UvA ECONOMETRICS

Discussion Paper: 2014/05

Unexplained factors and their effects on second pass R-squared’s

Frank Kleibergen and Zhaoguo Zhan

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Department of Economics & Econometrics
Valckenierstraat 65-67
1018 XE AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Unexplained factors and their effects on second pass R-squared's*

Frank Kleibergen† Zhaoguo Zhan‡

December 16, 2014

Abstract

We construct the large sample distributions of the OLS and GLS R^2's of the second pass regression of the Fama-MacBeth (1973) two pass procedure when the observed proxy factors are minorly correlated with the true unobserved factors. This implies an unexplained factor structure in the first pass residuals and, consequently, a large estimation error in the estimated beta's which is spanned by the beta's of the unexplained true factors. The average portfolio returns and the estimation error of the estimated beta's are then both linear in the beta's of the unobserved true factors which leads to possibly large values of the OLS R^2 of the second pass regression. These large values of the OLS R^2 are not indicative of the strength of the relationship. Our results question many empirical findings that concern the relationship between expected portfolio returns and (macro-) economic factors.

JEL Classification: G12

Keywords: Fama-MacBeth two pass procedure; factor pricing; stochastic discount factors; weak identification; (non-standard) large sample distribution; principal components

1 Introduction

An important part of the asset pricing literature is concerned with the relationship between portfolio returns and (macro-) economic factors. Support for such an relationship is often

*We thank the editor, associate editor and two anonymous referees for helpful comments and suggestions.

†Department of Economics, Brown University, Box B, Providence, RI 02912, United States, and Department of Economics & Econometrics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE Amsterdam, The Netherlands. Email: f.r.kleibergen@uva.nl. Homepage: http://www.econ.brown.edu/fac/Frank_Kleibergen.

‡Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China, 100084. Email: zhanzhg@sem.tsinghua.edu.cn.

established using the Fama-MacBeth (FM) two pass procedure, see *e.g.* Fama and MacBeth (1973), Gibbons (1982), Shanken (1992) and Cochrane (2001). The first pass of the FM two pass procedure estimates the β's of the (macro-) economic factors using a linear factor model, see *e.g.* Lintner (1965) and Fama and French (1992, 1993, 1996). In the second pass, the average portfolio returns are regressed on the estimated β's from the first pass to yield the estimated risk premia, see *e.g.* Jagannathan and Wang (1996, 1998), Lettau and Ludvigson (2001), Lustig and Van Niewerburgh (2005), Li *et. al.* (2006) and Santos and Veronesi (2006). The ordinary and generalized least squares R^2's of the second pass regression alongside t-statistics of the risk premia are used to gauge the strength of the relationship between the expected portfolio returns and the involved factors.

Recently, the appropriateness of these measures has been put into question when the β's are small. An early critique is Kan and Zhang (1999) who show that the second pass t-statistic increases with the sample size when the true β's are zero and the expected portfolio returns are non-zero, so there is no factor pricing. Kleibergen (2009) shows that the second pass t-statistic also behaves in a non-standard manner when the β's are non-zero but small and factor pricing is present so the expected portfolio returns are proportional to the (small) β's. To remedy these testing problems, Kleibergen (2009) proposes identification robust factor statistics that remain trustworthy even when the β's of the observed factors are small or zero.

Burnside (2010) does not focus on properties of second pass statistics, like R^2's and t-statistics, but argues that β's of observed factors which are close to zero, or which cannot be rejected to be equal to zero, invalidate a relationship between expected portfolio returns and involved factors. Daniel and Titman (2012) do not focus on the behavior of second pass statistics either but argue that the relationship between expected portfolio returns and involved factors depends on the manner in which the portfolios are constructed. When portfolios are not based on sorting with respect to book-to-market ratios and size, a relationship between expected portfolios returns and observed factors is often absent.

Lewellen *et. al.* (2010) criticize the use of the ordinary least squares (OLS) R^2 of the second pass regression. They show that it can be large despite that the β's of the observed factors are small or even zero and propose a few remedies. Lewellen *et. al.* (2010) do not provide a closed form expression of the large sample distribution of the OLS R^2 so it remains unclear why the OLS R^2 can be large despite that the β's of the observed factors are small or zero. The same argument applies to one of their remedies which is the generalized least squares (GLS) R^2. We therefore construct the expressions of the large sample distributions of both the OLS and GLS R^2's when the β's of the observed factors are small and possibly zero.

We derive the large sample distributions of the OLS and GLS R^2's starting out from factor pricing based on a small number of true possibly unknown factors. These factors imply an unobserved factor structure for the portfolio returns. The observed (proxy) factors used in the FM two pass procedure proxy for these unobserved true factors. When they are only minorly correlated with the true factors, a sizeable unexplained factor structure remains in the first pass residuals. Consequently also a sizeable estimation error in the estimated β's exists which is, as we show, to a large extent spanned by the β's of the unexplained factors. The expected portfolio returns are linear in the β's of the unobserved factors so both the average portfolio returns and the estimation error of the estimated β's are to a large extent linear in the β's of the unobserved true factors when the observed proxy and unobserved true factors are only minorly correlated. As further shown by the expression of the large sample distribution of the OLS R^2, this produces the large values of the OLS R^2 of the second pass regression when we regress the average portfolio returns on the estimated β's from the first pass regression and the observed proxy and unobserved true factors are only minorly correlated.

When the observed factors provide an accurate proxy of the unobserved true factors, the estimated β's from the first pass regression are spanned by the β's of the true factors and the OLS R^2 is large, see Lewellen *et. al.* (2010). Hence, both when the observed proxy factors are strongly or minorly correlated with the unobserved true factors, the OLS R^2 can be large. In the latter case, the large value, however, results from the estimation error in the estimated β's. An easy diagnostic for how a large value of the OLS R^2 should be interpreted therefore results from the unexplained factor structure in the first pass residuals. When this unexplained factor structure is considerable, a large value of the OLS R^2 is caused by it so the large value of the OLS R^2 is not indicative of the strength of the relationship between the expected portfolio returns and the (macro-) economic factors.

The expression of the large sample distribution of the GLS R^2 shows that it is small when the observed proxy factors are only minorly correlated with the unobserved true factors. It also shows, however, that the GLS R^2 is rather small in general so a small value of the GLS R^2 can result when the observed factors are strongly or minorly correlated with the unobserved true factors. This makes it difficult to gauge the strength of the relationship between the expected portfolio returns and the (macro-) economic factors using the GLS R^2.

To construct the expressions of the large sample distributions of the OLS and GLS R^2's which are representative for observed proxy factors that are minorly correlated with the unobserved true factors, we assume that the parameters in an (infeasible) linear regression of the true unknown factors on the observed proxy factors are decreasing/drifting with the sample size. Our assumption implies that statistics that test the significance of the observed proxy factors for explaining portfolio returns and the unobserved true factors do not increase with the sample size but stay constant/small when the sample size increases. This is in line with the values of these statistics that we typically observe in practice. Under the traditional assumption of strong correlation between the observed proxy and unobserved true factors, these statistics should all be large and proportional to the sample size. Since this is clearly not the case, the traditional assumption is out of line and provides an inappropriate base for statistical inference in such instances. Our assumption also implies that the estimated risk premia in the second pass regression converge to random variables so they cannot be used in a bootstrap procedure since such a procedure relies upon consistent estimators. The drifting assumption on the regression parameters provides inference which is closely related to so-called finite sample inference but it does not require the disturbances to be normally distributed, see *e.g.* MacKinlay (1987) and Gibbons *et. al.* (1989). It is akin to the weak instrument assumption made for the linear instrumental variables regression model in econometrics, see *e.g.* Staiger and Stock (1997).

Although we focus on the R^2's, the message conveyed in this paper in principle also applies to other second pass inference procedures like, for example, t-tests on the risk premia and tests of factor pricing using J-tests or Hansen and Jagannathan (1997) (HJ) distances. When the observed factors are minorly correlated with the unobserved true factors, these statistics no longer converge to their usual distributions when the sample size gets large, see *e.g.* Kleibergen (2009). The non-standard distributions of these statistics could further induce the spurious support for the observed factors that are substantially different from the unobserved true factors, see Gospodinov *et. al.* (2014) for results on the HJ distance.

The paper is organized as follows. We first in the second section lay out the factor structure in portfolio returns. We show that many of the (macro-) economic factors that are commonly used, like, for example, consumption and labor income growth, housing collateral, consumption-wealth ratio, labor income-consumption ratio, interactions of either one of the latter three with other factors, leave a strong unexplained factor structure in the first pass residuals. In the third section, we discuss the effects of the unexplained factor structure on the OLS and GLS R^2 by constructing expressions for their large sample distributions. The fourth section concludes.

2 Factor Model for Portfolio Returns

Portfolio returns exhibiting an (unobserved) factor structure with k factors result from a statistical model that is characterized by, see *e.g.* Merton (1973), Ross (1976), Roll and Ross (1980), Chamberlain and Rothschild (1983) and Connor and Korajczyk (1988, 1989):

$$r_{it} = \mu_{Ri} + \beta_{i1} f_{1t} + \ldots + \beta_{ik} f_{kt} + \varepsilon_{it}, \qquad i = 1, \ldots, N, \; t = 1, \ldots, T; \tag{1}$$

with r_{it} the return on the i-th portfolio in period t; μ_{Ri} the mean return on the i-th portfolio; f_{jt} the realization of the j-th factor in period t; β_{ij} the factor loading of the j-th factor for the

i-th portfolio, ε_{it} the idiosyncratic disturbance for the i-th portfolio return in the t-th period and N and T the number of portfolios and time periods. We can reflect the factor model in (1) as well using vector notation:

$$R_t = \mu_R + \beta F_t + \varepsilon_t, \tag{2}$$

with $R_t = (r_{1t} \dots r_{Nt})'$, $\mu_R = (\mu_{R1} \dots \mu_{RN})'$, $F_t = (f_{1t} \dots f_{kt})'$, $\varepsilon_t = (\varepsilon_{1t} \dots \varepsilon_{Nt})'$ and

$$\beta = \begin{pmatrix} \beta_{11} & \dots & \beta_{1k} \\ \vdots & \ddots & \vdots \\ \beta_{N1} & \dots & \beta_{Nk} \end{pmatrix}. \tag{3}$$

The vector notation of the factor model in (2) shows that, if the factors F_t, $t = 1, \dots, T$, are i.i.d. with finite variance and are uncorrelated with the disturbances ε_t, $t = 1, \dots, T$, which are i.i.d. with finite variance as well, the covariance matrix of the portfolio returns reads

$$V_{RR} = \beta V_{FF} \beta' + V_{\varepsilon\varepsilon}, \tag{4}$$

with V_{RR}, V_{FF} and $V_{\varepsilon\varepsilon}$ the $N \times N$, $k \times k$ and $N \times N$ dimensional covariance matrices of the portfolio returns, factors and disturbances respectively.

The factors affect many different portfolios simultaneously which allows us to identify the number of factors using principal components analysis, see *e.g.* Anderson (1984, Chap 11). When we construct the spectral decomposition of the covariance matrix of the portfolio returns,

$$V_{RR} = P \Lambda P', \tag{5}$$

with $P = (p_1 \dots p_N)$ the $N \times N$ orthonormal matrix of principal components or characteristic vectors (eigenvectors) and Λ the $N \times N$ diagonal matrix of characteristic roots (eigenvalues) which are in descending order on the main diagonal, the number of factors can be estimated as the number of characteristic roots that are distinctly larger than the other characteristic roots. The literature on selecting the number of factors is vast and contains further refinements of this factor selection procedure and settings with fixed and increasing number of portfolios. We do not contribute to this literature but just use some elements of it to shed light on the effect of the unexplained factor structure on the R^2 used in the FM two pass procedure.

2.1 Factor Structure in Observed Portfolio Returns

We use three different data sets to show the relevance of the factor structure. The first one is from Lettau and Ludvigson (2001). It consists of quarterly returns on twenty-five size and book-to-market sorted portfolios from the third quarter of 1963 to the third quarter of 1998 so $T = 141$ and $N = 25$. The second one is from Jagannathan and Wang (1996) and consists of monthly returns on one hundred size and beta sorted portfolios. The series are from July 1963 to December 1990 so $T = 330$ and $N = 100$. The third data set consists of quarterly returns on twenty-five size and book to market sorted portfolio's and is obtained from Ken French's website. The series are from the first quarter of 1952 to the fourth quarter of 2001 so $T = 200$ and $N = 25$.

	LL01	JW96	F52-01
1	2720	3116	2434
2	113.8	180.2	140.5
3	98.60	80.6	108.9
4	18.36	28.5	26.7
5	17.61	25.4	19.9
6	13.48	16.2	14.0
7	12.11	14.8	11.6
8	9.31	14.0	10.9
9	8.42	12.6	9.92
10	7.25	12.1	8.18
11	6.02	12.1	7.19
12	5.40	11.4	6.32
13	4.90	11.3	6.17
14	4.38	11.1	5.63
15	4.26	10.8	5.21
16	3.93	10.3	5.02
17	3.50	10.2	4.40
18	3.39	9.9	3.83
19	3.02	9.6	3.43
20	2.71	9.5	2.90
21	2.50	9.2	2.79
22	2.18	9.0	2.75
23	1.74	8.9	2.47
24	1.46	8.7	2.12
25	0.93	8.4	1.77
FACCHECK	95.5%	86%	94.3%

Table 1: Largest twenty five characteristic roots (in descending order) of the covariance matrix of the portfolio returns (LL01 stands for Lettau and Ludvigson (2001), JW96 stands for Jagannathan and Wang (1996) and F52-01 stands for the portfolio returns from Ken French's website during 1952-2001). FACCHECK equals the percentage of the variation explained by the three largest principal components.

Table 1 lists the (largest) twenty-five characteristic roots[1] of the three different sets of

[1] The data set from Jagannathan and Wang (1996) consists of one hundred portfolio returns so Table 1, for reasons of brevity, only shows the largest twenty-five characteristic roots.

portfolio returns. Table 1 shows that there is a rapid decline of the value of the roots from the largest to the third largest one and a much more gradual decline from the fourth largest one onwards. This indicates that the number of factors is (most likely) equal to three.

A measure/check for the presence of a factor structure (with three factors) is the fraction of the total variation of the portfolio returns that is explained by the three largest principal components. We measure the total variation by the sum of all characteristic roots.[2] The factor structure check then reads

$$\text{FACCHECK} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{\lambda_1 + \ldots + \lambda_N}, \tag{6}$$

with $\lambda_1 > \lambda_2 > \ldots > \lambda_N$ the characteristic roots in descending order. Table 1 shows that the factor structure check equals 95.5% for the Lettau-Ludvigson (LL01) data, 86% for the Jagannathan and Wang (JW96) data and 94.3% for the French (F52-01) data. Using the statistic proposed in, for example, Anderson (1984, Section 11.7.2), it can be shown that the hypothesis that the three largest principal components explain less than 80% of the variation of the portfolio returns is rejected with more than 95% significance for each of these three data sets.

