beta}) \right]^{-1} (\iota_n : \hat{\beta})' \hat{\Omega}^{-1} \bar{R} \\
&= \frac{1}{a} \left[ (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta))' \Omega^{-1} (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta)) \right]^{-1} (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta))' \Omega^{-1} (\iota_n \lambda_1 + \frac{1}{\sqrt{T}}(B \lambda_F + \psi_R)) \\
&= \frac{1}{a} \begin{pmatrix} a^{-1} \\ -\sqrt{T}((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} \frac{(B + \Psi_\beta)' \Omega^{-1} \iota_n}{\iota_n' \Omega^{-1} \iota_n} \\ -\sqrt{T} a^{-1} \iota_n' \Omega^{-1} (B + \Psi_\beta) ((B + \Psi_\beta)' \Omega^{-1} (B + \Psi_\beta))^{-1} \\ T((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} \end{pmatrix} \\
&\quad \begin{pmatrix} \iota_n' \Omega^{-1} \iota_n \lambda_1 + \frac{1}{\sqrt{T}} \iota_n' \Omega^{-1} (B \lambda_F + \psi_R) \\ \frac{1}{\sqrt{T}} (B + \Psi_\beta)' \Omega^{-1} \iota_n \lambda_1 + \frac{1}{T} (B + \Psi_\beta)' \Omega^{-1} (B \lambda_F + \psi_R) \end{pmatrix} \\
&= \frac{1}{a} \begin{pmatrix} a^{-1} (a \lambda_1 + \frac{1}{\sqrt{T}} a^{-1} b' \Omega^{-\frac{1}{2}} (B \lambda_F + \psi_R)) \\ ((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} (B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B \lambda_F + \psi_R) \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ ((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} (B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \end{pmatrix} \Omega^{-\frac{1}{2}} (B \lambda_F + \psi_R).
\end{aligned}$$

Hence,

$$\begin{pmatrix} \sqrt{T}(\hat{\lambda}_1 - \lambda_1) \\ \hat{\lambda}_F - \lambda_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ ((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} (B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \end{pmatrix} \Omega^{-\frac{1}{2}} (\psi_R - \Psi_\beta \lambda_F).$$

In this case  $(B + \Psi_\beta)'(B + \Psi_\beta)$  has a non-central Wishart distribution regardless of the value of  $\Omega$ . Hence, the moments of  $\hat{\lambda}_F$  exist up to order  $n - k$ .

**2b.**  $\beta = \frac{1}{\sqrt{T}}B$ , with  $B$  a  $n \times k$  matrix of full rank and  $E(\bar{R}) = c$ :

$$\begin{aligned}
\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_F \end{pmatrix} &= \left[ (\iota_n : \hat{\beta})' \hat{\Omega}^{-1} (\iota_n : \hat{\beta}) \right]^{-1} (\iota_n : \hat{\beta})' \bar{R} \\
&= \frac{1}{a} \left[ (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta))' \Omega^{-1} (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta)) \right]^{-1} (\iota_n : \frac{1}{\sqrt{T}}(B + \Psi_\beta))' \Omega^{-1} (c + \frac{1}{\sqrt{T}}\psi_R) \\
&= \frac{1}{a} \begin{pmatrix} a^{-1} \\ -\sqrt{T}((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} \frac{(B + \Psi_\beta)' \Omega^{-1} \iota_n}{\iota_n' \Omega^{-1} \iota_n} \\ -\sqrt{T} a^{-1} \iota_n' \Omega^{-1} (B + \Psi_\beta) ((B + \Psi_\beta)' \Omega^{-1} (B + \Psi_\beta))^{-1} \\ T((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} \end{pmatrix} \\
&\quad \begin{pmatrix} \iota_n' \Omega^{-1} c + \frac{1}{\sqrt{T}} \iota_n' \Omega^{-1} \psi_R \\ \frac{1}{\sqrt{T}} (B + \Psi_\beta)' \Omega^{-1} c + \frac{1}{T} (B + \Psi_\beta)' \Omega^{-1} \psi_R \end{pmatrix} \\
&= \frac{1}{a} \begin{pmatrix} a^{-1} (b' \Omega^{-\frac{1}{2}} c + \frac{1}{\sqrt{T}} b' \Omega^{-\frac{1}{2}} \psi_R) \\ ((B + \Psi_\beta)' M_{\Omega^{-\frac{1}{2}} \iota_n} (B + \Psi_\beta))^{-1} (B + \Psi_\beta)' M_{\Omega^{-\frac{1}{2}} \iota_n} (\sqrt{T} c + \psi_R) \end{pmatrix} \\
&= \frac{1}{a} \begin{pmatrix} \frac{1}{a} b' \Omega^{-\frac{1}{2}} c \\ \sqrt{T}((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} (B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} c \end{pmatrix} \\
&\quad + \begin{pmatrix} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ ((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} (B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \end{pmatrix} \Omega^{-\frac{1}{2}} \psi_R,
\end{aligned}$$

with  $a = \iota'_n \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}}(B+\Psi_\beta)} \Omega^{-\frac{1}{2}} \iota_n$ ,  $b = M_{\Omega^{-\frac{1}{2}}(B+\Psi_\beta)} \Omega^{-\frac{1}{2}} \iota_n$ , and

$$\begin{aligned} & \left( \begin{array}{c} \sqrt{T}(\hat{\lambda}_1 - \frac{1}{a} b' \Omega^{-\frac{1}{2}} c) \\ \hat{\lambda}_F - \sqrt{T}((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} (B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} c \end{array} \right) \\ & \xrightarrow{d} \left( \begin{array}{c} \frac{1}{\sqrt{T}} \frac{1}{a} b' \\ ((B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \Omega^{-\frac{1}{2}} (B + \Psi_\beta))^{-1} (B + \Psi_\beta)' \Omega^{-\frac{1}{2}} M_{\Omega^{-\frac{1}{2}} \iota_n} \end{array} \right) \Omega^{-\frac{1}{2}} \psi_R. \end{aligned}$$

**E. Proof of Lemma 2.** Assumption 1 states that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left( \begin{array}{c} 1 \\ F_t \end{array} \right) \otimes (R_t - \iota_n \lambda_1 - \beta(\bar{F}_t + \lambda_F)) \xrightarrow{d} \begin{pmatrix} \varphi_R \\ \varphi_\beta \\ \varphi_F \end{pmatrix},$$

with  $(\varphi'_R \ \varphi'_\beta \ \varphi'_F)' \sim N(0, V)$ . We take all asset returns in deviation from the  $n$ -th asset return so