Similar to the three data sets above, we also find evidence for a factor structure in several other commonly used data sets of financial assets. For example, one set is the conventional twenty-five size and book-to-market sorted portfolios augmented by thirty industry portfolios, as in Lewellen *et. al.* (2010), and another is the individual stock return data from the Center for Research in Security Prices (CRSP). We focus on the three data sets mentioned before and omit the other data sets for brevity since our results and findings extend to these data sets as well.

2.2 Factor Models with Observed Proxy Factors

Alongside describing portfolio returns using "unobserved factors", a large literature exists which explains portfolio returns using observed factors which are to proxy for the unobserved ones. The observed proxy factors that are used consist both of asset return based factors and (macro-) economic factors. The observed factor model is identical to the factor model in (2) but with a value of F_t that is observed and a known value of the number of factors, say m :

$$R_t = \mu + BG_t + U_t, \tag{7}$$

with $G_t = (g_{1t} \ldots g_{mt})'$ the m-dimensional vector of observed proxy factors, $U_t = (u_{1t} \ldots u_{Nt})'$ a N-dimensional vector with disturbances, μ a m-dimensional vector of constants and B the

[2] This corresponds with using the trace norm of the covariance matrix as a measure of the total variation.

$N \times m$ dimensional matrix that contains the β's of the portfolio returns with the observed proxy factors. In the sequel we discuss the observed proxy factors used in seven different articles: Fama and French (1993), Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Li *et. al*. (2006), Lustig and van Nieuwerburgh (2005), Santos and Veronesi (2006) and Yogo (2006).

Fama and French (1993) use the return on a value weighted portfolio, a "small minus big" (SMB) factor which consists of the difference in returns on a portfolio consisting of assets with a small market capitalization minus the return on a portfolio consisting of assets with a large market capitalization and a "high minus low" (HML) factor which consists of the difference in the returns on a portfolio consisting of assets with a high book to market ratio minus the return on a portfolio consisting of assets with a low book to market ratio. We use the portfolio returns on the twenty-five size and book to market sorted portfolio's from Ken French's website to estimate the observed factor model.

Table 2 shows the largest five characteristic roots of the covariance matrix of the portfolio returns and of the covariance matrix of the residuals that result from the observed factor model with the three Fama-French (FF) factors. The characteristic roots and factor structure check show that after incorporating the FF factors, there is no unexplained factor structure left in the residuals.

The characteristic roots of the covariance matrices can be used to test the significance of the parameters associated with the observed proxy factors. The likelihood ratio (LR) statistic for testing the null hypothesis that the parameters associated with the observed factors are all equal to zero, $\mathrm{H}_0 : B = 0$, against the alternative hypothesis that they are unequal to zero, $\mathrm{H}_1 : B \neq 0$, equals

$$\mathrm{LR} = T\left[\log(\left|\hat{V}_{Port}\right|) - \log(\left|\hat{V}_{\mathrm{Res}}\right|)\right] = T\sum_{i=1}^{N}\left[\log(\lambda_{i,\mathrm{port}}) - \log(\lambda_{i,\mathrm{res}})\right], \tag{8}$$

with $\hat{V}_{Port}$ and $\hat{V}_{\mathrm{Res}}$ estimators of the covariance matrix of the portfolio returns and the residual covariance matrix that results after regressing the portfolio returns on the observed factors G, and $\lambda_{i,\mathrm{port}}$, $i = 1, \ldots, N$, the characteristic roots of the covariance matrix of the portfolio returns, $\hat{V}_{Port}$, and $\lambda_{i,\mathrm{res}}$, $i = 1, \ldots, N$, the characteristic roots of the covariance matrix of the residuals of the observed factor model, $\hat{V}_{\mathrm{Res}}$.[3] Under H_0, the LR statistic in (8) has a $\chi^2(3N)$ distribution in large samples. The value of the LR statistic using the FF factors stated in Table 2 is highly significant,[4] see also Bai and Ng (2006).

[3] The expression of the LR statistic in the first part of (8) is standard, see *e.g.* Campbell, Lo and MacKinlay (1997, Eq (5.3.28)) in which there is a typo since the Likelihood Ratio statistic equals twice the difference between the log likelihoods of the restricted and unrestricted models. Upon conducting spectral decompositions of $\hat{V}_{Port}$ and $\hat{V}_{\mathrm{Res}}$, as in (5), the final expression in (8) results.

[4] Instead of using the LR statistic, we could also use a Wald statistic to test for the significance of the factors.

	F52-01		JW96			
	raw	FF factors	raw	FF factors	R_{vw}	JW96 factors
1	2434	54.94	3116	70.5	600.6	594.0
2	140.5	38.87	180.2	50.0	81.24	78.9
3	108.9	22.77	80.6	38.6	48.4	48.0
4	26.7	18.24	28.5	27.3	28.5	28.0
5	19.9	12.47	25.4	16.0	17.1	17.0
FACCHECK	94.3%	47.5%	86%	23%	57%	57%
LR against raw		2064		2994	1586	1845
		0.000		0.000	0.000	0.000
LR against R_{vw}		1285		1408		259
		0.000		0.000		0.004
F-stat HML-SMB						3.51
						0.476
pseudo-R^2 FF				1	0.627	0.794
pseudo-R^2		0.915		0.823	0.681	0.684

Table 2: The largest five characteristic roots (in descending order) of the covariance matrix of the portfolio returns and residuals that result using FF factors (French's website data 1952-2001) and those that result from using the Jagannathan and Wang (1996) data with different observed factors. The likelihood ratio (LR) statistic tests against the indicated specification (p-value is listed below). The F-statistics at the bottom of the table result from testing the significance of the indicated factors in a regression of either the FF factors or only the HML-SMB factors on them. The pseudo-R^2's of the regression of the FF factors or the portfolio returns on the observed factors are listed at the bottom of the table. FACCHECK equals the percentage of the variation explained by the three largest principal components.

Alongside the LR statistic that tests the significance of all the parameters associated with the FF factors, Table 2 also lists three more statistics: another LR statistic, an F-statistic and a goodness of fit measure to which we refer as the pseudo-R^2.

The other LR statistic in Table 2 tests the significance of the parameters associated only with the SMB and HML factors. The expression for this LR statistic is identical to that in (8) when we replace the characteristic roots of the covariance matrix of the raw portfolio returns, $\lambda_{i,\text{port}}$, with the characteristic roots of the covariance matrix of the residuals of an observed factor model that has the value weighted return as the only factor. This LR statistic is highly significant so the parameters of the SMB and HML factors are significant.

The F-statistic reported in Table 2 is the F-statistic (times number of tested parameters) that results from regressing either the FF or just the HML and SMB factors on other observed proxy factors. The F-statistic then results from testing $\mathrm{H}_0 : \delta = 0$ in the linear model:

$$F_t = \quad \mu_F + \delta G_t + V_t, \tag{9}$$

Under homoscedastic independent normal errors, the Wald statistic has an exact F-distribution, see MacKinlay (1987) and Gibbons *et. al.* (1989). We use the LR statistic, for whose distribution we have to rely on a large sample argument, since it is directly connected to the characteristic roots.

with F_t a 3×1 vector that contains the FF factors or a 2×1 vector that consists of the HML and SMB factors and G_t a $m \times 1$ vector containing other observed proxy factors.[5]

The pseudo-R^2 reported in Table 2 is a goodness of fit measure that reflects the percentage of the total variation of the portfolio returns that is explained by the observed proxy factors. We measure the total variation of the portfolio returns by the sum of the characteristic roots of its covariance matrix and similarly for the total variation of the portfolio returns explained by the observed proxy factors. Since the latter equals the total variation of the portfolio returns minus the total variation of the residuals of the regression of the portfolio returns on the observed factors proxy, the pseudo-R^2 reads[6]

$$\text{pseudo-}R^2 = 1 - \frac{\sum_{i=1}^{N} \lambda_{i,res}}{\sum_{i=1}^{N} \lambda_{i,port}}. \tag{10}$$

The pseudo-R^2 in Table 2 shows that the FF factors explain 91.5% of the total variation of the portfolio returns.

Jagannathan and Wang (1996) propose a conditional version of the capital asset pricing model which they estimate using three observed factors: the return on a value weighted portfolio, a corporate bond yield spread and a measure of per capita labor income growth. The characteristic roots in Table 2 show that the latter two factors do not explain any of the (unobserved) factors. This is further emphasized by: the (insignificant) small F-statistic in the regression of the HML and SMB factors on these factors and the value weighted return, the small change in the pseudo-R^2 from just using the value weighted return to all three factors.

Lettau and Ludvigson (2001) use a number of specifications of an observed factor model to estimate different conditional asset pricing models. The observed proxy factors that they consider are the value weighted return (R_{vw}), the consumption-wealth ratio (cay), consumption growth (Δc), labor income growth (Δy), the FF factors and interactions between the consumption-wealth ratio and consumption growth ($cay\Delta c$), the value weighted return ($cayR_{vw}$) and labor income growth ($cay\Delta y$). Our results for the Lettau Ludvigson (2001) data are listed in Table 3.

The characteristic roots in Table 3 show that only the FF factors, which include the value

[5] The F-statistic in (9) assumes that the unobserved factors are well approximated by the FF factors. This is mainly done for expository purposes and might not stand up to more formal testing, see Onatski (2012).

[6] The pseudo-R^2 equals the total variation of the explained sum of squares over the total variation of the portfolio returns so pseudo-$R^2 = \frac{\text{trace}(\hat{V}_{\mu+\hat{B}G})}{\text{trace}(\hat{V}_R)} = 1 - \frac{\text{trace}(\hat{V}_{R-\mu-\hat{B}G})}{\text{trace}(\hat{V}_R)} = 1 - \frac{\sum_{i=1}^{N} \lambda_{i,res}}{\sum_{i=1}^{N} \lambda_{i,port}}$, where the last result is obtained using the spectral decomposition of $\hat{V}_R$ and $\hat{V}_{R-\mu-\hat{B}G}$, see (5), and we used that $\hat{V}_R = \hat{V}_{\mu+\hat{B}G} + \hat{V}_{R-\mu-\hat{B}G}$.

	LL01						
	raw	R_{vw}	Δc	FF factors	cay, R_{vw}, $cayR_{vw}$	$cay, \Delta c$, $cay\Delta c$	$cay, R_{vw}, \Delta y$, $cayR_{vw}, cay\Delta y$
1	2720	435	2676	26.5	433	2414	412
2	114	99.5	111	22.3	98.0	105	97.2
3	98.6	26.2	98.6	14.3	26.0	96.0	25.6
4	18.4	18.36	18.1	13.9	17.9	17.9	17.8
5	17.6	13.8	16.8	11.2	12.9	16.7	12.8
FACCHECK	95.5%	82.1%	95.5%	38.2%	82.5%	95.2%	82.1%
LR against raw		765 0.000	36.4 0.064	1940 0.000	856.4 0.000	128.2 0.001	902.5 0.000
LR against R_{vw}				1175 0.000	91.4 0.000		138.5 0.007
LR against Δc						91.8 0.000	
LR against cay, R_{vw}, $cayR_{vw}$							46.1 0.630
F-stat FF factors		80.1 0.000	3.73 0.292		81.8 0.000	28.9 0.001	90.3 0.000
F-stat HML-SMB					1.91 0.928		10.7 0.381
pseudo-R^2 FF		0.63	0.01	1	0.63	0.094	0.64
pseudo-R^2		0.78	0.016	0.95	0.78	0.10	0.79

Table 3: The largest five characteristic roots (in descending order) of the covariance matrix of the portfolio returns and residuals that result using different specifications from Lettau and Ludvigson (2001). The likelihood ratio (LR) statistic tests against the indicated specification (p-value is listed below). The F-statistics at the bottom of the table result from testing the significance of the indicated factors in a regression of either the FF factors or only the HML-SMB factors on them. The pseudo-R^2's of the regression of the FF factors or the portfolio returns on the observed factors are listed at the bottom of the table. FACCHECK equals the percentage of the variation explained by the three largest principal components.

weighted return, explain any of the unobserved factors. Statistics which are functions of the characteristic roots therefore also show that the other observed factors have minor explanatory power. For example, the LR statistic shows that only the FF factors and value weighted return are strongly significant while it is always less than twice the number of tested parameters for all other observed factors.[7] This indicates that although the LR statistic might be significant at the 95% significance level, the values of the parameters associated with the observed factors are all close to zero.

The F-statistics of the regression of either the FF factors or just the HML and SMB factors on the observed factors reiterate the observation from the LR statistic. They only come out large when the observed factors include one of the FF factors and otherwise at most equal a small multiple times the number of tested parameters. This shows that the parameters are close to zero in such a regression.

[7] For the linear instrumental variables regresssion model Stock and Yogo (2005) have shown that first stage/pass statistics, like, for example, the LR statistic, have to be ten fold the number of tested parameters to yield standard inference for second stage/pass statistics.

	F52-01							
	raw	R_{vw}	$\Delta c_{nondurable}$	FINAN	LVX06	LN05	SV06	Y06
1	2434	465.2	2250	2422	2404	2204	439.7	465.1
2	140.5	140.3	139.5	138.9	137.7	137.0	122.7	139.1
3	108.9	36.7	108.8	105.9	103.9	108.0	36.3	36.1
4	26.7	21.6	26.6	26.6	25.1	26.5	21.5	21.6
5	19.9	16.8	19.9	19.5	19.2	19.7	16.3	16.1
FACCHECK	94.3%	81.4%	93.9%	94.3%	94.3%	93.8%	80.5%	81.5%
LR against raw		854 0.000	41.9 0.019	35.2 0.085	111 0.004	93.5 0.07	972 0.000	904 0.00
LR against FINAN					75.87 0.011			
LR against $\Delta c_{nondurable}$						51.6 0.41		863 0.000
LR against R_{vw}							118 0.000	50.4 0.46
F-stat FF factors		–	22.4 0.000	6.9 0.074	11.5 0.240	31.2 0.000	–	–
pseudo-R^2 FF		0.321	0.028	0.015	0.026	0.043	0.374	0.326
pseudo-R^2		0.723	0.065	0.007	0.015	0.083	0.739	0.724

Table 4: The largest five characteristic roots (in descending order) of the covariance matrix of the portfolio returns and residuals that result using different specifications from Li et. al. (2006) (LVX06), Lustig and Van Nieuwerburgh (2005) (LN05), Santos and Veronesi (2006) (SV06) and Yogo (2006) (Y06). The likelihood ratio (LR) statistic tests against the indicated specification (p-value is listed below). The F-statistics at the bottom of the table result from testing the significance of the indicated factors in a regression of the FF factors on them. The pseudo-R^2 of these regressions are listed at the bottom of the table. FACCHECK equals the percentage of the variation explained by the three largest principal components.

Li, Vassalou and Xing (2006) use investment growth rate in the household sector (HHOLD), nonfinancial corporate firms (NFINCO) and financial companies (FINAN) as factors in an observed factor model. We estimate this model using the quarterly portfolio returns from French's website. The results in Table 4 show that none of these factors explain any of the unobserved factors.

Lustig and Van Nieuwerburgh (2005) employ an observed factor model that contains nondurable consumption growth (Δc_{nondur}), a housing collateral ratio (myfa) and the interaction between nondurable consumption growth and the housing collateral ratio ($\Delta c_{nondur} \times$myfa). We estimate this model using the quarterly portfolio returns from French's website. The results in Table 4 show that these factors do not explain the unobserved factors.

Santos and Veronesi (2006) use adaptations of the factors from Lettau and Ludvigson (2001). Alongside the value weighted return, Santos and Veronesi (2006) use both the consumption-wealth ratio (cay), previously used by Lettau and Ludvigson (2001), and a labor income to consumption ratio (s_w) interacted with the value weighted return as factors. We estimate their specification using the portfolio returns from French's website. Table 4 shows that except for the value weighted return none of these factors explains any of the unobserved factors.

Yogo (2006) considers a specification of the observed factor model that alongside the value weighted return has consumption growth in durables (Δc_{dur}) and nondurables (Δc_{nondur}) as the three observed factors. We estimate this specification using the portfolio returns from French's website. Table 4 again shows that except for the value weighted returns, these factors do not capture any of the factor structure in the portfolio returns.

3 Implications of Missed Factors for the FM Two Pass Procedure

Stochastic discount factor models, see *e.g.* Cochrane (2001), stipulate a relationship between the expected returns on the portfolios and the β's of the portfolio returns with their (unobserved) factors:

$$E(R_t) = \iota_N \lambda_0 + \beta \lambda_F, \tag{11}$$

with ι_N the N-dimensional vector of ones, λ_0 the zero-β return and λ_F the k-dimensional vector of factor risk premia. To estimate the risk premia, Fama and MacBeth (1973) propose a two pass procedure:

1. Estimate the observed factor model in (7) by regressing the portfolio returns R_t on the observed factors G_t to obtain the least squares estimator:

$$\hat{B} = \sum_{t=1}^{T} \bar{R}_t \bar{G}_t' \left(\sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right)^{-1}, \tag{12}$$

   with $\bar{G}_t = G_t - \bar{G}$, $G = \frac{1}{T}\sum_{t=1}^{T} G_t$, $\bar{R}_t = R_t - \bar{R}$ and $\bar{R} = \frac{1}{T}\sum_{t=1}^{T} R_t$.

2. Regress the average returns, $\bar{R}$, on the vector of constants ι_N and the estimated B, to obtain estimates of the zero-β return λ_0 and the risk premia λ_F:

$$\begin{pmatrix} \hat{\lambda}_0 \\ \hat{\lambda}_F \end{pmatrix} = \left[(\iota_N \vdots \hat{B})'(\iota_N \vdots \hat{B}) \right]^{-1} (\iota_N \vdots \hat{B})'\bar{R}. \tag{13}$$

The FM two pass procedure uses the least squares estimator that results from the observed factor model to estimate the risk premia. The adequacy of the results that stem from the FM two pass regression hinges on the ability of the observed factor model to capture the factor structure of the portfolio returns. To highlight this, we specify an (infeasible) linear regression model for the unobserved factors F_t that uses the observed proxy factors G_t as explanatory variables:

$$\begin{aligned} F_t = & \ \mu_F + \delta G_t + V_t \\ \delta = & \ V_{FG} V_{GG}^{-1} \end{aligned} \tag{14}$$

with V_{FG} the covariance between the unobserved and observed factors, $V_{FG} = cov(F_t, G_t)$, and V_{GG} the covariance matrix of the observed factors, $V_{GG} = var(G_t)$, and G_t and V_t are assumed to be uncorrelated with ε_t since F_t is uncorrelated with ε_t.[8] We substitute (14) into (2) to obtain

$$R_t = \ \mu_R + \beta\mu_F + \beta\delta G_t + \beta V_t + \varepsilon_t = \mu + \beta\delta G_t + U_t, \tag{15}$$

with $\mu = \mu_R + \beta\mu_F$, $U_t = \beta V_t + \varepsilon_t$. When the observed proxy factors do not explain the unobserved factors well, δ is small or zero and V_t is large and proportional to the unobserved factor F_t. The large value of V_t then implies an unexplained factor structure in the residuals U_t of the observed factor model (15) since $U_t = \beta V_t + \varepsilon_t$. Alongside the unexplained factor structure, the small value of δ also implies that the estimand of $\hat{B}$ in (12), *i.e.* $\beta\delta$, is small. The traditional results for the FM two pass procedure are derived under the assumption that the estimand of $\hat{B}$ is a full rank matrix so

$$\hat{B} \underset{p}{\rightarrow} \beta\delta, \tag{16}$$

is a full rank matrix, see *e.g.* Fama and MacBeth (1973) and Shanken (1992).

Tables 2-4 in Section 2 show that for many of the observed (macro-) economic factors used in the literature, the estimand of $\hat{B}$, $\beta\delta$, is such that we cannot reject that at least some or even all of its columns are close to zero. Table 1, however, shows that a strong factor structure is present in portfolio returns which can be explained by the FF factors. It implies that all columns of β are non-zero so the proximity to zero of $\beta\delta$ results from a small value of δ. This is also reflected by the F-statistics in Tables 2-4. They test the hypothesis that δ, or some of its rows, is equal to zero. Since F_t is unknown, we approximate it by the FF factors. Tables 2-4 show that, when the elements of δ being tested do not concern the value weigthed return, the F-statistics are either insignificant or just barely significant. The assumption that $\beta\delta$ has a full rank value implies that δ has a full rank value as well. But when δ has a full rank value, the F-statistics in Table 2-4 should all be proportional to the sample size just as they are when we use them to test the significance of elements of δ that are associated with the value weighted

[8] We could allow for correlation between (G_t, V_t) and ε_t. This would not alter our main results but complicate the exposition. We therefore refrained from doing so.

return. The assumption of a full rank value of δ is therefore not supported by the data when it is associated with factors other then the FF factors. A more appropriate assumption is to assume a value of δ that leads to the smallish values of the F-statistics reported in Tables 2-4.

Assumption 1. *When the sample size T increases, the parameter δ in the (infeasible) linear regression model for the unobserved factors that uses the observed proxy factors as explanatory variables (14) is drifting to zero:*

$$\delta = \tfrac{d}{\sqrt{T}}, \tag{17}$$

with d a fixed $k \times m$ dimensional full rank matrix, while the number of portfolios N stays fixed.

Traditional large sample inference requires that both β and δ are full rank matrices which is not realistic in many applications. In so-called finite sample inference, no assumptions are made with respect to β and δ and instead the disturbances of (15) are assumed to be i.i.d. normal, see $e.g.$ MacKinlay (1987) and Gibbons $et.\ al.$ (1989). Traditional large sample inference generalizes finite sample inference in the sense that it does not require the disturbances to be normally distributed. The price paid for this is that β and δ have to have fixed full rank values. Assumption 1 provides a generalization to both finite sample and traditional large sample inference since it neither assumes fixed full rank values for β and δ nor normally distributed disturbances. Identical to finite sample inference, the results obtained from it therefore apply to small values of β and δ but do not require normality of the disturbances. Assumption 1 is similar to the weak-instrument assumption made in econometrics, see $e.g.$ Staiger and Stock (1997). Assumption 1 seems unrealistic but must solely be seen from the perspective that it leads to the smallish values of the F-statistics that test the significance of δ in (14) as reported in Tables 2-4.

Theorem 1. *Under Assumption 1, the (infeasible) F-statistic testing the significance of δ in (14) converges, when the sample size T goes to infinity, to a non-central χ^2 distributed random variable with km degrees of freedom and non-centrality parameter trace($d^{*\prime}d^*$), $d^* = V_{VV}^{\frac{1}{2}} d V_{GG}^{-\frac{1}{2}}$, $V_{VV} = var(V_t)$.*

Proof. Results straightforwardly from Assumption 1, see also the Supplementary Appendix. ■

Theorem 2. *Under Assumption 1 and portfolio returns that are generated by (15), the LR-statistic testing the significance of B in (7) converges, when the sample size T goes to infinity, to a non-central χ^2 distributed random variable with Nm degrees of freedom and non-centrality parameter trace($d^{+\prime}d^+$), $d^+ = V_{RR}^{-\frac{1}{2}} \beta d V_{GG}^{\frac{1}{2}}$.*

Proof. Results straightforwardly from Assumption 1, see also the Supplementary Appendix. ■

The large sample properties of the F and LR statistics stated in Theorems 1 and 2 are in line with the realized values of the F and LR statistics stated in Tables 2-4 for all factors except the FF ones. The assumption of weak correlation between observed and unobserved factors made in Assumption 1 is therefore more appropriate for deriving the large sample properties of statistics in the FM two pass approach. This is especially relevant since these properties are considerably different from those derived under the traditional assumption. We focus on one kind of statistics which are commonly used in the FM two pass approach: R^2's.

It is common practice to measure the explanatory power of a regression using a goodness of fit measure like the R^2. Both the OLS and GLS R^2's of the second pass regression of the FM two pass procedure are used for this purpose. We discuss them both and start with the most commonly used one which is the OLS R^2.

OLS R^2. The OLS R^2 equals the explained sum of squares over the total sum of squares when we only use a constant term so its expression reads

$$R^2_{OLS} = \frac{\bar{R}' P_{M_{\iota_N}\hat{B}} \bar{R}}{\bar{R}' M_{\iota_N} \bar{R}} = \frac{\bar{R}' M_{\iota_N} \hat{B}(\hat{B}' M_{\iota_N} \hat{B})^{-1} \hat{B}' M_{\iota_N} \bar{R}}{\bar{R}' M_{\iota_N} \bar{R}}, \tag{18}$$

with $P_A = A(A'A)^{-1}A'$, $M_A = I_N - P_A$ for a full rank matrix A and I_N the $N \times N$ dimensional identity matrix. We analyze the behavior of R^2_{OLS} under the assumption that the observed and unobserved factors are only minorly correlated as stated in Assumption 1.

Theorem 3. *Under Assumption 1, portfolio returns that are generated by (15) and mean returns that are characterized by (11), the behavior of R^2_{OLS} in (18) is in large samples characterized by:*

$$\frac{[\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]' P_{M_{\iota_N}(\beta(d+\psi_{VG})+\psi_{\varepsilon G})} [\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]}{[\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]' M_{\iota_N} [\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]}, \tag{19}$$

where $\psi_{\iota F} = V_{FF}^{\frac{1}{2}}\psi^*_{\iota F}$, $\psi_{\iota\varepsilon} = V_{\varepsilon\varepsilon}^{\frac{1}{2}}\psi^*_{\iota\varepsilon}$, $\psi_{VG} = V_{VV}^{\frac{1}{2}}\psi^*_{VG}V_{GG}^{-\frac{1}{2}}$ and $\psi_{\varepsilon G} = V_{\varepsilon\varepsilon}^{\frac{1}{2}}\psi^*_{\varepsilon G}V_{GG}^{-\frac{1}{2}}$ and $\psi^*_{\iota F}$, $\psi^*_{\iota\varepsilon}$, ψ^*_{VG} and $\psi^*_{\varepsilon G}$ are $k \times 1$, $N \times 1$, $k \times m$ and $N \times m$ dimensional random matrices whose elements are independently standard normally distributed.

Proof. see Appendix. ■

When the correlation between the observed and unobserved factors is large and their number is the same, so d in (17) and (19) is a square invertible matrix and large compared to ψ_{VG} and $\psi_{\varepsilon G}$, R^2_{OLS} is equal to one when the sample size goes to infinity, see also Lewellen *et. al.* (2010).

Corollary 1. *When the number of observed and unobserved factors is the same and they are highly correlated, so d in (19) is a large invertible matrix which is of a larger order of magnitude that ψ_{VG} and $\psi_{\varepsilon G}$, R^2_{OLS} converges to one when the sample size T increases.*

Corollary 1 shows the behavior of R^2_{OLS} under the conventional assumption of a full rank value of the estimand of $\hat{B}$. The R^2_{OLS} is then a consistent estimator of its population value.

Corollary 2. *When the number of observed factors is less than the number of unobserved factors but the observed factors explain the unobserved factors well, so d in (17) is a large full rank rectangular $k \times m$ dimensional matrix with $m < k$, R^2_{OLS} is asymptotically equivalent to*

$$\frac{[\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]' P_{M_{\iota_N}\beta d}[\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]}{[\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]' M_{\iota_N}[\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]}, \tag{20}$$

which converges, when the sample size T goes to infinity, to

$$\frac{\lambda_F'\beta' P_{M_{\iota_N}\beta d}\beta\lambda_F}{\lambda_F'\beta' M_{\iota_N}\beta\lambda_F}. \tag{21}$$

The scenarios stated in Corollaries 1 and 2 are also discussed in Lewellen *et. al.* (2010). The cases for which Lewellen *et. al.* (2010) do not provide any analytical results are those where:

1. the observed factors are only minorly correlated with the unobserved factors and
2. when only a few of the observed factors are strongly correlated with the unobserved factors and the number of correlated observed factors is less than the number of unobserved factors.

These are highly relevant cases since they apply to the (macro-) economic factors discussed previously. It is therefore important to have an analytical expression for the large sample behavior of R^2_{OLS} so we understand where its properties result from.

The first important property Theorem 3 shows is that, under Assumption 1, R^2_{OLS} converges to a random variable. When d is of a larger order of magnitude than the random variables ψ_{VG} and $\psi_{\varepsilon G}$, the latter two do not affect the large sample behavior of R^2_{OLS} so R^2_{OLS} is a consistent estimator of its population value. This results in the behavior stated in Corollaries 1 and 2, see also Lewellen *et. al.* (2010). When d is of a similar order of magnitude than ψ_{VG} and $\psi_{\varepsilon G}$, R^2_{OLS} is, however, no longer a consistent estimator of its population value since it converges to a random variable. Under case 2, the part of R^2_{OLS} associated with the strongly correlated observed factors converges to its population value while the remaining part converges to a random variable. In total, R^2_{OLS} is therefore also not consistent and converges to a random variable.