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left( \begin{array}{c} 1 \\ F_t \end{array} \right) \otimes (\mathcal{R}_t - \mathcal{B}(\bar{F}_t + \lambda_F)) \xrightarrow{d} \begin{pmatrix} \varphi_{\mathcal{R}} \\ \varphi_{\mathcal{B}} \\ \psi_F \end{pmatrix},$$

with  $\varphi_{\mathcal{R}} = J_n \varphi_R$ ,  $\varphi_{\mathcal{B}} = (I_k \otimes J_n) \varphi_\beta$  with  $J_n = (I_{N-1} \vdots -\iota_{n-1})$  so  $(\varphi'_{\mathcal{R}} \vdots \varphi'_{\mathcal{B}} \vdots \varphi'_F)' \sim N(0, \text{diag}(Q \otimes \Sigma, V_{FF}))$  with  $\Sigma = J_n \Omega J'_n$ . We pre-multiply the result from Assumption 1 with

$$\left( \begin{array}{cc} 1 & 0 \\ \left[ \frac{1}{T} \sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})' \right]^{-1} (\lambda_{F,0} - \bar{F}) & \left[ \frac{1}{T} \sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})' \right]^{-1} \otimes I_n & 0 \\ 0 & & I_k \end{array} \right)$$

to obtain under  $H_0 : \lambda_F = \lambda_{F,0}$  that

$$\sqrt{T} \begin{pmatrix} \bar{\mathcal{R}} - \mathcal{B} \lambda_F \\ \text{vec}(\tilde{\mathcal{B}} - \mathcal{B}) \\ \bar{F} - \mu_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_{\mathcal{R}} \\ \xi_{\mathcal{B}} \\ \xi_F \end{pmatrix},$$

where  $\tilde{\mathcal{B}} = \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_{F,0})' (\sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})')^{-1}$  and  $(\xi'_{\mathcal{R}} \vdots \xi'_{\mathcal{B}} \vdots \xi'_F) \sim N(0, \text{diag}(Q(\lambda_F) \otimes \Sigma, V_{FF}))$  and

$$\begin{aligned} \mathcal{Q}(\lambda_F) &= \lim_{T \rightarrow \infty} E \left[ \begin{pmatrix} 1 & \lambda'_F \left[ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F) \right]^{-1} \\ \left[ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F) \right]^{-1} \lambda_F & \left[ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F) \right]^{-1} \end{pmatrix} \right] \\ &= \begin{pmatrix} \mathcal{Q}(\lambda_F)_{11} & \mathcal{Q}(\lambda_F)_{1F} \\ \mathcal{Q}(\lambda_F)_{F1} & \mathcal{Q}(\lambda_F)_{FF} \end{pmatrix}. \end{aligned}$$

Hence, under  $H_0 : \lambda_F = \lambda_{F,0}$ ,  $\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0} = \bar{\mathcal{R}} - \mathcal{B} \lambda_F - (\tilde{\mathcal{B}} - \mathcal{B}) \lambda_F$  so  $\sqrt{T}(\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}) \xrightarrow{d} \xi_{\mathcal{R}} - (\lambda'_F \otimes I_{n-1}) \xi_{\mathcal{B}}$ , which is independent of  $\xi_{\mathcal{B}}$  since  $\begin{pmatrix} 1 - \lambda'_{F,0} \\ 0 \quad I_k \end{pmatrix} \mathcal{Q}(\lambda_F) \begin{pmatrix} 1 - \lambda'_{F,0} \\ 0 \quad I_k \end{pmatrix}' = \text{diag}(1 - \lambda'_{F,0} \mathcal{Q}(\lambda_F)_{FF} \lambda_{F,0}, \mathcal{Q}(\lambda_F)_{FF})$ , so

$$\sqrt{T} \begin{pmatrix} \bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_F \\ \text{vec}(\tilde{\mathcal{B}} - \mathcal{B}) \\ \bar{F} - \mu_F \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_{\mathcal{R}} \\ \psi_{\mathcal{B}} \\ \psi_F \end{pmatrix},$$

where  $\psi_{\mathcal{R}}$ ,  $\psi_{\mathcal{B}}$  and  $\psi_F$  are independent  $(n-1) \times 1$ ,  $(n-1)k \times 1$  and  $k \times 1$  normal distributed random vectors with mean zero and covariance matrices,  $(1 - \lambda'_{F,0} \mathcal{Q}(\lambda_F)_{FF} \lambda_{F,0}) \otimes \Sigma$ ,  $\mathcal{Q}(\lambda_F)_{FF} \otimes \Sigma$  and  $V_{FF}$ .

**F. The derivative of the FAR statistic with respect to  $\lambda_{F,0}$ .** To obtain the derivative of

$$\text{FAR}(\lambda_{F,0}) = \frac{T}{1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0}} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}),$$

we first construct the derivatives of its different elements with respect to  $\lambda_{F,0}$ .

$$\begin{aligned} \frac{\partial \text{vec}(\tilde{\mathcal{B}})}{\partial \lambda'_{F,0}} &= \frac{\partial}{\partial \lambda'_{F,0}} \text{vec} \left[ \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_{F,0})' (\sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})')^{-1} \right] \\ &= \left[ (\sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})')^{-1} \otimes I_{n-1} \right] \frac{\partial}{\partial \lambda'_{F,0}} \text{vec} \left[ \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_{F,0})' \right] \\ &\quad + \left[ I_k \otimes \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_{F,0})' \right] \frac{\partial}{\partial \lambda'_{F,0}} \text{vec} \left[ (\sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})')^{-1} \right] \\ &= \left[ (\sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})')^{-1} \otimes T \bar{\mathcal{R}} \right] \\ &\quad + \left[ I_k \otimes \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_{F,0})' \right] \frac{\partial}{\partial \lambda'_{F,0}} \text{vec} \left[ (\sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})')^{-1} \right] \\ &= \left[ (\sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})')^{-1} \otimes T \bar{\mathcal{R}} \right] \\ &\quad - 2 \left[ (\sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})')^{-1} \otimes T \tilde{\mathcal{B}} \lambda_{F,0} \right] \\ &= T \left[ (\sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})')^{-1} \otimes (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}) \right] \\ \frac{\partial}{\partial \lambda'_{F,0}} \left[ 1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0} \right] &= -2 \lambda'_{F,0} \left( \frac{1}{T} \sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})' \right)^{-1} - \\ &\quad (\lambda'_{F,0} \otimes \lambda'_{F,0}) \frac{\partial}{\partial \lambda'_{F,0}} \text{vec} \left[ \left( \frac{1}{T} \sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})' \right)^{-1} \right] \\ &= -2 \lambda'_{F,0} \left( \frac{1}{T} \sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})' \right)^{-1}_{FF} \\ &\quad (1 - \lambda'_{F,0} \left( \frac{1}{T} \sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})' \right)^{-1} \lambda_{F,0}) \\ \frac{\partial \text{vec}(\tilde{\Sigma})}{\partial \lambda'_{F,0}} &= -2 \left( \frac{1}{T-k-1} \sum_{t=1}^T (\bar{\mathcal{R}}_t - \tilde{\mathcal{B}} (\bar{F}_t + \lambda_{F,0})) \otimes \tilde{\mathcal{B}} \right) \\ &\quad - 2 \left( \frac{1}{T-k-1} \sum_{t=1}^T (\bar{\mathcal{R}}_t - \tilde{\mathcal{B}} (\bar{F}_t + \lambda_{F,0})) (\bar{F}_t + \lambda_{F,0})' \otimes I_{n-1} \right) \\ &= -2 \left( \frac{1}{T-k-1} \sum_{t=1}^T (\bar{\mathcal{R}}_t - \tilde{\mathcal{B}} (\bar{F}_t + \lambda_{F,0})) \otimes \tilde{\mathcal{B}} \right) \\ &= -2 \left( \frac{T}{T-k-1} ((\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}) \otimes \tilde{\mathcal{B}}) \right) \end{aligned}$$

since  $\sum_{t=1}^T (\bar{\mathcal{R}}_t - \tilde{\mathcal{B}} (\bar{F}_t + \lambda_{F,0})) (\bar{F}_t + \lambda_{F,0})' = 0$  and  $\bar{F} = 0$ .