Corollary 3. *Under Assumption 1 and when only the first m_1 observed factors are strongly correlated with the unobserved factors and m_1 is less than k, so $d = (d_1 \vdots d_2)$, $d_1 : k \times m_1$, $d_2 : k \times m_2$, $m_1 + m_2 = m$, with d_1 large and d_2 small, the large sample behavior of R^2_{OLS} is characterized by*

$$\frac{[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]' P_{M_{\iota_N}\beta d_1}[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]}{[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]' M_{\iota_N}[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]}+ \frac{[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]' P_{M_{(\iota_N \,:\, \beta d_1)}\left(\beta\left(d_2+\psi_{VG,2}\right)+\psi_{\varepsilon G,2}\right)}[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]}{[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]' M_{\iota_N}[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]}, \tag{22}$$

where we use that $P_{(A\,:\,B)} = P_A + P_{M_A B}$ and $\psi_{VG} = (\psi_{VG,1} \vdots \psi_{VG,2})$, $\psi_{\varepsilon V} = (\psi_{\varepsilon V,1} \vdots \psi_{\varepsilon V,2})$ and $\psi_{VG,1} : k \times m_1$, $\psi_{VG,2} : k \times m_2$, $\psi_{\varepsilon V,1} : N \times m_1$, $\psi_{\varepsilon V,2} : N \times m_2$.

Without loss of generality, we have assumed in Corollary 3 that only the first m_1 observed factors are correlated with the unobserved ones. A similar result is obtained when more than m_1 of the observed factors are correlated with the unobserved ones but they are correlated in an identical manner. In that case d_1 would be a matrix which is of reduced rank for which we can adapt the expression in Corollary 3 accordingly.

Corollary 3 shows that the large sample behavior of R^2_{OLS} consists of two components, one which converges to the population value of R^2_{OLS} when we use only those observed factors that are strongly correlated with the unobserved ones and the other random component results from those observed factors that are minorly correlated with the unobserved factors. Hence overall R^2_{OLS} converges to a random variable as well so it is not a consistent estimator of its population value.

Having now established that R^2_{OLS} converges to a random variable in cases which are reminiscent of using (macro-) economic proxy factors other than the FF factors, it is important to establish the behavior of this random variable. The expression of the limiting behavior of R^2_{OLS} is such that only the numerator is random since the denominator of R^2_{OLS} converges to its population value. Theorem 3 shows that the numerator consists of the projection of

$$M_{\iota_N}[\beta\lambda_F + \tfrac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})] \qquad \text{on} \qquad M_{\iota_N}(\beta\,(d + \psi_{VG}) + \psi_{\varepsilon G}).$$

The first element of the part where you project on, *i.e.* $M_{\iota_N}\beta(d+\psi_{VG})$, is tangent to $M_{\iota_N}\beta(\lambda_F + \frac{1}{\sqrt{T}}\psi_{\iota F})$ since both are linear combinations of $M_{\iota_N}\beta$. This implies that the numerator of R^2_{OLS} is big whenever $M_{\iota_N}\beta(d+\psi_{VG})$ is relatively large compared to $M_{\iota_N}\psi_{\varepsilon G}$ regardless of whether this results from a large value of d or not.

When the observed proxy factors G_t explain the unobserved factors well, d is large and V_t is small. When V_t is small, there is no unexplained factor structure in the residuals of (15),

U_t, that results from regressing the portfolio returns on the observed proxy factors. When we use factors other than the FF factors, the F-statistics and pseudo-R^2's, indicated by pseudo-R^2 FF, in Tables 2-4 show that d is small and V_t often explains more than ten times as much of the variation in F_t, measured by pseudo-R^2 FF, than the observed proxy factors G_t. The same reasoning applies when the observed proxy factors include the value weighted return and we consider the increment in the pseudo-R^2 that results from adding observed proxy factors other than the FF factors. Hence for all these observed proxy factors, d is small and V_t is large and causes, since it is multiplied by β, an unexplained factor structure in the residuals of (15). This unexplained factor structure also indicates that βV_t is large compared to ε_t in (15). The weighted averages of these components converge to ψ_{VG} and $\psi_{\varepsilon G}$. The small values of the pseudo-R^2's thus imply that d is small relative to ψ_{VG} while the unexplained factor structure indicates that $\beta\psi_{VG}$ is large relative to $\psi_{\varepsilon G}$. Taken all together this implies that large values of R^2_{OLS} result from the projection of $M_{\iota_N}\beta(\lambda + \frac{1}{\sqrt{T}}\psi_{\iota F})$ on $M_{\iota_N}\beta\psi_{VG}$ since $M_{\iota_N}\beta\psi_{VG}$ is large compared to both $M_{\iota_N}\beta d$ and $M_{\iota_N}\psi_{\varepsilon G}$. Hence, since $\beta\psi_{VG}$ is part of the estimation error of $\hat{B}$, it is the estimation error of $\hat{B}$ that leads to the large values of R^2_{OLS} when d is small. These large values of R^2_{OLS} are therefore not indicative of the strength of the relationship between expected portfolio returns and observed proxy factors.

The same reasoning that applies to R^2_{OLS} in case 1, as described above, holds for case 2 as well. Corollary 3 shows that R^2_{OLS} then converges to the sum of two components. The first of these two components converges to the population value of R^2_{OLS} that results from only using the strongly correlated observed factors. The second component has a similar expression as R^2_{OLS} in case 1. Identical to R^2_{OLS} in case 1, its large values when the observed factors do not explain the unobserved factors therefore result from the estimation error in $\hat{B}$.

The above shows that the unexplained factor structure in the residuals of (15) can lead to large values of the R^2_{OLS} when the observed proxy and unobserved true factors are minorly correlated. We have discussed several statistics, like, for example, F and LR statistics, pseudo-R^2's and our FACCHECK measure, to shed light on the small correlation between observed and unobserved factors. Of all these statistics, FACCHECK (6) directly measures the unexplained factor structure in the residuals or put differently the relative size of $\beta\psi_{VG}$ compared to $\psi_{\varepsilon G}$. Consequently, applying the FACCHECK statistic to the residuals of the observed factor model helps gauge the reliability of R^2_{OLS}. When analyzing twenty five portfolios, a value of FACCHECK of around 0.95, implies that this relative size is around 20, it is around 4 when FACCHECK is 0.8 and around 1.5 when FACCHECK is 0.6. Hence for values of FACCHECK around 0.5-0.6, the influence of the factor structure on R^2_{OLS} is comparable to that of the idiosyncratic components. This would make a sensible rule of thumb for applying FACCHECK to assess the extent to which a large value of R^2_{OLS} is indicative of the strength of the second pass cross sectional regression. When FACCHECK is small, R^2_{OLS} can be straightforwardly interpreted but not so if FACCHECK is large in which case we should interpret it cautiously.

Simulation experiment

We conduct a simulation experiment to further illustrate the properties of R^2_{OLS} and the accuracy of the large sample distribution stated in Theorem 3. Our simulation experiment is calibrated to data from Lettau and Ludvigson (2001). We use the FM two pass procedure to estimate the risk premia on the three FF factors using their returns on twenty-five size and book to market sorted portfolios from 1963 to 1998. We then generate portfolio returns from the factor model in (2), with $\mu = \iota_N\lambda_0 + \beta\lambda_F$, and $E(F_t) = 0$ using the estimated values of β, λ_0 and λ_F as the true values and factors F_t and disturbances ε_t that are generated as i.i.d. normal with mean zero and covariance matrices $\hat{V}_{FF}$ and $\hat{V}_{\varepsilon\varepsilon}$ with $\hat{V}_{FF}$ the covariance matrix of the three FF factors and $\hat{V}_{\varepsilon\varepsilon}$ the residual covariance matrix that results from regressing the portfolio returns on the three FF factors. The number of time series observations is the same as in Lettau and Ludvigson (2001).

We use the simulated portfolio returns to compute the density functions of R^2_{OLS} in (18) using an observed factor G_t that initially only consists of the first (observed) factor, then of the first two factors and then of all three factors. Alongside the density function of R^2_{OLS} that results from simulating from the model, we also use the approximation that results from Theorem 3. Figure 1.1 in Panel 1 shows that the density functions of R^2_{OLS} that result from simulating from the model and from the approximation in Theorem 3 are almost identical. The figures in Panel 1 further show that, as expected, the distribution of R^2_{OLS} moves to the right when we add an additional true factor. Figure 1.1 also shows that R^2_{OLS} is close to one when we use all three factors as stated in Corollary 1.

To show the extent to which the observed factor model explains the factor structure of the portfolio returns, Panel 1 also reports the density function of FACCHECK. Figure 1.2 shows that when we use only one factor, the three largest principal components explain around 81% of the variation which is roughly equal to the 82% that we stated in Tables 3 and 4 when we use the value weighted return as the only factor.[9] The variation explained by the three largest principal components decreases to 58% when we use two factors and 38% when we use all three factors. The last percentage is similar to the percentage in Table 3 when we use all three FF factors.

[9] We note that the Jagannathan-Wang data contains one hundred portfolio returns so the explained percentage of the variation is not comparable with that which results when we use the value weighted return as the only factor for the Jagannathan-Wang data.

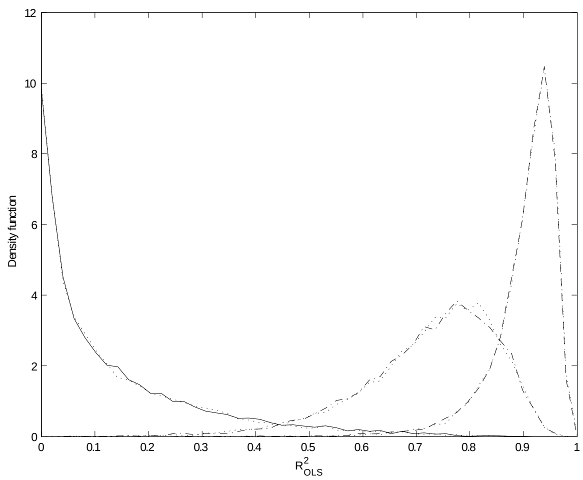

Figure 1.1. Density functions of R^2_{OLS}

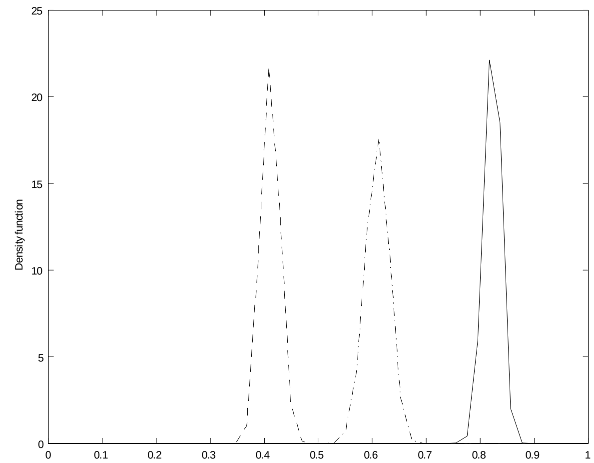

Figure 1.2. Density functions of FACCHECK

Panel 1. Density functions of R^2_{OLS} and FACCHECK (the ratio of the sum of the three largest characteristic roots of the residual covariance matrix over the sum of all characteristic roots) when we use one of the three factors (solid), two (dashed-dotted) and all three (dashed). Figure 1.1 also shows the large sample distributions from Theorem 3 (dotted lines).

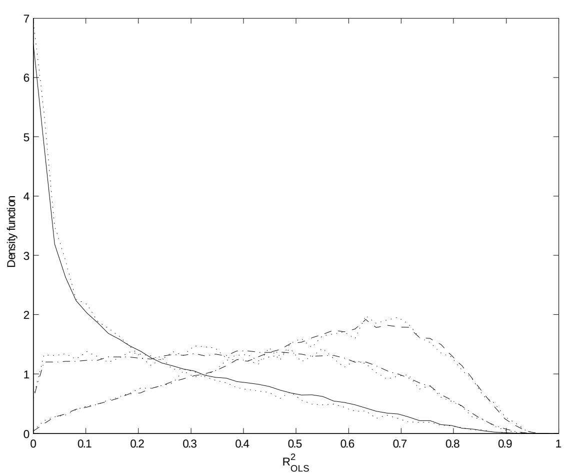

Figure 2.1. Density functions of R^2_{OLS}

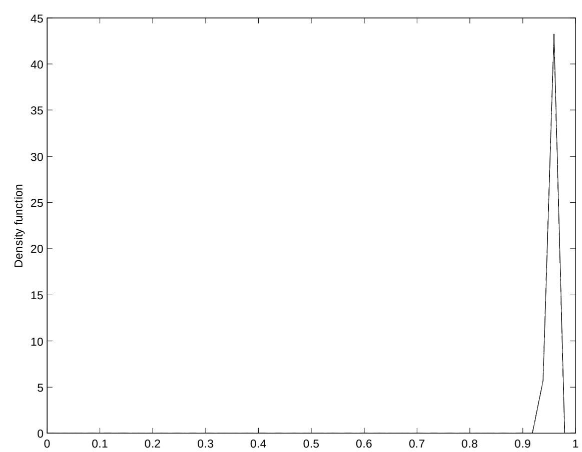

Figure 2.2. Density functions of FACCHECK

Panel 2. Density functions of R^2_{OLS} and FACCHECK (the ratio of the sum of the three largest characteristic roots of the residual covariance matrix over the sum of all characteristic roots) when we use one useless factor (solid), two (dashed-dotted) and three (dashed). Figure 2.1 also shows the large sample distributions from Theorem 3 (dotted lines).

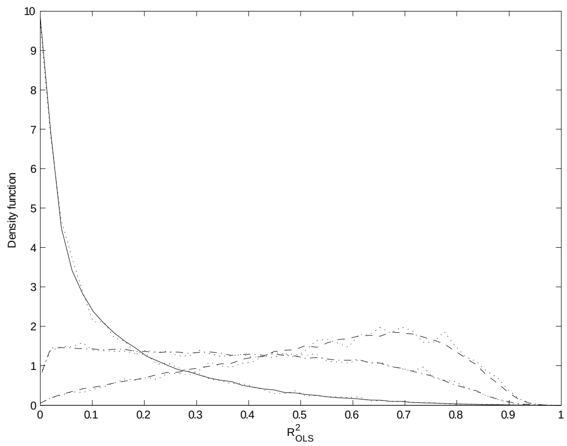

Figure 3.1. Density functions of R^2_{OLS}

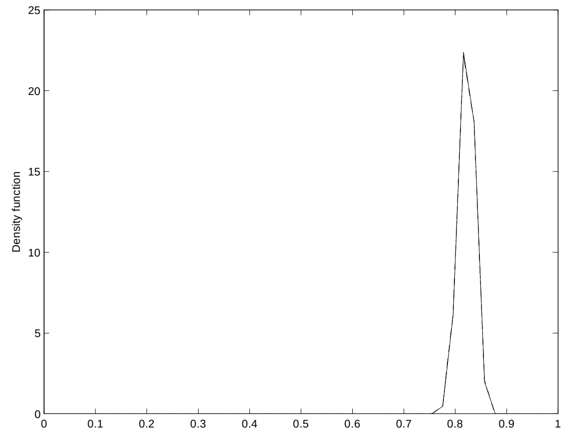

Figure 3.2. Density functions of FACCHECK

Panel 3. Density functions of R^2_{OLS} and FACCHECK (the ratio of the sum of the three largest characteristic roots of the residual covariance matrix over the sum of all characteristic roots) when we use one valid factor (solid), one valid factor and one irrelevant factor (dash-dotted) and one valid factor and two irrelevant factors (dashed). Figure 3.1 also shows the large sample distributions from Theorem 3 (dotted lines).

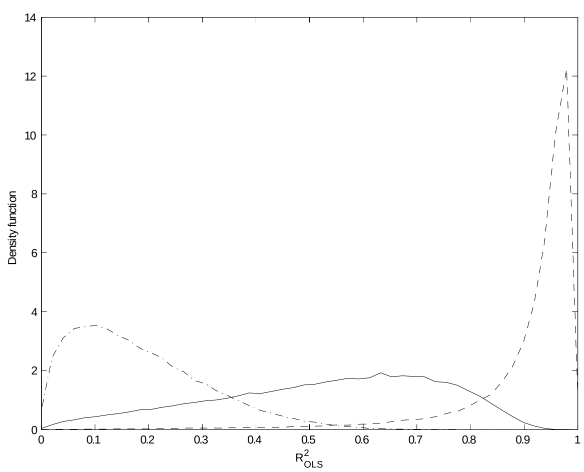

Figure 4.1. Density functions of the R^2

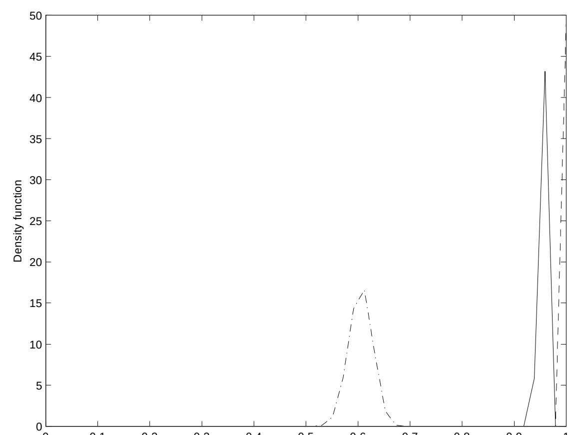

Figure 4.2. Density functions of FACCHECK

Panel 4. Density functions of R^2_{OLS} and FACCHECK (the ratio of the sum of the three largest characteristic roots of the residual covariance matrix over the sum of all characteristic roots) when we use three useless factors and there is a factor structure (solid line), strong factor structure (dashed line) and weak factor structure (dashed-dotted line).

Panel 2 shows the density functions that result from another simulation experiment where we simulate from the same model as used previously but now we estimate an observed factor model with only useless factors. We start out with an observed factor model with one useless factor and then add one or two additional useless factors. Again we obtain virtually the same distributions from simulating from the model and using the approximation from Theorem 3.

The density functions of R^2_{OLS} in Figure 2.1 are surprising. They dominate the distribution of R^2_{OLS} in case we only use one of the true factors. Hence, based on R^2_{OLS}, observed factor models with useless factors outperform an observed factor model which just has one of the three true factors. It is even such that the R^2_{OLS} that results from using three useless factors often exceeds the R^2_{OLS} when we use two valid factors. This becomes even more pronounced when we add more useless factors which we do not show. To reveal that the observed factor models with the useless factors do not explain anything, we also computed the density function of FACCHECK. As expected, its density functions that result from the three specifications with the useless factors all lie on top of one another at 95% which is identical to the value of the ratio in Tables 2 and 4 when the observed factors matter very little.

Similar results are shown in Panel 3 where we use a setting with one valid factor and then add one or two irrelevant factors. The figures in Panel 3 show that the distribution of R^2_{OLS} in case of one valid factor and one or two irrelevant factors is similar to the one that results from two or three irrelevant factors. The main difference between the distributions for these settings occurs for the density of FACCHECK which shows that the unexplained factor structure in Panel 3 is less pronounced than in Panel 2.

The expression of the large sample distribution of R^2_{OLS} in Theorem 3 states the importance of the unexplained factor structure for R^2_{OLS}. This is further shown by the simulation results in Panels 1-3. It all shows that R^2_{OLS} cannot be interpreted appropriately without some diagnostic statistic that reports on the unexplained factor structure. Hence, R^2_{OLS} is only indicative for a relationship between portfolio returns and the observed factors when there is no unexplained factor structure in the residuals. To further emphasize this, we conduct another simulation experiment where we specifically analyze the influence of the unexplained factor structure. We therefore estimate an observed factor model that has three useless factors. To show the sensitivity of R^2_{OLS} to the unexplained factor structure, we simulate from the same model as used previously but we now use three different settings of the covariance matrix $V_{\varepsilon\varepsilon}$ of the disturbances in the original factor model: $V_{\varepsilon\varepsilon} = 25\hat{V}_{\varepsilon\varepsilon}$ (weak factor structure), $V_{\varepsilon\varepsilon} = \hat{V}_{\varepsilon\varepsilon}$ (factor structure) and $V_{\varepsilon\varepsilon} = 0.04\hat{V}_{\varepsilon\varepsilon}$ (strong factor structure) with $\hat{V}_{\varepsilon\varepsilon}$ the residual covariance matrix that results from regressing the portfolio returns on the three FF factors. No changes are made to the specification of the risk premia or the β's so the factor pricing in the model where we simulate from remains unaltered except for the covariance matrix of the disturbances. The results are reported in Panel 4.

The figures in Panel 4 reiterate the sensitivity of the distribution of R^2_{OLS} to the unexplained factor structure in the residuals. Figure 4.1 shows that for the same irrelevant explanatory power of the observed factor model, R^2_{OLS} varies greatly. Figure 4.2 shows that for the observed factor models where R^2_{OLS} is high in Figure 4.1 also the unexplained factor structure in the residuals is very strong. For the observed factor model where the factor structure in the residuals is rather mild, the density of R^2_{OLS} is as expected and close to zero. Hence, for the models where there is still a strong unexplained factor structure in the residuals, R^2_{OLS} is not indicative of a relationship between expected portfolio returns and the observed factors.

	LL01					
	R_{vw}	Δc	FF factors	$cay, R_{vw},$ $cayR_{vw}$	$cay, \Delta c,$ $cay\Delta c$	$cay, R_{vw}, \Delta y,$ $cayR_{vw}, cay\Delta y$
R^2_{OLS}	0.01	0.16	0.80	0.31	0.70	0.77
FACCHECK	82.1%	95.5%	38.2%	82.5%	95.2%	82.1%
pseudo-R^2	0.78	0.016	0.95	0.78	0.10	0.79

Table 5: R-squared of the second pass regression of the FM two pass procedure, FACCHECK (the percentage of the variation explained by the three largest principal components) and the pseudo-R-squared for different specifications from Lettau and Ludvigson (2001).

	F52-01						
	R_{vw}	Δc	FINAN	LVX06	LN05	SV06	Y06
R^2_{OLS}	0.07	0.04	0.51	0.58	0.74	0.65	0.54
FACCHECK	81.4%	93.9%	94.3%	94.3%	93.8%	80.5%	81.5%
pseudo-R^2	0.723	0.065	0.007	0.015	0.083	0.739	0.724

Table 6: R-squared of the second pass regression of the FM two pass procedure, FACCHECK (the percentage of the variation explained by the three largest principal components) and pseudo-R-squared using the factors from Li et. al. (2006) (LVX06), Lustig and Van Nieuwerburgh (2005) (LN05), Santos and Veronesi (2006) (SV06) and Yogo (2006) (Y06). All use the quarterly portfolio returns from French's website.

Tables 5 and 6 report R^2_{OLS}, FACCHECK and pseudo-R^2 for the specifications in Tables 3 and 4. Many of the specifications stated in Tables 5 and 6 have high values of R^2_{OLS}. Except for the specification using the FF factors, all of these specifications also have large values of the factor structure check, which indicates that there is an unexplained factor structure in the first pass residuals, and small values of the pseudo-R^2's in Tables 2-4 which indicate a small value of d. We just showed that R^2_{OLS} is then not indicative of a relationship between expected portfolio returns and observed factors since these large values result from the estimation error in the estimated β's of the observed proxy factors. Tables 5 and 6 correspond with Lettau and Ludvigson (2001), Li *et. al.* (2006), Lustig and Van Nieuwerburgh (2006), Santos and Veronesi

(2006) and Yogo (2006), so the reported R^2_{OLS}'s are not indicative of a relationship between expected portfolio returns and observed proxy factors.

GLS R^2. The GLS R^2 equals the explained sum of squares over the total sum of squares in a GLS regression where we weight by the inverse of the covariance matrix of $\bar{R}$:

$$\begin{aligned} R^2_{GLS} &= \frac{\bar{R}'\bar{M}\hat{B}(\hat{B}'\bar{M}\hat{B})^{-1}\hat{B}'\bar{M}\bar{R}}{\bar{R}'\bar{M}\bar{R}} \\ &= \frac{(V_{RR}^{-\frac{1}{2}}\bar{R})'P_{M_{V_{RR}^{-\frac{1}{2}}\iota_N}V_{RR}^{-\frac{1}{2}}\hat{B}}(V_{RR}^{-\frac{1}{2}}\bar{R})}{(V_{RR}^{-\frac{1}{2}}\bar{R})'M_{V_{RR}^{-\frac{1}{2}}\iota_N}(V_{RR}^{-\frac{1}{2}}\bar{R})}, \end{aligned} \tag{23}$$

with $\bar{M} = V_{RR}^{-1} - V_{RR}^{-1}\iota_N(\iota_N'V_{RR}^{-1}\iota_N)^{-1}\iota_N'V_{RR}^{-1}$.

Under the conventional assumption of a full rank value of the estimand of $\hat{B}$, R^2_{OLS} is a consistent estimator of its population value. For many observed proxy factors, this assumption is not realistic. To accommodate such instances, we made Assumption 1 using which Theorem 3 shows that the R^2_{OLS} then converges to a random variable. Alongside the explanatory power of the observed proxy factors for the unobserved factors, the large sample behavior of R^2_{GLS} crucially depends on the scaled risk premia on the unobserved factors, $\left(\frac{V_{FF}}{T}\right)^{-\frac{1}{2}}\lambda_F$. We assume that these relative risk premia do not change with the sample size which is in line with their relatively small values reported in Table 7. When we do not make this assumption and just assume that the risk premia are constant, the R^2_{GLS} always converges to one when the sample size increases which we deem unrealistic.

Assumption 2. *The scaled risk premia $\left(\frac{V_{FF}}{T}\right)^{-\frac{1}{2}}\lambda_F$ remain constant when the sample size increases so*

$$\left(\frac{V_{FF}}{T}\right)^{-\frac{1}{2}}\lambda_F = l, \tag{24}$$

with l a k dimensional fixed vector, for different values of the sample size T.

	LL01			JW96			F52-01		
$\lambda_F'\left(\frac{V_{FF}}{T}\right)^{-1}\lambda_F$	24.1			10.1			15.2		
	λ_F	$V_{FF}^{-\frac{1}{2}}\lambda_F$	$\left(\frac{V_{FF}}{T}\right)^{-\frac{1}{2}}\lambda_F$	λ_F	$V_{FF}^{-\frac{1}{2}}\lambda_F$	$\left(\frac{V_{FF}}{T}\right)^{-\frac{1}{2}}\lambda_F$	λ_F	$V_{FF}^{-\frac{1}{2}}\lambda_F$	$\left(\frac{V_{FF}}{T}\right)^{-\frac{1}{2}}\lambda_F$
R_{VW}	1.32	0.22	2.62	-0.51	-0.12	-2.21	-1.12	-0.13	-1.84
SMB	0.47	0.024	0.28	0.21	0.10	1.86	0.47	0.12	1.78
HML	1.46	0.35	4.14	0.24	0.072	1.30	1.32	0.21	2.94

Table 7: Estimates of the risk premia and scaled risk premia that result from the FM two pass procedure for data from Lettau-Lutvigson (2001) (LL01), Jaganathan and Wang (1996) (JW96) and from French's website (F52-01).

Theorem 4. *Under Assumptions 1, 2, portfolio returns that are generated by (15) and mean returns on the portfolios that are characterized by (11), the behavior of R^2_{GLS} in (23) is in large samples characterized by:*

$$\frac{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' P_{M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}dV_{GG}^{\frac{1}{2}} \\ 0 \end{pmatrix}+\varphi^*\right\}}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}, \tag{25}$$

where ψ^* and φ^* are independent $N \times 1$ and $N \times m$ dimensional random matrices whose elements have independent standard normal distributions, W is an orthonormal $k \times k$ dimensional matrix which contains the eigenvectors of

$$\left[(\beta'\beta)^{\frac{1}{2}\prime}V_{FF}(\beta'\beta)^{\frac{1}{2}} + (\beta'\beta)^{-\frac{1}{2}\prime}\beta'V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-\frac{1}{2}}\right] \tag{26}$$

and $M_{V_{RR}^{-\frac{1}{2}}\iota_N}$ is characterized by

$$\begin{aligned} &I_N - \begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta'\iota_N \\ (\beta_\perp'V_{\varepsilon\varepsilon}\beta_\perp)^{-\frac{1}{2}}\beta_\perp'\iota_N \end{pmatrix} \\ &\{\iota_N'[\beta(\beta'\beta)^{-1}V_{FF}^{-1}(\beta'\beta)^{-1}\beta' + \beta_\perp(\beta_\perp'V_{\varepsilon\varepsilon}\beta_\perp)^{-1}\beta_\perp']\iota_N\}^{-1}\begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta'\iota_N \\ (\beta_\perp'V_{\varepsilon\varepsilon}\beta_\perp)^{-\frac{1}{2}}\beta_\perp'\iota_N \end{pmatrix}', \end{aligned} \tag{27}$$

with $\beta_\perp$ the $N \times (N-k)$ dimensional orthogonal complement of β, so $\beta_\perp'\beta \equiv 0$, $\beta_\perp'\beta_\perp \equiv I_{N-k}$.

Proof. see Appendix. ■

We use Theorem 4 to classify the different kinds of behavior of R^2_{GLS}. We start with a strong observed proxy factor setting.

Corollary 4. *When the number of observed factors equals the number of unobserved factors and they explain them well, so $d = \sqrt{T}V_{FG}V_{GG}^{-1}$, the large sample behavior of R^2_{GLS} is characterized by*

$$\frac{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}' P_{M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}V_{FG}V_{GG}^{-\frac{1}{2}} \\ 0 \end{pmatrix} + \frac{1}{\sqrt{T}}V_{RR}^{-\frac{1}{2}}V_{\varepsilon\varepsilon}^{\frac{1}{2}}\varphi^*\right\}}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}' M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}}. \tag{28}$$

Furthermore, when the observed factors are an invertible linear combination of the true factors, $V_{FF}^{-\frac{1}{2}}V_{FG}V_{GG}^{-\frac{1}{2}} = I_k$.