Hence,

$$\begin{aligned} \frac{\partial}{\partial \lambda'_{F,0}} \text{FAR}(\lambda_{F,0}) &= -2 \left\{ \frac{T}{1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0}} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} \tilde{\mathcal{B}} + \right. \\ &\quad \frac{T}{1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0}} (\lambda'_{F,0} \otimes (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1}) \frac{\partial \text{vec}(\tilde{\mathcal{B}})}{\partial \lambda'_{F,0}} \\ &\quad \frac{T}{(1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0})^2} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}) \\ &\quad \frac{\partial}{\partial \lambda'_{F,0}} \left[ 1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0} \right] + \frac{T}{1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0}} \\ &\quad \left. ((\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} \otimes (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1}) \frac{\partial \text{vec}(\tilde{\Sigma})}{\partial \lambda'_{F,0}} \right\} \\ &= -2 \frac{T}{1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0}} \left\{ (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} \tilde{\mathcal{B}} + \right. \\ &\quad \left[ \lambda'_{F,0} \left( \frac{1}{T} \sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})' \right)^{-1} \otimes (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}) \right] \\ &\quad - \left[ \lambda'_{F,0} \left( \frac{1}{T} \sum_{j=1}^T (\bar{F}_j + \lambda_{F,0}) (\bar{F}_j + \lambda_{F,0})' \right)^{-1}_{FF} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}) \right] \\ &\quad \left. \left( \frac{T}{T-k-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}) \otimes (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} \tilde{\mathcal{B}} \right) \right\} \\ &= -2 \frac{T}{1 - \lambda'_{F,0} \hat{\mathcal{Q}}(\lambda_F)_{FF}^{-1} \lambda_{F,0}} \left[ 1 + \frac{1}{T-k-1} \text{FAR}(\lambda_{F,0}) \right] (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \tilde{\Sigma}^{-1} \tilde{\mathcal{B}} \end{aligned}$$



**G. The specification of FAR( $\lambda_{F,0}$ ) and CLR( $\lambda_{F,0}$ ).** After concentrating with respect to  $B = (c : \mathcal{B})$ , the log-likelihood of the unrestricted model

$$\mathcal{R}_t = B\left(\frac{1}{\bar{F}_t}\right) + \varepsilon_t,$$

under independent normal distributed disturbances with mean zero and covariance matrix  $\Sigma$ , is proportional to

$$\ln \mathcal{L}_{\text{unres}}(\Sigma) = c_\pi - \frac{T}{2} \ln |\Sigma| - \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T \left( \mathcal{R}_t - \hat{B}\left(\frac{1}{\bar{F}_t}\right) \right) \left( \mathcal{R}_t - \hat{B}\left(\frac{1}{\bar{F}_t}\right) \right)' \right],$$

with  $\hat{B} = \sum_{t=1}^T \mathcal{R}_t \left(\frac{1}{\bar{F}_t}\right)' \left[ \sum_{j=1}^T \left(\frac{1}{\bar{F}_j}\right) \left(\frac{1}{\bar{F}_j}\right)' \right]^{-1}$  and  $c_\pi = -\frac{T}{2} \ln(2\pi)$ . To determine the difference with the log-likelihood of the restricted model, we specify  $\hat{B}\left(\frac{1}{\bar{F}_t}\right)$  as

$$\begin{aligned} \hat{B}\left(\frac{1}{\bar{F}_t}\right) &= \sum_{i=1}^T \mathcal{R}_i \left(\frac{1}{\bar{F}_i}\right)' \left[ \sum_{j=1}^T \left(\frac{1}{\bar{F}_j}\right) \left(\frac{1}{\bar{F}_j}\right)' \right]^{-1} \left(\frac{1}{\bar{F}_t}\right) \\ &= \sum_{i=1}^T \mathcal{R}_i \left(\frac{1}{\bar{F}_i}\right)' \begin{pmatrix} 1 & 0 \\ \lambda_{F,0} & I_k \end{pmatrix}' \left[ \begin{pmatrix} 1 & 0 \\ \lambda_{F,0} & I_k \end{pmatrix} \sum_{j=1}^T \left(\frac{1}{\bar{F}_j}\right) \left(\frac{1}{\bar{F}_j}\right)' \begin{pmatrix} 1 & 0 \\ \lambda_{F,0} & I_k \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 & 0 \\ \lambda_{F,0} & I_k \end{pmatrix} \left(\frac{1}{\bar{F}_t}\right) \\ &= \sum_{i=1}^T \mathcal{R}_i \begin{pmatrix} 1 \\ \bar{F}_i + \lambda_{F,0} \end{pmatrix}' \left[ \sum_{j=1}^T \begin{pmatrix} 1 \\ \bar{F}_j + \lambda_{F,0} \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_j + \lambda_{F,0} \end{pmatrix}' \right]^{-1} \begin{pmatrix} 1 \\ \bar{F}_t + \lambda_{F,0} \end{pmatrix} \\ &= \sum_{i=1}^T \mathcal{R}_i \begin{pmatrix} 1 \\ \bar{F}_i + \lambda_{F,0} \end{pmatrix}' \left[ \sum_{j=1}^T \begin{pmatrix} 1 \\ \bar{F}_j + \lambda_{F,0} \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_j + \lambda_{F,0} \end{pmatrix}' \right]^{-1} \begin{pmatrix} 1 \\ \bar{F}_t + \lambda_{F,0} \end{pmatrix} \\ &= \sum_{i=1}^T \mathcal{R}_i \begin{pmatrix} 1 \\ \bar{F}_i + \lambda_{F,0} \end{pmatrix}' \left[ \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{T} \hat{Q}_{FF}(\lambda_{F,0})^{-1} \end{pmatrix} + \begin{pmatrix} 1 \\ -\hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0} \end{pmatrix} \right. \\ &\quad \left. \frac{1}{T} (1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0})^{-1} \begin{pmatrix} 1 \\ -\hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0} \end{pmatrix}' \right] \begin{pmatrix} 1 \\ \bar{F}_t + \lambda_{F,0} \end{pmatrix} \\ &= \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) + (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0})(1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0})^{-1} \\ &\quad (1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} (\bar{F}_t + \lambda_{F,0})), \end{aligned}$$

with  $\hat{Q}_{FF}(\lambda_{F,0}) = \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_{F,0})(\bar{F}_t + \lambda_{F,0})'$ ,  $\tilde{\mathcal{B}} = \frac{1}{T} \sum_{i=1}^T \mathcal{R}_i \begin{pmatrix} 1 \\ \bar{F}_i + \lambda_{F,0} \end{pmatrix}' \hat{Q}_{FF}(\lambda_{F,0})^{-1}$ . Hence,

$$\begin{aligned} \ln \mathcal{L}_{\text{unres}}(\Sigma) + \frac{T}{2} \ln |\Sigma| - c_\pi &= -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T \left( \mathcal{R}_t - \hat{B}\left(\frac{1}{\bar{F}_t}\right) \right) \left( \mathcal{R}_t - \hat{B}\left(\frac{1}{\bar{F}_t}\right) \right)' \right] \\ &= -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right) \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right)' \right] - \frac{1}{2} \frac{1}{1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0}} \times \\ &\quad \left\{ -\text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0})(1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} (\bar{F}_t + \lambda_{F,0})) \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right)' \right] \right. \\ &\quad \left. -\text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right) (1 - (\bar{F}_t + \lambda_{F,0})' \hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0}) (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0}) \right] \right. \\ &\quad \left. + \frac{1}{1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0}} \times \text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0})(1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} (\bar{F}_t + \lambda_{F,0})) \right. \right. \\ &\quad \left. \left. (1 - (\bar{F}_t + \lambda_{F,0})' \hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0}) (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0}) \right] \right\} \\ &= -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right) \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right)' \right] - \frac{1}{2} \frac{1}{1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0}} \times \\ &\quad \left\{ -\text{tr} \left[ \Sigma^{-1} (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0}) T (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0} - \lambda'_{F,0} \tilde{\mathcal{B}}' + \lambda'_{F,0} \tilde{\mathcal{B}}') \right] \right. \\ &\quad \left. -\text{tr} \left[ \Sigma^{-1} (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0} - \tilde{\mathcal{B}}\lambda_{F,0} + \tilde{\mathcal{B}}\lambda_{F,0}) T (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0})' \right] + \frac{1}{1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0}} \times \right. \\ &\quad \left. +\text{tr} \left[ \Sigma^{-1} (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0}) T (1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0}) (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0})' \right] \right\} \\ &= -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right) \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right)' \right] + \frac{1}{2} \frac{T}{1 - \lambda'_{F,0} \hat{Q}_{FF}(\lambda_{F,0})^{-1} \lambda_{F,0}} \times \\ &\quad \text{tr} \left[ \Sigma^{-1} (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0}) (\tilde{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0})' \right] \end{aligned}$$

where we used that  $\sum_{t=1}^T \bar{F}_t = 0$ . Since the log-likelihood of the restricted model reads

$$\ln \mathcal{L}_{\text{res}}(\Sigma, \lambda_F) = c_\pi - \frac{T}{2} \ln |\Sigma| - \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right) \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right)' \right],$$

since  $\text{tr} \left[ \Sigma^{-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0}) (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \right] = (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})' \Sigma^{-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}} \lambda_{F,0})$ , the FAR statistic equals:

$$\text{FAR}(\lambda_{F,0}) = 2 [\ln \mathcal{L}_{\text{unres}}(\Sigma) - \ln \mathcal{L}_{\text{res}}(\Sigma, \lambda_F)],$$

where we use  $\tilde{\Sigma}$  as an estimator of the unknown covariance matrix  $\Sigma$ .

**CLR**( $\lambda_{F,0}$ ). We construct the maximal value of the likelihood of the restricted model. We therefore concentrate the likelihood with respect to the different parameters. The log-likelihood reads

$$\ln \mathcal{L}_{\text{res}}(\Sigma, \lambda_F, \mathcal{B}) = c_\pi - \frac{T}{2} \ln |\Sigma| - \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T \left( \mathcal{R}_t - \mathcal{B}(\bar{F}_t + \lambda_F) \right) \left( \mathcal{R}_t - \mathcal{B}(\bar{F}_t + \lambda_F) \right)' \right].$$

When we substitute the maximum likelihood estimator for  $\mathcal{B}$  into the expression of the likelihood, the concentrated likelihood of  $(\Sigma, \lambda_F)$  becomes

$$\ln \mathcal{L}_{\text{res}}(\Sigma, \lambda_F) = c_\pi - \frac{T}{2} \ln |\Sigma| - \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \sum_{t=1}^T \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_F) \right) \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_F) \right)' \right].$$

The maximum likelihood estimator of  $\Sigma$  is

$$\tilde{\Sigma} = \frac{1}{T} \sum_{t=1}^T \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_F) \right) \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_F) \right)'$$

and the concentrated likelihood of  $\lambda_F$  reads

$$\begin{aligned} \ln \mathcal{L}_{\text{res}}(\lambda_F) &= c_\pi - \frac{T}{2} \ln |\tilde{\Sigma}| - \frac{1}{2} \text{tr} [I_n T^{-1}] \\ &= c_T - \frac{T}{2} \ln |\tilde{\Sigma}|, \end{aligned}$$

with  $c_T = c_\pi - \frac{n}{2T}$ . The determinant of  $\tilde{\Sigma}$  can be decomposed as

$$\begin{aligned} |\tilde{\Sigma}| &= \left| \frac{1}{T} \sum_{t=1}^T \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_F) \right) \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_F) \right)' \right| \\ &= \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' - \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_F)' \right] \left[ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F)' \right]^{-1} \right. \\ &\quad \left. \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_F)' \right]' \right| \\ &= \left| \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' & \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_F)' \\ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) \mathcal{R}_t' & \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F)' \end{pmatrix} \right| \\ &= \left| \begin{pmatrix} I_n & 0 \\ -\left[ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F)' \right]^{-1} \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) \mathcal{R}_t' & \left[ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F)' \right]^{-1} \end{pmatrix} \right| \\ &= \left| \begin{pmatrix} I_n & 0 \\ -\frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) \mathcal{R}_t' \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right]^{-1} & I_{kf} \end{pmatrix} \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' & \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_F)' \\ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) \mathcal{R}_t' & \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F)' \end{pmatrix} \right| \\ &= \left| \begin{pmatrix} I_n & 0 \\ -\left[ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F)' \right]^{-1} \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) \mathcal{R}_t' & \left[ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F)' \right]^{-1} \end{pmatrix} \right| \\ &= \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right| \left| \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F)' - \sum_{t=1}^T \frac{1}{T} (\bar{F}_t + \lambda_F) \mathcal{R}_t' \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right]^{-1} \right. \\ &\quad \left. \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_F)' \right| \left| \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_F) (\bar{F}_t + \lambda_F)' \right|^{-1}, \end{aligned}$$

where we used that the multiplications by the triangular matrices do not alter the overall value of the determinants. We can now use the  $(k+1) \times 1$  vectors  $q_i : i = 1, \dots, k+1$  which are such that

$$\begin{aligned} 1. & & q'_i \left( \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right) q_j &= 0 & i \neq j \\ 2. & & &= 1 & i = j \\ 3. & q'_i \left( \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \mathcal{R}'_t \right] \Sigma^{-1} \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] \right) q_j &= 0 & i \neq j \\ 4. & & &= \mu_i & i = j. \end{aligned}$$