The large sample behavior of R^2_{GLS} in Corollary 4 differs considerably from that of R^2_{OLS}. Corollary 1 states that R^2_{OLS} converges to one when the observed factors explain the unobserved factors well and their numbers are the same. Because $W'l$ is of the same order of magnitude as the standard normal random variables in ψ^*, this is not the case for R^2_{GLS}. Only when the scaled risk premia are very large, R^2_{GLS} is approximately equal to one.

Corollary 5. *When the relative size of the risk premia is very large and the number of observed factors equals the number of unobserved factors and they explain them well,* R^2_{GLS} *is approximately equal to one.*

Another interesting aspect of the large sample distribution of R^2_{GLS} is that it depends on the number of portfolios N. For the same values of the other parameters, a larger number of portfolios implies a smaller value of R^2_{GLS}.

Corollary 6. *When the observed factors consist of the first* m *of the true factors, the large sample behavior of* R^2_{GLS} *is characterized by*

$$\frac{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}' P_{M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'\begin{pmatrix} I_m \\ \phi^*_{VG}/\sqrt{T} \end{pmatrix} \\ 0 \end{pmatrix} + \frac{1}{\sqrt{T}}V_{RR}^{-\frac{1}{2}}V_{\varepsilon\varepsilon}^{\frac{1}{2}}\varphi^*\right\}}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}' M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}}, \tag{29}$$

with ϕ^*_{VG} *a* $(k-m)\times m$ *dimensional random matrix whose elements are standard normally distributed and independent of* φ^* *and* ψ^**.*

Corollary 6 shows that when the observed factors explain fewer of the true factors, the R^2_{GLS} goes down on average. This argument extends to the case where the relative risk premia are large.

Corollary 7. *When the observed factors consist of the first m of the true factors and the relative size of the risk premia is large, R^2_{GLS} converges to*

$$\frac{\begin{pmatrix} W'l \\ 0 \end{pmatrix}' P_{V_{RR}^{-\frac{1}{2}}\iota_N\left\{\begin{pmatrix} W'\begin{pmatrix} I_m \\ 0 \end{pmatrix} \\ 0 \end{pmatrix}\right\}} \begin{pmatrix} W'l \\ 0 \end{pmatrix}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}\right\}' M_{V_{RR}^{-\frac{1}{2}}\iota_N} \left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}\right\}}. \tag{30}$$

Simulation experiment

We use our previous simulation experiment, calibrated to data from Lettau and Ludvigson (2001), to further illustrate the properties of R^2_{GLS} and the accuracy of the large sample distribution stated in Theorem 4. Panel 5 contains the density function of R^2_{GLS} for different settings of the explanatory power of the observed proxy factors and the size of the relative risk premia.

Figures 5.1 and 5.3 use the data generating process that corresponds with the estimated factor model which uses the three FF factors and their risk premia. The observed proxy factors in Figure 5.1 correspond with the true ones while they are irrelevant in Figure 5.3. Figures 5.2 and 5.4 use the settings as used for Figures 5.1 and 5.3 except that the risk premia are ten times as large. The observed proxy factors used for Figure 5.2 correspond with the true ones while the observed proxy factors used for Figure 5.4 are irrelevant.

Figure 5.1 shows the density function of R^2_{GLS} when we use one, two or three of the true factors. We compute these three density functions by simulating from the model and using the large sample approximation stated in Theorem 4. When we use one or all three of the true factors as proxy factors, the resulting density functions are almost indistinguishable. When we just use two of the true factors as proxy factors, there is some discrepancy between the density function which results from simulation and the one which results from the large sample approximation. It shows that the approximation by the large sample distribution is less accurate compared to the one for R^2_{OLS}. This was to be expected because of the inversion of the $N \times N$ dimensional covariance matrix of the portfolio returns which is also, given the factor structure, badly scaled. The large sample approximation remains quite accurate though and is also important since it reveals the dependence of R^2_{GLS} on the scaled risk premia (24).

The density functions show that R^2_{GLS} is well below one even if we use all three factors. When we use only one or two of the three factors, R^2_{GLS} is close to zero. This all results from the small size of the relative risk premia. When we multiply these risk premia by ten as in Figure 5.2, the density of R^2_{GLS} when we use all three factors is close to one.

Figures 5.3 and 5.4 show the density of R^2_{GLS} when we use one, two or three useless factors. Figure 5.3 uses the setting where the risk premia correspond with those from Lettau and

Ludvigson (2001) and Figure 5.4 uses risk premia which are ten fold the estimated ones. Unlike when we use the true factors, the larger risk premia have no effect on the density of R^2_{GLS}.

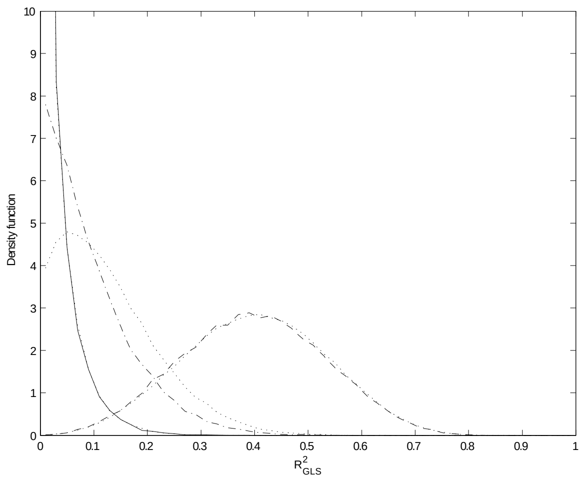

Figure 5.1. True factors, standard premia

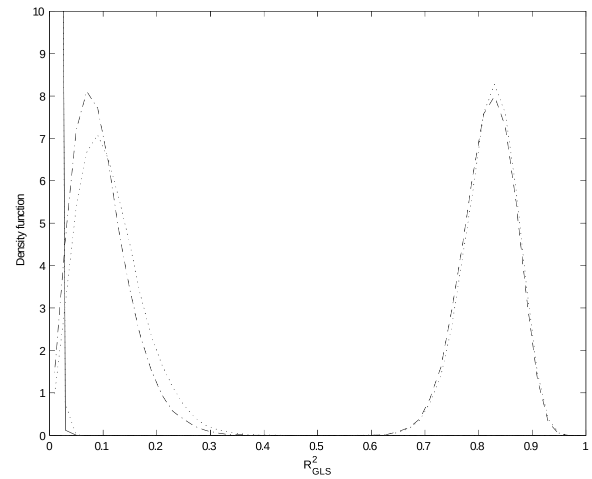

Figure 5.2. True factors, large premia

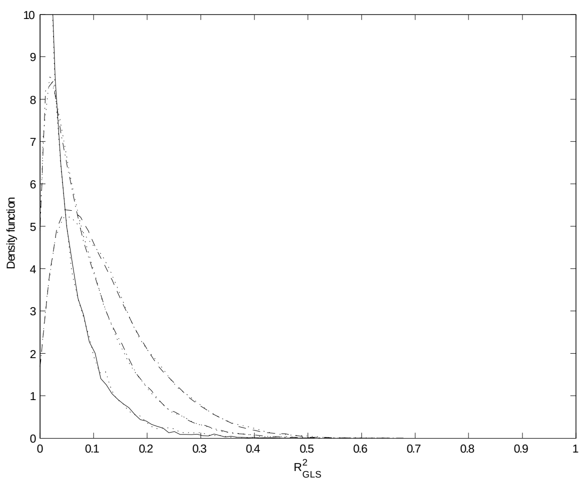

Figure 5.3. Irrelevant factors, standard premia

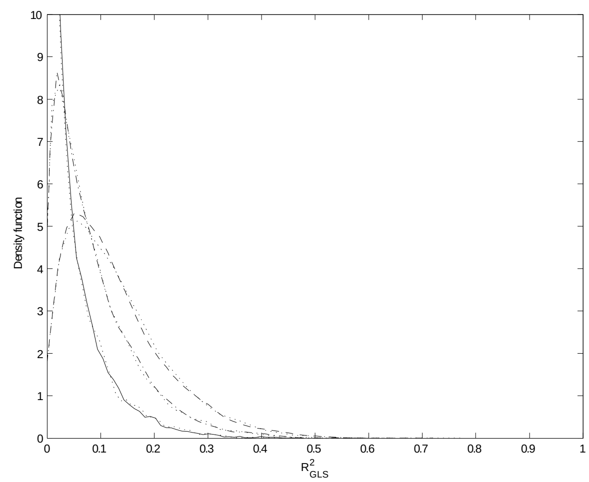

Figure 5.4. Irrelevant factors, large premia

Panel 5. Density of R^2_{GLS} for simulation experiment calibrated to Lettau and Ludvigson (2001). One factor (solid), two factors (dash-dot), three factors (dashed). The dotted lines result from the large sample approximation from Theorem 4.

The density of R^2_{GLS} when we use one, two or three useless factors all lie quite close to zero. However, the density of R^2_{GLS} when we use all three true factors does not lie far from zero either. The density of R^2_{GLS} when we use three useless factors therefore has a lot of probability mass in the area where the density of R^2_{GLS} when we use the three true factors has a sizeable probability mass. This implies that we just based on R^2_{GLS} cannot make a trustworthy statement about the quality of the second pass regression. Identical to R^2_{OLS}, we can use a measure which indicates the unexplained factor structure in the first pass residuals to assess R^2_{GLS} more decisively.

The large sample distributions of R^2_{OLS} and R^2_{GLS} stated in Theorems 3 and 4 depend on the parameters d and l. When these parameters are small, as is the case for all observed proxy factors different from the FF factors, the large sample distributions of R^2_{OLS} and R^2_{GLS} are not normal. Since we cannot estimate d and l consistently, it is then not possible to conduct reliable inference on R^2_{OLS} and R^2_{GLS}, for example, using t-tests as in Kan *et. al.* (2013). Also the bootstrap is only valid when d and l can be estimated consistently so it cannot be applied either.

4 Conclusions

The results from the R^2_{OLS} and the FM t-statistic can line up nicely in favor of a hypothesized factor pricing relationship despite that such a relationship is absent. These statistics can generate such results when the observed proxy factors do not capture the factor structure in portfolio returns. The remaining factor structure in the first pass residuals can then lead to a large value of the R^2_{OLS} while the standard limiting distribution of the FM t-statistic does not apply because of the small correlation between the observed proxy factor and the unobserved factors, see Kleibergen (2009).

To gauge the adequacy of the R^2_{OLS}, we propose to measure the unexplained factor structure in the first pass residuals. When such a factor structure is absent, we can straightforwardly interpret the R^2_{OLS} but we have to do so carefully if this is not the case.

Many observed proxy factors proposed in the literature, like, for example, consumption and labor income growth, housing collateral, consumption-wealth ratio, labor income-consumption ratio, interactions of either one of the latter three with other factors, etc., leave a considerable unexplained factor structure in the first pass residuals. The high R^2's and significant t-statistics that are reported for these factors therefore have to be interpreted judiciously.

Previously suggested solutions to the inferential issues with second pass R^2's and t-statistics do not work well for different reasons. One suggestion is to use the bootstrap. The bootstrap, however, relies on consistent estimation of the risk premia of the observed proxy factors. It fails therefore for the same reason as why the large sample distribution of the second pass t-statistic no longer applies. Another suggestion is to add other portfolios to the typically used pool. Although this reduces the factor structure, a sizeable factor structure typically remains.

Appendix

Proof of Theorem 3. The expression of R^2_{OLS} :

$$R^2_{OLS} = \frac{\bar{R}' M_{\iota_N} \hat{B}(\hat{B}' M_{\iota_N} \hat{B})^{-1} \hat{B}' M_{\iota_N} \bar{R}}{\bar{R}' M_{\iota_N} \bar{R}},$$

shows that it is a function of $\bar{R}$ and $\hat{B}$. To construct the large sample behavior of $\hat{B}$:

$$\hat{B} = \sum_{t=1}^T \bar{R}_t \bar{G}_t' \left(\sum_{t=1}^T \bar{G}_t \bar{G}_t' \right)^{-1}.$$

we use that under the models in (2), (14) and Assumption 1, we can specify it as

$$\begin{aligned} \hat{B} = & \sum_{t=1}^T \left(\beta(\tfrac{d}{\sqrt{T}} \bar{G}_t + \bar{V}_t) + \bar{\varepsilon}_t \right) \bar{G}_t' \left(\sum_{t=1}^T \bar{G}_t \bar{G}_t' \right)^{-1} \\ = & \tfrac{1}{\sqrt{T}} \left[\beta \left(d(\tfrac{1}{T} \sum_{t=1}^T \bar{G}_t \bar{G}_t') + \tfrac{1}{\sqrt{T}} \sum_{t=1}^T \bar{V}_t \bar{G}_t' \right) + \tfrac{1}{\sqrt{T}} \sum_{t=1}^T \bar{\varepsilon}_t \bar{G}_t' \right] \left(\tfrac{1}{T} \sum_{t=1}^T \bar{G}_t \bar{G}_t' \right)^{-1}. \end{aligned}$$

We now use that $\frac{1}{T} \sum_{t=1}^T \bar{G}_t \bar{G}_t' \underset{p}{\to} V_{GG}$, $\frac{1}{\sqrt{T}} \sum_{t=1}^T \bar{V}_t \bar{G}_t' \left(\frac{1}{T} \sum_{t=1}^T \bar{G}_t \bar{G}_t' \right)^{-1} \underset{d}{\to} \psi_{VG} = V_{VV}^{\frac{1}{2}} \psi_{VG}^* V_{GG}^{-\frac{1}{2}}$, $\frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t \bar{G}_t' \left(\frac{1}{T} \sum_{t=1}^T \bar{G}_t \bar{G}_t' \right)^{-1} \underset{d}{\to} \psi_{\varepsilon G} = V_{\varepsilon\varepsilon}^{\frac{1}{2}} \psi_{\varepsilon G}^* V_{GG}^{-\frac{1}{2}}$, and ψ_{VG}^* and $\psi_{\varepsilon G}^*$ are independent $k \times m$ and $N \times m$ dimensional random variables whose elements are independently standard normally distributed. ψ_{VG} and $\psi_{\varepsilon G}$ are independent since F_t and ε_t are uncorrelated so the same applies for V_t then as well since it is an element of F_t. Combining all elements, we obtain the limiting behavior of $\hat{B}$:

$$\sqrt{T} \hat{B} \underset{d}{\to} \beta \left(d + \psi_{VG} \right) + \psi_{\varepsilon G}.$$