The total number of restrictions<sup>10</sup> is  $\frac{1}{2}k(k+1) + k + 1 + \frac{1}{2}k(k+1) = (k+1)^2$  which equals the total number of elements of the  $q_i$  vectors. We order the vectors  $q_i$  in a descending order according to the value of  $\mu_i$ :  $(1 >) \mu_1 \geq \mu_2 \geq \dots \geq \mu_{k+1} (> 0)$  and specify  $w_1 : 1 \times k$ ,  $W_2 : k \times k$  as  $\begin{pmatrix} w_1 \\ W_2 \end{pmatrix} = (q_1 \dots q_k)$ . The MLE of  $\lambda_F$  then reads:  $\tilde{\lambda}'_F = w_1 W_2^{-1}$ .

With this value of the MLE:  $(\bar{F}_t + \tilde{\lambda}_F) = \begin{pmatrix} \tilde{\lambda}'_F \\ I_k \end{pmatrix}' \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} = \begin{pmatrix} w_1 W_2^{-1} \\ I_k \end{pmatrix}' \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} = W_2^{-1'} \begin{pmatrix} w_1 \\ W_2 \end{pmatrix}' \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}$  such that

$$\begin{aligned} & \left| \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \tilde{\lambda}_F)(\bar{F}_t + \tilde{\lambda}_F)' - \sum_{t=1}^T \frac{1}{T} (\bar{F}_t + \tilde{\lambda}_F) \mathcal{R}'_t \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}'_t \right]^{-1} \right. \\ & \left. \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \tilde{\lambda}_F)' \right| \left| \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \tilde{\lambda}_F)(\bar{F}_t + \tilde{\lambda}_F)' \right|^{-1} &= \\ & \left| W_2^{-1'} \begin{pmatrix} w_1 \\ W_2 \end{pmatrix}' \left\{ \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] - \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \mathcal{R}'_t \right] \Sigma^{-1} \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] \right\} \begin{pmatrix} w_1 \\ W_2 \end{pmatrix} W_2^{-1} \right| &= \\ & \left| W_2^{-1'} \begin{pmatrix} w_1 \\ W_2 \end{pmatrix}' \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] \begin{pmatrix} w_1 \\ W_2 \end{pmatrix} W_2^{-1} \right|^{-1} &= \\ & \left| (q_1 \dots q_k)' \left\{ \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] - \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \mathcal{R}'_t \right] \Sigma^{-1} \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] \right\} (q_1 \dots q_k) \right| &= \\ & \left| (q_1 \dots q_k)' \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] (q_1 \dots q_k) \right| &= \\ & \left| \begin{pmatrix} 1 - \mu_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 - \mu_K \end{pmatrix} \right| &= \\ & \prod_{i=1}^k (1 - \mu_i), & \end{aligned}$$

so the logarithm of the determinant of  $|\tilde{\Sigma}|$  when we use the MLE  $\tilde{\lambda}_F$  equals

$$\log |\tilde{\Sigma}| = \log \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}'_t \right| + \sum_{i=1}^k \ln(1 - \mu_i) \approx \log \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}'_t \right| + \frac{T}{2} \sum_{i=1}^k \mu_i,$$

since  $\mu_i$  lies between zero and one, and the logarithm of the likelihood becomes

$$\ln \mathcal{L}_{\text{res}}(\tilde{\lambda}_F) \approx c_T - \frac{T}{2} \log \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}'_t \right| + \frac{T}{2} \sum_{i=1}^k \mu_i.$$

This expression shows that the largest  $k$  values of  $\mu_i, \mu_1, \dots, \mu_k$ , maximize the likelihood so  $\tilde{\lambda}_F$  is indeed the MLE. Thus the MLE results from the characteristic polynomial

$$\left| \theta \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] - \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \mathcal{R}'_t \right] \Sigma^{-1} \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] \right| = 0,$$

where  $\Sigma = \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}'_t$ , and  $\lambda_F$  results from the  $k$  eigenvectors  $(q_1 \dots q_k) : (k+1) \times k$  of the polynomial that are associated with the  $k$  largest eigenvalues. When we express these as  $(q_1 \dots q_k) = \begin{pmatrix} w_1 \\ W_2 \end{pmatrix}$ ,  $w_1 : 1 \times k$ ,  $W_2 : k \times k$ , the maximum likelihood estimator of  $\lambda_F$  equals:  $\tilde{\lambda}_F = w_1 W_2^{-1}$ .

<sup>10</sup>The number of restrictions in 1 and 3 are equal to  $\frac{1}{2}k(k+1)$  while the number of restrictions from 2 equals  $k+1$ .

To obtain the likelihood ratio statistic, we express the characteristic polynomial as a function of  $\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{F,0}$  and  $\tilde{\mathcal{B}}$  which does not affect the roots  $\theta_i$ ,  $i = 1, \dots, k$ .