The independent large sample behavior of $\bar{R}$ is characterized by (the asymptotic independence of $\bar{R}$ and $\hat{B}$ is shown in Shanken (1992) and Kleibergen (2009))

$$\begin{aligned} \bar{R} = & \tfrac{1}{T} \sum_{t=1}^T \mu_R + \beta F_t + \varepsilon_t \\ = & \tfrac{1}{T} \sum_{t=1}^T (\mu_R + \beta \mu_F) + \beta(F_t - \mu_F) + \varepsilon_t \end{aligned}$$

with $E(\bar{R}) = \mu_R + \beta \mu_F = \iota_N \lambda_0 + \beta \lambda_F$ as stated in (11) so

$$M_{\iota_N} \bar{R} = \tfrac{1}{T} \sum_{t=1}^T M_{\iota_N} (\beta \lambda_F + \beta(F_t - \mu_F)) + \tfrac{1}{T} \sum_{t=1}^T M_{\iota_N} \varepsilon_t$$

and

$$\sqrt{T}(M_{\iota_N} \bar{R} - M_{\iota_N} \beta \lambda_F) \underset{d}{\to} M_{\iota_N} \beta \psi_{\iota F} + M_{\iota_N} \psi_{\iota\varepsilon},$$

where $\frac{1}{\sqrt{T}} \sum_{t=1}^T (F_t - \mu_F) \underset{d}{\to} \psi_{\iota F}$ and $\frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t \underset{d}{\to} \psi_{\iota\varepsilon}$ with $\psi_{\iota F}$ and $\psi_{\iota\varepsilon}$ independently normally distributed k and N dimensional random vectors with mean 0 and covariance matrices V_{FF} and

$V_{\varepsilon\varepsilon}$ which are independent of ψ_{VG} and $\psi_{\varepsilon G}$ as well, or

$$M_{\iota_N}\bar{R} = \quad M_{\iota_N}\beta\lambda_F + \tfrac{1}{\sqrt{T}}(M_{\iota_N}\beta\psi_{\iota F} + M_{\iota_N}\psi_{\iota\varepsilon}) + O_p(T^{-1}).$$

We insert the expressions of the large sample behaviors of $M_{\iota_N}\bar{R}$ and $\hat{B}$ into the expression of R^2_{OLS} to obtain its large sample behavior:

$$\frac{[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]' P_{M_{\iota_N}(\beta(d+\psi_{VG})+\psi_{\varepsilon G})}[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]}{[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]' M_{\iota_N}[\beta\lambda_F+\frac{1}{\sqrt{T}}(\beta\psi_{\iota F}+\psi_{\iota\varepsilon})]}.$$

Proof of Theorem 4. The spectral decomposition of the covariance matrix of the portfolio returns in (5) can be specified as

$$V_{RR} = \quad P_1\Lambda_1 P_1' + P_2\Lambda_2 P_2',$$

with Λ_1 and Λ_2 the $k\times k$ and $(N-k)\times(N-k)$ diagonal matrices that hold respectively the largest k and smallest $N-k$ characteristic roots. The orthonormal $N\times k$ and $N\times(N-k)$ dimensional matrices P_1 and P_2 contain the principal components/eigenvectors. Because of the factor structure,

$$\begin{aligned} P_1 = &\quad \beta Q \\ P_2 = &\quad \beta_\perp, \end{aligned}$$

with $\beta_\perp$ the $N\times(N-k)$ dimensional orthogonal complement of β, so $\beta_\perp'\beta\equiv 0$, $\beta_\perp'\beta_\perp\equiv I_{N-k}$, and Q is a $k\times k$ dimensional matrix which makes P_1 orthonormal, so

$$\begin{gathered} Q = (\beta'\beta)^{-\frac{1}{2}}W \\ W'\left[(\beta'\beta)^{\frac{1}{2}\prime}V_{FF}(\beta'\beta)^{\frac{1}{2}} + (\beta'\beta)^{-\frac{1}{2}\prime}\beta' V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-\frac{1}{2}}\right]W = \Lambda_1, \end{gathered}$$

with W an orthonormal $k\times k$ dimensional matrix. We use the spectral decomposition of V_{RR} to construct the inverse of its square root, so $V_{RR}^{-\frac{1}{2}}V_{RR}V_{RR}^{-\frac{1}{2}\prime} = I_N$:

$$\begin{aligned} V_{RR}^{-\frac{1}{2}} = &\begin{pmatrix} \Lambda_1^{-\frac{1}{2}}P_1' \\ \Lambda_2^{-\frac{1}{2}}P_2' \end{pmatrix} \\ = &\begin{pmatrix} W'\left[(\beta'\beta)^{\frac{1}{2}\prime}V_{FF}(\beta'\beta)^{\frac{1}{2}} + (\beta'\beta)^{-\frac{1}{2}\prime}\beta' V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-\frac{1}{2}}\right]^{-\frac{1}{2}}(\beta'\beta)^{-\frac{1}{2}}\beta' \\ \Lambda_2^{-\frac{1}{2}}\beta_\perp' \end{pmatrix} \\ = &\begin{pmatrix} W'\left[V_{FF} + (\beta'\beta)^{-1\prime}\beta' V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-1}\right]^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta' \\ \Lambda_2^{-\frac{1}{2}}\beta_\perp' \end{pmatrix}. \end{aligned}$$

We can further approximate $[V_{FF}+(\beta'\beta)^{-1\prime}\beta' V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-1}]^{-\frac{1}{2}}$ by

$$V_{FF}^{-\frac{1}{2}}-\tfrac{1}{2}V_{FF}^{-1}(\beta'\beta)^{-1\prime}\beta' V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-1}V_{FF}^{-\frac{1}{2}}$$

which results from a first order Taylor approximation. Because of the factor structure, the second component of the approximation of $[V_{FF}+(\beta'\beta)^{-1\prime}\beta' V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-1}]^{-\frac{1}{2}}$ is much smaller than the first component and we can approximate $[V_{FF}+(\beta'\beta)^{-1\prime}\beta' V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-1}]^{-\frac{1}{2}}$ by $V_{FF}^{-\frac{1}{2}}$.

To construct the large sample behavior of $M_{V_{RR}^{-\frac{1}{2}}\iota_N} V_{RR}^{-\frac{1}{2}}\bar{R}$ and $M_{V_{RR}^{-\frac{1}{2}}\iota_N} V_{RR}^{-\frac{1}{2}}\hat{B}$, we first construct the large sample expressions for $V_{RR}^{-\frac{1}{2}}\beta V_{FF}^{\frac{1}{2}}$ and $M_{V_{RR}^{-\frac{1}{2}}\iota_N}$:

$$\begin{aligned}
V_{RR}^{-\frac{1}{2}}\beta V_{FF}^{\frac{1}{2}} &= \begin{pmatrix} W'\left[V_{FF}+(\beta'\beta)^{-1\prime}\beta' V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-1}\right]^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta' \\ \Lambda_2^{-\frac{1}{2}}\beta_\perp' \end{pmatrix}\beta V_{FF}^{\frac{1}{2}} \\
&= \begin{pmatrix} W'\left[V_{FF}+(\beta'\beta)^{-1\prime}\beta' V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-1}\right]^{-\frac{1}{2}} V_{FF}^{\frac{1}{2}} \\ 0 \end{pmatrix} \\
&\approx \begin{pmatrix} W'[V_{FF}^{-\frac{1}{2}}]V_{FF}^{\frac{1}{2}} \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} W' \\ 0 \end{pmatrix}.
\end{aligned}$$

To obtain $M_{V_{RR}^{-\frac{1}{2}}\iota_N}$, we note that $V_{RR}^{-\frac{1}{2}}\iota_N$, with ι_N an M-dimensional vector of ones, reads:

$$V_{RR}^{-\frac{1}{2}}\iota_N \approx \begin{pmatrix} W' V_{FF}^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta'\iota_N \\ (\beta_\perp' V_{\varepsilon\varepsilon}\beta_\perp)^{-\frac{1}{2}}\beta_\perp'\iota_N \end{pmatrix}$$

so $M_{V_{RR}^{-\frac{1}{2}}\iota_N}$ is characterized by

$$\begin{aligned}
& I_N - \begin{pmatrix} W' V_{FF}^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta'\iota_N \\ (\beta_\perp' V_{\varepsilon\varepsilon}\beta_\perp)^{-\frac{1}{2}}\beta_\perp'\iota_N \end{pmatrix} \\
& \{\iota_N'[\beta(\beta'\beta)^{-1}V_{FF}^{-1}(\beta'\beta)^{-1}\beta' + \beta_\perp(\beta_\perp' V_{\varepsilon\varepsilon}\beta_\perp)^{-1}\beta_\perp']\iota_N\}^{-1} \begin{pmatrix} W' V_{FF}^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta'\iota_N \\ (\beta_\perp' V_{\varepsilon\varepsilon}\beta_\perp)^{-\frac{1}{2}}\beta_\perp'\iota_N \end{pmatrix}'.
\end{aligned}$$

The specification of GLS R^2 reads

$$R^2_{GLS} = \frac{(M_{V_{RR}^{-\frac{1}{2}}\iota_N} V_{RR}^{-\frac{1}{2}}\bar{R})' P_{M_{V_{RR}^{-\frac{1}{2}}\iota_N} V_{RR}^{-\frac{1}{2}}\hat{B}} (M_{V_{RR}^{-\frac{1}{2}}\iota_N} V_{RR}^{-\frac{1}{2}}\bar{R})}{(V_{RR}^{-\frac{1}{2}}\bar{R})' M_{V_{RR}^{-\frac{1}{2}}\iota_N} (V_{RR}^{-\frac{1}{2}}\bar{R})}.$$

We proceed with constructing expressions for the large sample behavior of the components of the GLS R^2 : $V_{RR}^{-\frac{1}{2}}\bar{R}$ and $V_{RR}^{-\frac{1}{2}}\hat{B}$ where we use both strong and weak factor settings for the latter.

$\mathbf{V}_{RR}^{-\frac{1}{2}}\bar{\mathbf{R}}$. The large sample behavior of $\bar{R}$ is constructed in the proof of Theorem 3:

$$\begin{aligned}\bar{R} = &\ \iota_N\lambda_0 + \beta(\lambda_F + \tfrac{1}{\sqrt{T}}\psi_{\iota F}) + \tfrac{1}{\sqrt{T}}\psi_{\iota\varepsilon} + O_p(T^{-1}) \\ = &\ \iota_N\lambda_0 + \beta V_{FF}(V_{FF}^{-\frac{1}{2}}\lambda_F) + \tfrac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon}) + O_p(T^{-1})\end{aligned}$$

Under Assumption 2, $l = V_{FF}^{-\frac{1}{2}}\lambda_F\sqrt{T}$ is constant and we can specify the large sample behavior of $\sqrt{T}M_{V_{RR}^{-\frac{1}{2}}\iota_N}V_{RR}^{-\frac{1}{2}}\bar{R}$ as:

$$\begin{aligned}\sqrt{T}M_{V_{RR}^{-\frac{1}{2}}\iota_N}V_{RR}^{-\frac{1}{2}}\bar{R} = &\ M_{V_{RR}^{-\frac{1}{2}}\iota_N}V_{RR}^{-\frac{1}{2}}\left\{\beta V_{FF}^{\frac{1}{2}}l + (\beta\psi_{\iota F} + \psi_{\iota\varepsilon})\right\} + O_p(T^{-1}) \\ = &\ M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + V_{RR}^{-\frac{1}{2}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})\right\} + O_p(T^{-1}) \\ = &\ M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\} + O_p(T^{-1})\end{aligned}$$

with $\psi^* = V_{RR}^{-\frac{1}{2}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon}) \sim N(0, I_N)$.

$\mathbf{V}_{RR}^{-\frac{1}{2}}\hat{\mathbf{B}}$. For the large sample behavior of $V_{RR}^{-\frac{1}{2}}\hat{B}$, we distinguish between strong and weak factors.

Strong factors. When the observed factors are strong and their number equals the true number of unobserved factors, the large sample behavior of $\hat{B}$ is characterized by (16):

$$\hat{B} = \ \beta V_{FG}V_{GG}^{-1} + \tfrac{1}{\sqrt{T}}\psi_{\varepsilon G}.$$

It results in a large sample behavior of $M_{V_{RR}^{-\frac{1}{2}}\iota_N}V_{RR}^{-\frac{1}{2}}\hat{B}$ which is characterized by:

$$\begin{aligned}M_{V_{RR}^{-\frac{1}{2}}\iota_N}V_{RR}^{-\frac{1}{2}}\hat{B} = &\ M_{V_{RR}^{-\frac{1}{2}}\iota_N}V_{RR}^{-\frac{1}{2}}\left\{\beta V_{FF}^{\frac{1}{2}}V_{FF}^{-\frac{1}{2}}V_{FG}V_{GG}^{-1} + \tfrac{1}{\sqrt{T}}\psi_{\varepsilon G}\right\} + O_p(T^{-1}) \\ = &\ M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}V_{FG}V_{GG}^{-\frac{1}{2}} \\ 0 \end{pmatrix} + \tfrac{1}{\sqrt{T}}V_{RR}^{-\frac{1}{2}}V_{\varepsilon\varepsilon}^{\frac{1}{2}}\varphi^*\right\}V_{GG}^{-\frac{1}{2}} + O_p(T^{-1}),\end{aligned}$$

with φ^* a $N \times m$ dimensional random matrix whose elements are independently standard normally distributed.

Weak factors. When the observed proxy factors are minorly correlated with the observed true factors as outlined in Assumption 1, the large sample behavior of $\sqrt{T}\hat{B}$ is:

$$\sqrt{T}\hat{B} \underset{d}{\to} \quad \beta\left(d+\psi_{VG}\right)+\psi_{\varepsilon G}$$

and results in large sample behavior of $\sqrt{T} M_{V_{RR}^{-\frac{1}{2}}\iota_N} V_{RR}^{-\frac{1}{2}}\hat{B}$ which is characterized by:

$$\begin{aligned}
\sqrt{T} M_{V_{RR}^{-\frac{1}{2}}\iota_N} V_{RR}^{-\frac{1}{2}}\hat{B} &\underset{d}{\to} \quad M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{V_{RR}^{-\frac{1}{2}}\beta V_{FF}^{\frac{1}{2}}\left(V_{FF}^{-\frac{1}{2}}d\right)+V_{RR}^{-\frac{1}{2}}\left(\beta\psi_{VG}+\psi_{\varepsilon G}\right)\right\} \\
&\underset{d}{\to} \quad M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W' V_{FF}^{-\frac{1}{2}} d V_{GG}^{\frac{1}{2}} \\ 0 \end{pmatrix}+\varphi^*\right\} V_{GG}^{-\frac{1}{2}}
\end{aligned}$$

with $\varphi^* = V_{RR}^{-\frac{1}{2}}\left(\beta\psi_{VG}+\psi_{\varepsilon G}\right)V_{GG}^{\frac{1}{2}}$ a $N\times m$ dimensional random matrix whose elements are independently standard normally distributed. The identity covariance matrix of φ^* results since $\hat{B} \underset{p}{\to} 0$ under Assumption 1.