$$\begin{aligned}
& \left| \theta \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] - \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \mathcal{R}'_t \right] \Sigma^{-1} \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] \right| &= 0 \Leftrightarrow \\
& \left| \begin{pmatrix} 1 & 0 \\ \lambda_F & I_k \end{pmatrix} \left[ \theta \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' - \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix} \mathcal{R}'_t \Sigma^{-1} \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \begin{pmatrix} 1 \\ \bar{F}_t \end{pmatrix}' \right] \begin{pmatrix} 1 & 0 \\ \lambda_F & I_k \end{pmatrix}' \right| &= 0 \Leftrightarrow \\
& \left| \theta \begin{pmatrix} 1 & \lambda'_F \\ \lambda_F & \hat{Q}_{FF}(\lambda_F) \end{pmatrix} - \left( \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \bar{\mathcal{R}}' \\ \bar{F}_t + \lambda_F \end{pmatrix} \mathcal{R}'_t \right) \Sigma^{-1} \left( \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \bar{\mathcal{R}}' \\ \bar{F}_t + \lambda_F \end{pmatrix} \mathcal{R}'_t \right)' \right| &= 0 \Leftrightarrow \\
& \left| \begin{pmatrix} 1 & -\lambda'_F \hat{Q}_{FF}(\lambda_F)^{-1} \\ 0 & I_k \end{pmatrix} \left[ \theta \begin{pmatrix} 1 & \lambda'_F \\ \lambda_F & \hat{Q}_{FF}(\lambda_F) \end{pmatrix} - \left( \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \bar{\mathcal{R}} \\ \bar{F}_t + \lambda_F \end{pmatrix} \mathcal{R}'_t \right) \Sigma^{-1} \right. \right. \\
& \left. \left. \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \bar{\mathcal{R}} \\ \bar{F}_t + \lambda_F \end{pmatrix} \mathcal{R}'_t \right) \right] \begin{pmatrix} 1 & -\lambda'_F \hat{Q}_{FF}(\lambda_F)^{-1} \\ 0 & I_k \end{pmatrix}' \right| &= 0 \Leftrightarrow \\
& \left| \theta \begin{pmatrix} (1-\lambda'_F \hat{Q}_{FF}(\lambda_F)^{-1} \lambda_F) & 0 \\ 0 & \hat{Q}_{FF}(\lambda_F) \end{pmatrix} - \left( \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_F \\ \bar{F}_t + \lambda_F \end{pmatrix} \mathcal{R}'_t \right) \Sigma^{-1} \left( \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_F \\ \bar{F}_t + \lambda_F \end{pmatrix} \mathcal{R}'_t \right)' \right| &= 0 \Leftrightarrow \\
& \left| \theta I_{k+1} - \left( \Sigma^{-\frac{1}{2}} (\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_F) \frac{1}{\sqrt{1-\lambda'_F \hat{Q}_{FF}(\lambda_F) \lambda_F}} \Sigma^{-\frac{1}{2}} \tilde{\mathcal{B}} \hat{Q}_{FF}(\lambda_F)^{\frac{1}{2}} \right)' \right. \\
& \left. \left( \Sigma^{-\frac{1}{2}} (\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_F) \frac{1}{\sqrt{1-\lambda'_F \hat{Q}_{FF}(\lambda_F) \lambda_F}} \Sigma^{-\frac{1}{2}} \tilde{\mathcal{B}} \hat{Q}_{FF}(\lambda_F)^{\frac{1}{2}} \right) \right| &= 0 \Leftrightarrow \\
& \left| \theta I_{k+1} - \frac{1}{T} \begin{pmatrix} \text{FAR}(\lambda_F) & s(\lambda_F)' \\ s(\lambda_F) & r(\lambda_F) \end{pmatrix} \right| &= 0,
\end{aligned}$$

with

$$\begin{aligned}
r(\lambda_F) &= T \hat{Q}_{FF}(\lambda_F)^{\frac{1}{2}} \tilde{\mathcal{B}}' \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}'_t \right]^{-1} \tilde{\mathcal{B}} \hat{Q}_{FF}(\lambda_F)^{\frac{1}{2}}, \\
s(\lambda_F) &= T \hat{Q}_{FF}(\lambda_F)^{\frac{1}{2}} \tilde{\mathcal{B}}' \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}'_t \right]^{-1} (\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_F) \frac{1}{\sqrt{1-\lambda'_F \hat{Q}_{FF}(\lambda_F) \lambda_F}}.
\end{aligned}$$

When  $k = 1$ , the characteristic polynomial reads

$$\begin{aligned}
& \left| \theta I_{k+1} - \frac{1}{T} \begin{pmatrix} \text{FAR}(\lambda_F) & s(\lambda_F)' \\ s(\lambda_F) & r(\lambda_F) \end{pmatrix} \right| &= 0 \Leftrightarrow \\
& (\theta - \frac{1}{T} \text{FAR}(\lambda_F)) (\theta - \frac{1}{T} r(\lambda_F)) - s(\lambda_F) s(\lambda_F)' &= 0 \Leftrightarrow \\
& \theta^2 - \theta \frac{1}{T} (\text{FAR}(\lambda_F) + r(\lambda_F)) + \frac{1}{T^2} [\text{FAR}(\lambda_F) r(\lambda_F) - s(\lambda_F) s(\lambda_F)'] &= 0,
\end{aligned}$$

and the roots of the polynomial are characterized by

$$\theta = \frac{1}{2T} [\text{FAR}(\lambda_F) + r(\lambda_F) \pm \sqrt{(\text{FAR}(\lambda_F) + r(\lambda_F))^2 - 4r(\lambda_F) (\text{FAR}(\lambda_F) - \text{GLS-LM}(\lambda_F))}].$$

The largest root, which leads to the minimal value of  $|\tilde{\Sigma}|$ , is therefore equal to

$$\theta_{\max} = \frac{1}{2T} [\text{FAR}(\lambda_F) + r(\lambda_F) + \sqrt{(\text{FAR}(\lambda_F) + r(\lambda_F))^2 - 4r(\lambda_F) (\text{FAR}(\lambda_F) - \text{GLS-LM}(\lambda_F))}].$$

and the maximal value of the likelihood of the restricted model is

$$\begin{aligned}
\max_{\lambda_F} \ln \mathcal{L}_{\text{res}}(\Sigma, \lambda_F) &= c_T - \frac{T}{2} \ln [\min_{\lambda_F} |\Sigma|] \\
&= c_T - \frac{T}{2} \ln \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}'_t \right| - \frac{T}{2} \ln(1 - \theta_{\max}) \\
&\approx c_T - \frac{T}{2} \ln \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}'_t \right| + \frac{T}{2} \theta_{\max}.
\end{aligned}$$

To simplify the value of the likelihood under  $H_0 : \lambda_F = \lambda_{F,0}$ , we decompose the determinant of the estimator of  $\Sigma$  under  $H_0$  :

$$\begin{aligned}
& \left| \frac{1}{T} \sum_{t=1}^T \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right) \left( \mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}) \right)' \right| \\
&= \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' - \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_{F,0}) \right] \left[ \frac{1}{T} \sum_{t=1}^T (\bar{F}_t + \lambda_{F,0}) (\bar{F}_t + \lambda_{F,0})' \right]^{-1} \right. \\
&\quad \left. \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_{F,0}) \right]' \right| \\
&= \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right| \left| \hat{Q}_{FF}(\lambda_{F,0}) \right|^{-1} \\
&\quad \left| \hat{Q}_{FF}(\lambda_{F,0}) - \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_{F,0})' \right]' \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right]^{-1} \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t (\bar{F}_t + \lambda_{F,0})' \right] \right| \\
&= \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right| \left| I_{k_f} - \hat{Q}_{FF}(\lambda_F)^{\frac{1}{2}} \tilde{\mathcal{B}}' \left[ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right]^{-1} \tilde{\mathcal{B}} \hat{Q}_{FF}(\lambda_F)^{\frac{1}{2}} \right| \\
&= \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right| \left| I_{k_f} - \frac{1}{T} \mathfrak{r}(\lambda_F) \right|.
\end{aligned}$$

When  $k = 1$ , we can therefore characterize the likelihood ratio statistic for testing  $H_0 : \lambda_F = \lambda_{F,0}$  against  $H_1 : \lambda_F \neq \lambda_{F,0}$  by

$$\begin{aligned}
& \text{LR}(\lambda_{F,0}) \\
&= 2 \left[ \max_{\lambda_F} \ln \mathcal{L}_{\text{res}}(\lambda_F) - \ln \mathcal{L}_{\text{res}}(\lambda_{F,0}) \right] \\
&\approx 2 \left[ c_T - \frac{T}{2} \ln \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right| + \frac{T}{2} \theta_{\max} - c_T + \frac{T}{2} \ln \left( \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right| \left| I_{k_f} - \frac{1}{T} \mathfrak{r}(\lambda_F) \right| \right) \right] \\
&\approx 2 \left[ c_T - \frac{T}{2} \ln \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right| + \frac{T}{2} \theta_{\max} - c_T + \frac{T}{2} \ln \left| \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \right| + \frac{T}{2} \ln \left| I_{k_f} - \frac{1}{T} \mathfrak{r}(\lambda_F) \right| \right] \\
&\approx 2 \left[ \frac{T}{2} \theta_{\max} - \frac{1}{2} \mathfrak{r}(\lambda_F) \right] \\
&\approx \left[ \text{FAR}(\lambda_{F,0}) - \mathfrak{r}(\lambda_{F,0}) + \sqrt{(\text{FAR}(\lambda_{F,0}) + \mathfrak{r}(\lambda_{F,0}))^2 - 4\mathfrak{r}(\lambda_{F,0}) (\text{FAR}(\lambda_{F,0}) - \text{GLS-LM}(\lambda_{F,0}))} \right].
\end{aligned}$$

**H. Size distortions of statistics that result from the inversion of a large dimensional covariance matrix.** The expression of the GLS Wald statistic is

$$\text{GLS-W}(\lambda_{F,0}) = (\bar{R} - \iota_n \hat{\lambda}_1 - \hat{\beta} \lambda_{F,0})' \hat{\Omega}^{-1} \hat{\beta} \hat{\Theta}^{-1} \hat{\beta}' \hat{\Omega}^{-1} (\bar{R} - \iota_n \hat{\lambda}_1 - \hat{\beta} \lambda_{F,0}),$$

with  $\hat{\Theta} = \hat{\beta}' \hat{\Omega}^{-1} \hat{\beta} - \hat{\beta} \hat{\Omega}^{-1} \iota_n (\iota_n' \hat{\Omega}^{-1} \iota_n)^{-1} \iota_n' \hat{\Omega}^{-1} \hat{\beta}$ . The expression for  $\hat{\Omega}$  is:

$$\hat{\Omega} = \frac{1}{T-k-1} \sum_{t=1}^T (\bar{R}_t - \hat{\beta} \bar{F}_t) (\bar{R}_t - \hat{\beta} \bar{F}_t)'.$$

To understand the size distortion of GLS-W, we specify it using matrix notation:

$$\text{GLS-W}(\lambda_{F,0}) = \left[ \frac{1}{T} \iota_T' (R - \iota_T (\hat{\lambda}_1 \iota_n' + \lambda_{F,0} \hat{\beta}') - \hat{F} \hat{\beta}') \right] \hat{\Omega}^{-1} \hat{\beta} \hat{\Theta}^{-1} \hat{\beta}' \hat{\Omega}^{-1} \left[ \frac{1}{T} (R - \iota_T (\hat{\lambda}_1 \iota_n' + \lambda_{F,0} \hat{\beta}') - \hat{F} \hat{\beta}')' \iota_T \right],$$

with  $R = (R_1' \dots R_T')'$ ,  $\hat{F} = (\bar{F}_1' \dots \bar{F}_T')'$ , so  $\iota_T' \hat{F} = 0$ , and  $\hat{\Omega} = \frac{1}{T-k-1} R' M_{(\iota_T \hat{F})} R$ . When  $T$  equals  $n$ ,  $\hat{\Omega}$  is a  $T \times T$  matrix of rank  $T - 2$  since  $M_{(\iota_T \hat{F})}$  is a  $T \times T$  matrix of rank  $T - 2$ . The eigenvectors associated with the zero eigenvalues of  $M_{(\iota_T \hat{F})}$  are spanned by  $\iota_T$  and  $\hat{F}$ . The eigenvectors associated with the zero eigenvalues of  $\hat{\Omega}$  are obtained by specifying  $R$  as

$$R = P_{(\iota_T \hat{F})} R + M_{(\iota_T \hat{F})} R.$$

The first part of this specification of  $R$ ,  $P_{(\iota_T \hat{F})} R$ , is mapped onto zero when  $R$  is multiplied by  $M_{(\iota_T \hat{F})}$ . When  $T = n$ , the eigenvectors that are associated with a zero value of  $\hat{\Omega}$  are therefore spanned by

$R'(\nu_T \hat{F})$  and the eigenvectors that belong to the non-zero eigenvalues are spanned by the orthogonal complement of  $R'(\nu_T \hat{F})$ . For large values of  $n$  smaller than  $T$ ,  $\hat{\Omega}$  is non-singular but will have two relatively small eigenvalues whose eigenvectors are spanned by  $R'(\nu_T \hat{F})$ . In the GLS-Wald statistic,  $\hat{\Omega}^{-1}$  is post-multiplied by  $(R - \nu_T(\hat{\lambda}_1 + \hat{\beta}'\lambda_{F,0}) - \hat{F}\hat{\beta}')'\nu_T$ . This vector lies in the span of  $R'(\nu_T \hat{F})$  which is the eigenvector associated with the smallest eigenvalues of  $\hat{\Omega}$  or put differently the largest eigenvalues of  $\hat{\Omega}^{-1}$ . This shows that the large size distortion of the GLS-Wald statistics results from the association between  $(R - \nu_T(\hat{\lambda}_1 + \hat{\beta}'\lambda_{F,0}) - \hat{F}\hat{\beta}')'\nu_T$  and the eigenvectors that belong to the largest eigenvalues of  $\hat{\Omega}^{-1}$ .

When we use the covariance matrix estimator,

$$\tilde{\Sigma} = \frac{1}{T-k-1} \sum_{t=1}^T (\mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}))(\mathcal{R}_t - \tilde{\mathcal{B}}(\bar{F}_t + \lambda_{F,0}))',$$

the matrix expression of this covariance matrix estimator reads

$$\tilde{\Omega} = \frac{1}{T-k-1} \mathcal{R}' M_{\hat{F} + \nu_T \lambda'_{F,0}} \mathcal{R}.$$

with  $\mathcal{R} = (\mathcal{R}'_1 \dots \mathcal{R}'_T)'$ . Since  $\nu'_T(\mathcal{R} - (\hat{F} + \nu_T \lambda'_{F,0})\tilde{\mathcal{B}}')$  is not spanned by  $\mathcal{R}'(\hat{F} + \nu_T \lambda'_{F,0})$ ,  $\nu'_T(\mathcal{R} - (\hat{F} + \nu_T \lambda'_{F,0})\tilde{\mathcal{B}}')$  is not associated with the eigenvector of the largest eigenvalue of  $\tilde{\Sigma}^{-1}$  and therefore all statistics that use  $\tilde{\Sigma}^{-1}$  perform appropriately despite the inversion of a large dimensional matrix.