GLS $\mathbf{R}^2$. Combining the large sample behaviors of $\bar{R}$ and $\hat{B}$, we obtain the large sample behavior of the GLS R^2 under weak and strong factors.

Strong factors:

$$\frac{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' P_{M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W' V_{FF}^{-\frac{1}{2}} V_{FG} V_{GG}^{-\frac{1}{2}} \\ 0 \end{pmatrix}+\frac{1}{\sqrt{T}} V_{RR}^{-\frac{1}{2}} V_{\varepsilon\varepsilon}^{\frac{1}{2}}\varphi^*\right\}}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}},$$

which results since $W' V_{FF}^{-\frac{1}{2}} V_{FG} V_{GG}^{-\frac{1}{2}}$ is an invertible $k\times k$ matrix.

Weak factors:

$$\frac{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' P_{M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W' V_{FF}^{-\frac{1}{2}} d V_{GG}^{\frac{1}{2}} \\ 0 \end{pmatrix}+\varphi^*\right\}}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}.$$

References

[1] Anderson, T.W. *An Introduction to Multivariate Statistical Analysis.* John Wiley (New York), second edition, 1984.

[2] Bai, J. and S. Ng. Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics*, **131**:507–537, 2006.

[3] Burnside, C. The Cross Section of Foreign Currency Risk Premia and Consumption Growth Risk: Comment. *American Economic Review*, **101**:3456–3476, 2011.

[4] Campbell, J.Y., A.L. Lo and A.C. MacKinlay. *The Econometrics of Financial Markets.* Princeton University Press, 1997.

[5] Chamberlain, G. and M. Rothschild. Arbitrage, Factor Structure, and Mean-Variancce Analysis on Large Asset Markets. *Econometrica*, **51**:1281–1302, 1983.

[6] Cochrane, J.H. *Asset Pricing.* Princeton University Press, 2001.

[7] Connor, G. and R. Korajczyk. Risk and Return in an Equillibrium APT: Application of a New Test Methodology. *Journal of Financial Economics*, **21**:255–289, 1988.

[8] Connor, G. and R. Korajczyk. An Intertemporal Equillibrium Beta Pricing Model. *Review of Financial Studies*, **2**:373–392, 1989.

[9] Daniel, K. and S. Titman. Testing Factor-Model Explanations of Market Anomalies. *Critical Finance Review*, **1**:103–139, 2012.

[10] Fama, E.F. and J.D. MacBeth. Risk, Return and Equillibrium: Empirical Tests. *Journal of Political Economy*, **81**:607–636, 1973.

[11] Fama, E.F. and K.R. French. The Cross-Section of Expected Stock Returns. *Journal of Finance*, **47**:427–465, 1992.

[12] 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.

[13] Fama, E.F. and K.R. French. Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance*, **51**:55–84, 1996.

[14] Gibbons, M. Multivariate Tests of Financial Models: A New Approach. *Journal of Financial Economics*, **10**:3–27, 1982.

[15] Gibbons, M., S. Ross and J. Shanken. A Test of the Efficiency of a Given Portfolio. *Econometrica*, **57**:1121–1152, 1989.

[16] Hansen, L.P. and R. Jagannathan. Assessing Specification Errors in Stochastic Discount Factor Models. *Journal of Finance*, **52**:557–590, 1997.

[17] Jagannathan, R. and Z. Wang. The Conditional CAPM and the Cross-Section of Expected Returns. *Journal of Finance*, **51**:3–53, 1996.

[18] Jagannathan, R. and Z. Wang. An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression. *Journal of Finance*, **53**:1285–1309, 1998.

[19] Kan, R. and C. Zhang. Two-Pass Tests of Asset Pricing Models with Useless Factors. *Journal of Finance*, **54**:203–235, 1999.

[20] Kan, R., C. Robotti and J. Shanken. Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology. *Journal of Finance*, **68**:2617–2649, 2013.

[21] Kleibergen, F. Tests of Risk Premia in Linear Factor Models. *Journal of Econometrics*, **149**:149–173, 2009.

[22] Lettau, M. and S. Ludvigson. Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia are Time-Varying. *Journal of Political Economy*, **109**:1238–1287, 2001.

[23] Lewellen, J., S. Nagel and J. Shanken. A Skeptical Appraisal of Asset-Pricing Tests. *Journal of Financial Economics*, **96**:175–194, 2010.

[24] Li, Q., M. Vassalou and Y. Xing. Sector investment growth rates and the cross section of equity returns. *The Journal of Business*, **79**:1637–1665, 2006.

[25] 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.

[26] Lustig, H. and S. Van Nieuwerburgh. Housing collateral, consumption insurance, and risk premia: An empirical perspective. *Journal of Finance*, **60**:1167–1219, 2005.

[27] MacKinlay, A.C. On Multivariate Tests of the CAPM. *Journal of Financial Economics*, **18**:341–372, 1987.

[28] Merton, R.C. An intertemporal capital asset pricing model. *Econometrica*, **41**:867–887, 1973.

[29] Onatski, A. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, **168**:244–258, 2012.

[30] Roll, R. and S.A. Ross. An Empirical Investigation of the Arbitrage Pricing Theory. *Journal of Finance*, **5**:1073–1103, 1980.

[31] Ross, S. The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, **13**:341–360, 1976.

[32] Santos, T. and P. Veronesi. Labor income and predictable stock returns. *Review of Financial Studies*, **19**:1–44, 2006.

[33] Shanken, J. On the Estimation of Beta-Pricing Models. *Review of Financial Studies*, **5**:1–33, 1992.

[34] Staiger, D. and J.H. Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65**:557–586, 1997.

[35] Stock, J.H. and M. Yogo. Testing for Weak Instruments in Linear IV Regression. In D.W.K. Andrews and J.H. Stock, editor, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge: Cambridge University Press, 2005. Prepared for Festschrift in honor of Thomas Rothenberg.

[36] M. Yogo. A consumption-based explanation of expected stock returns. *Journal of Finance*, **61**:539–580, 2006.

Supplementary Appendix.

Proof of Theorem 1. The least squares estimator of δ reads

$$\hat{\delta} = \sum_{t=1}^{T} \bar{F}_t \bar{G}_t' \left(\sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right)^{-1}$$

and the F-statistic (times number of parameters tested) testing if the factors G_t have an effect on F_t reads

$$F\text{-stat} = \text{trace}\left[\hat{V}_{VV}^{-1} \hat{\delta} \left(\sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right) \hat{\delta}' \right],$$

with $\hat{V}_{VV}$ an estimator of the covariance of the residuals, $\hat{V}_{VV} = \frac{1}{T-m-1} \sum_{t=1}^{T} (\bar{F}_t - \hat{\delta}\bar{G}_t)(\bar{F}_t - \hat{\delta}\bar{G}_t)'$. Under Assumption 1 and since $\hat{V}_{VV} \underset{p}{\rightarrow} V_{VV}$, $\frac{1}{T}\sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \underset{p}{\rightarrow} V_{GG}$, we have that

$$\begin{aligned} \sqrt{T}\hat{\delta} = & \sqrt{T} \sum_{t=1}^{T} \left(\tfrac{d}{\sqrt{T}} \bar{G}_t + \bar{V}_t \right) \bar{G}_t' \left(\sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right)^{-1} \\ = & \left(d \tfrac{1}{T} \sum_{t=1}^{T} \bar{G}_t \bar{G}_t' + \tfrac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{V}_t \bar{G}_t' \right) \left(\tfrac{1}{T} \sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right)^{-1} \underset{d}{\rightarrow} d + \psi_{VG}, \end{aligned}$$

with $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{V}_t \bar{G}_t' \left(\frac{1}{T} \sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right)^{-1} \underset{d}{\rightarrow} \psi_{VG} = V_{VV}^{\frac{1}{2}} \psi_{VG}^* V_{GG}^{-\frac{1}{2}}$ and ψ_{VG}^* is a $k \times m$ dimensional random matrix whose elements are independently standard normally distributed,

$$\begin{aligned} & F\text{-stat} = \text{trace}\left[\hat{V}_{VV}^{-1} (\sqrt{T}\hat{\delta}) \left(\tfrac{1}{T} \sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right) (\sqrt{T}\hat{\delta})' \right] \\ & \underset{d}{\rightarrow} \text{trace}\left[(d^* + \psi_{VG}^*)'(d^* + \psi_{VG}^*) \right] \sim \chi^2(\text{trace}(d^{*\prime} d^*), km), \end{aligned}$$

where $d^* = V_{VV}^{\frac{1}{2}} d V_{GG}^{-\frac{1}{2}}$ and $\chi^2(a, h)$ is a non-central χ^2 distributed random variable with h degrees of freedom and non-centrality parameter a.

Proof of Theorem 2. The least squares estimator $\hat{B}$ reads

$$\hat{B} = \sum_{t=1}^{T} \bar{R}_t \bar{G}_t' \left(\sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right)^{-1}.$$

Under the models in (2), (14) and Assumption 1, we can specify it as

$$\begin{aligned} \hat{B} = & \sum_{t=1}^{T} \left(\beta (\tfrac{d}{\sqrt{T}} \bar{G}_t + \bar{V}_t) + \bar{\varepsilon}_t \right) \bar{G}_t' \left(\sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right)^{-1} \\ = & \tfrac{1}{\sqrt{T}} \left[\beta \left(d (\tfrac{1}{T} \sum_{t=1}^{T} \bar{G}_t \bar{G}_t') + \tfrac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{V}_t \bar{G}_t' \right) + \tfrac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{\varepsilon}_t \bar{G}_t' \right] \left(\tfrac{1}{T} \sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right)^{-1}. \end{aligned}$$

We now use that $\frac{1}{T} \sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \underset{p}{\rightarrow} V_{GG}$, $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{V}_t \bar{G}_t' \left(\frac{1}{T} \sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right)^{-1} \underset{d}{\rightarrow} \psi_{VG} = V_{VV}^{\frac{1}{2}} \psi_{VG}^* V_{GG}^{-\frac{1}{2}}$, $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \bar{G}_t' \left(\frac{1}{T} \sum_{t=1}^{T} \bar{G}_t \bar{G}_t' \right)^{-1} \underset{d}{\rightarrow} \psi_{\varepsilon G} = V_{\varepsilon\varepsilon}^{\frac{1}{2}} \psi_{\varepsilon G}^* V_{GG}^{-\frac{1}{2}}$, and ψ_{VG}^* and $\psi_{\varepsilon G}^*$ are independent $k \times m$ and $N \times m$ dimensional random variables whose elements are independently standard normally distributed. ψ_{VG} and $\psi_{\varepsilon G}$ are independent since F_t and ε_t are uncorrelated so the same applies for V_t then as well since it is an element of F_t. Combining all elements, we obtain the limiting behavior of $\hat{B}$:

$$\sqrt{T}\hat{B} \underset{d}{\rightarrow} \quad \beta\left(d+\psi_{VG}\right)+\psi_{\varepsilon G}.$$

The Likelihood ratio statistic equals

$$\begin{aligned} \mathrm{LR} = & \; T\left[\log(\left|\tilde{\Sigma}\right|)-\log(\left|\hat{\Sigma}\right|)\right] \\ = & \; T\left[\log(\left|\tilde{\Sigma}\hat{\Sigma}^{-1}\right|)\right] \\ = & \; T\left[\log(\left|I_N+\hat{B}\left(\tfrac{1}{T-1}\textstyle\sum_{t=1}^{T}\bar{G}_t\bar{G}_t'\right)\hat{B}'\hat{\Sigma}^{-1}\right|)\right] \end{aligned}$$

where we used that the restricted covariance matrix estimator,

$$\begin{aligned} \tilde{\Sigma} = & \; \tfrac{1}{T-1}\textstyle\sum_{t=1}^{T}\bar{R}_t\bar{R}_t' \\ = & \; \tfrac{1}{T-1}\textstyle\sum_{t=1}^{T}(\bar{R}_t-\hat{B}\bar{G}_t)(\bar{R}_t-\hat{B}\bar{G}_t)'+\hat{B}\left(\tfrac{1}{T-1}\textstyle\sum_{t=1}^{T}\bar{G}_t\bar{G}_t'\right)\hat{B}' \\ = & \; \hat{\Sigma}+\hat{B}\left(\tfrac{1}{T-1}\textstyle\sum_{t=1}^{T}\bar{G}_t\bar{G}_t'\right)\hat{B}', \end{aligned}$$

with $\hat{\Sigma}=\frac{1}{T-1}\sum_{t=1}^{T}(\bar{R}_t-\hat{B}\bar{G}_t)(\bar{R}_t-\hat{B}\bar{G}_t)'$. Upon conducting a second order mean value expansion around $\log|I_N|$, the Likelihood ratio statistic can be approximated by

$$\begin{aligned} \mathrm{LR} = & \; T\log(|I_N|)+\mathrm{vec}\left[\hat{B}\left(\tfrac{1}{T-1}\textstyle\sum_{t=1}^{T}\bar{G}_t\bar{G}_t'\right)\hat{B}'\hat{\Sigma}^{-1}\right]'\mathrm{vec}(I_N)+O_p(T^{-2}) \\ = & \; \mathrm{trace}\left[\hat{B}\left(\tfrac{1}{T-1}\textstyle\sum_{t=1}^{T}\bar{G}_t\bar{G}_t'\right)\hat{B}'\hat{\Sigma}^{-1}\right]+O_p(T^{-2}) \\ \underset{d}{\rightarrow} & \; \mathrm{trace}\left[\left(\beta\left(d+\psi_{VG}\right)+\psi_{\varepsilon G}\right)V_{GG}(\beta\left(d+\psi_{VG}\right)+\psi_{\varepsilon G})'V_{RR}^{-1}\right] \\ = & \; \mathrm{trace}\left[(d^{+}+\psi_{RG}^{*})'(d^{+}+\psi_{RG}^{*})\right] \\ \sim & \; \chi^2(\mathrm{trace}(d^{+\prime}d^{+}),Nm) \end{aligned}$$

since $\frac{1}{T}\sum_{t=1}^{T}\bar{G}_t\bar{G}_t' \underset{p}{\rightarrow} V_{GG}$, $\hat{\Sigma} \underset{p}{\rightarrow} V_{RR}=\beta V_{FF}\beta'+V_{\varepsilon\varepsilon}$, and we used that $d^{+}=V_{RR}^{-\frac{1}{2}}\beta dV_{GG}^{\frac{1}{2}}$, $V_{RR}^{-\frac{1}{2}}(\beta\psi_{VG}+\psi_{\varepsilon G})V_{GG}^{\frac{1}{2}}=\psi_{RG}^{*}$ with ψ_{RG}^{*} a $N\times m$ random matrix whose elements are independently normally distributed, $\mathrm{vec}(A)$ is the column vectorization of the matrix A.