**I. Covariance matrix maximum likelihood estimator.** The asymptotic covariance matrix of  $(\text{vec}(\tilde{\mathcal{B}}), \tilde{\lambda}_F)$ , where  $\tilde{\mathcal{B}}$  results from (20) using  $\lambda_{F,0}$  equal to  $\tilde{\lambda}_F$ , reads

$$\text{cov}((\text{vec}(\tilde{\mathcal{B}}), \tilde{\lambda}_F)) = \begin{pmatrix} \Sigma^{-1} \otimes \sum_{t=1}^T (\bar{F}_t + \tilde{\lambda}_F)(\bar{F}_t + \tilde{\lambda}_F)' & \Sigma^{-1} \tilde{\mathcal{B}} \otimes \sum_{t=1}^T (\bar{F}_t + \tilde{\lambda}_F) \\ \tilde{\mathcal{B}}' \Sigma^{-1} \otimes \sum_{t=1}^T (\bar{F}_t + \tilde{\lambda}_F)' & \tilde{\mathcal{B}}' \Sigma^{-1} \tilde{\mathcal{B}} \otimes T \end{pmatrix}^{-1},$$

where we used that the covariance matrix equals the inverse of the information matrix. The covariance matrix of  $\tilde{\lambda}_F$  then equals (using properties of the partitioned inverse):

$$\begin{aligned} \text{cov}(\tilde{\lambda}_F) &= \left( \tilde{\mathcal{B}}' \Sigma^{-1} \tilde{\mathcal{B}} \right)^{-1} \otimes \left[ T - \sum_{t=1}^T (\bar{F}_t + \tilde{\lambda}_F)' \left( \sum_{t=1}^T (\bar{F}_t + \tilde{\lambda}_F)(\bar{F}_t + \tilde{\lambda}_F)' \right)^{-1} \sum_{t=1}^T (\bar{F}_t + \tilde{\lambda}_F) \right]^{-1} \\ &= \left( \tilde{\mathcal{B}}' \Sigma^{-1} \tilde{\mathcal{B}} \right)^{-1} \otimes \left( T \left[ 1 - \tilde{\lambda}'_F \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t + \tilde{\lambda}_F \tilde{\lambda}'_F \right)^{-1} \tilde{\lambda}_F \right] \right)^{-1} \\ &= \left( \tilde{\mathcal{B}}' \Sigma^{-1} \tilde{\mathcal{B}} \right)^{-1} \otimes \left( T \left[ 1 - \tilde{\lambda}'_F \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} \tilde{\lambda}_F + \tilde{\lambda}'_F \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} \tilde{\lambda}_F \right. \right. \\ &\quad \left. \left. \left[ 1 + \tilde{\lambda}'_F \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} \tilde{\lambda}_F \right]^{-1} \tilde{\lambda}'_F \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} \tilde{\lambda}_F \right] \right)^{-1} \\ &= \left( \tilde{\mathcal{B}}' \Sigma^{-1} \tilde{\mathcal{B}} \right)^{-1} \otimes \left( \frac{T}{1 + \tilde{\lambda}'_F \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} \tilde{\lambda}_F} \right)^{-1} \\ &= \frac{1}{T} \left( 1 + \tilde{\lambda}'_F \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} \tilde{\lambda}_F \right) \left( \tilde{\mathcal{B}}' \Sigma^{-1} \tilde{\mathcal{B}} \right)^{-1}, \end{aligned}$$

where we used that  $\sum_{t=1}^T \bar{F}_t = 0$  so  $\sum_{t=1}^T \bar{F}_t + \tilde{\lambda}_F = T\tilde{\lambda}_F$  and that

$$\begin{aligned} \left( \tilde{\lambda}_F \tilde{\lambda}'_F + \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} &= \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} - \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} \tilde{\lambda}_F \\ &\quad \left[ 1 + \tilde{\lambda}'_F \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1} \tilde{\lambda}_F \right]^{-1} \tilde{\lambda}'_F \left( \frac{1}{T} \sum_{t=1}^T \bar{F}_t \bar{F}'_t \right)^{-1}. \end{aligned}$$

## References

- [1] Anderson, T.W. and H. Rubin. Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics*, **21**:570–582, (1949).
- [2] Bekker, P.A. Alternative Approximations to the Distributions of Instrumental Variable Estimators. *Econometrica*, **62**:657–681, 1994.
- [3] Cochrane, J.H. *Asset Pricing*. Princeton University Press, 2001.
- [4] Drèze, J.H. and J.F. Richard. Bayesian Analysis of Simultaneous Equations systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1*. Elsevier Science (Amsterdam), (1983).
- [5] Fama, E.F. and J.D. MacBeth. Risk, Return and Equilibrium: Empirical Tests. *Journal of Political Economy*, **81**:607–636, 1973.
- [6] Fama, E.F. and K.R. French. The Cross-Section of Expected Stock Returns. *Journal of Finance*, **47**:427–465, 1992.
- [7] Fama, E.F. and K.R. French. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, **33**:3–56, 1993.
- [8] Fama, E.F. and K.R. French. Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance*, **51**:55–84, 1996.
- [9] Gibbons, M. Multivariate Tests of Financial Models: A New Approach. *Journal of Financial Economics*, **10**:3–27.
- [10] Hansen, L.P. Large Sample Properties of Generalized Method Moments Estimators. *Econometrica*, **50**:1029–1054, 1982.
- [11] Hansen, L.P. and R. Jagannathan. Assessing Specification Errors in Stochastic Discount Factor Models. *Journal of Finance*, **52**:557–590, 1997.
- [12] Jagannathan, R. and Z. Wang. The Conditional CAPM and the Cross-Section of Expected Returns. *Journal of Finance*, **51**:3–53, 1996.
- [13] Jagannathan, R. and Z. Wang. An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression. *Journal of Finance*, **53**:1285–1309, 1998.
- [14] Kan, R. and C. Zhang. Two-Pass Tests of Asset Pricing Models with Useless Factors. *Journal of Finance*, **54**:203–235, 1999.
- [15] Kleibergen, F. Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, **70**:1781–1803, 2002.
- [16] Kleibergen, F. Testing Parameters in GMM without assuming that they are identified. *Econometrica*, **73**:1103–1124, 2005.
- [17] Kleibergen, F. Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, 2006. Forthcoming.

- [18] Kleibergen, F. Subset Statistics in the Linear IV Regression Model. 2006. Technical Report, Brown University.
- [19] Lintner, J. The Valuation of Risk Assets and the Selection of Risky Investment in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, **47**:13–37, 1965.
- [20] Moreira, M.J.,. A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, **71**:1027–1048, 2003.
- [21] Muirhead, R.J. *Aspects of Multivariate Statistical Theory*. John Wiley (New York), 1982.
- [22] Newey, W.K. and R.J. Smith. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica*, **72**:219–255, 2004.
- [23] Phillips, P.C.B. Partially Identified Econometric Models. *Econometric Theory*, **5**:181–240, (1989).
- [24] Shanken, J. On the Estimation of Beta-Pricing Models. *Review of Financial Studies*, **5**:1–33, 1992.
- [25] Staiger, D. and J.H. Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65**:557–586, 1997.
- [26] Zellner, A. *An Introduction to Bayesian Inference in Econometrics*. Wiley, (1971).