Unexplained factors and their effects on second pass R-squared's and t-tests

Frank Kleibergen[*]   Zhaoguo Zhan[†]

March 11, 2013

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. The small correlation implies a sizeable unexplained factor structure in the first pass residuals and, consequently, a sizeable 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 between the expected portfolio returns and the (macro-) economic factors. We propose an easy manner for diagnosing it using a statistic that reflects the unexplained factor structure in the first pass residuals. Similar arguments apply to the second pass t-statistic which are resolved using the identification robust factor statistics of Kleibergen (2009). Our results put into question many of the empirical findings that concern the relationship between expected portfolio returns and (macro-) economic factors. We discuss some prominent ones in passing.

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 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), Santos and Veronesi (2006) and Yogo (2006). The ordinary and generalized least squares R^2's of the second pass regression alongside the t-statistics of the risk premia are used to gauge the strength of the relationship between the expected portfolio returns and the involved factors.

---

[*]Department of Economics, Brown University, Box B, Providence, RI 02912, United States. Email: Frank_Kleibergen@brown.edu. 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.

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 (2011) 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 are the proxies 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. Since both the average portfolio returns and the estimated β's are basically linear in the β's of the unobserved true factors, 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 then 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 also akin to the weak instrument assumption made for the linear instrumental variables regression model in econometrics, see *e.g.* Staiger and Stock (1997).

The paper is organized as follows. Following *e.g.* Merton (1973), Ross (1976), Roll and Ross (1980), Chamberlain and Rothschild (1983) and Connor and Korajczyk (1988, 1989), we first in the second section lay out the factor structure in portfolio returns. We then propose a measure for the unexplained factor structure in the first pass residuals of the FM two pass procedure. We use it to 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 of their large sample distributions. We show that the unexplained factor structure is likely to have explained the large OLS R^2 documented in many empirical studies, like, Jagannathan and Wang (1996, 1998), Lettau and Ludvigson (2001), Lustig and Van Niewerburgh (2005), Li *et. al.* (2006), Santos and Veronesi (2006) and Yogo (2006). The large OLS R^2 in these studies are then not indicative of the strength of the relationship between expected portfolio returns and the involved (macro-) economic factors. The third section also briefly discusses the FM t-statistic and the identification robust factor statistics. 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):

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

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

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

The vector notation of the factor model in (2) shows that, if the factors F_t, $t = 1, \ldots, T$, are i.i.d. with finite variance and are uncorrelated with the disturbances ε_t, $t = 1, \ldots, T$, 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 $P = (p_1 \ldots p_N)$ 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 statistics used in the FM two pass procedure.

2.1 Factor Structure in Observed Portfolio Returns

We use three different data sets to show that the effects of the unexplained factor structure on the FM two pass procedure are highly relevant. The first data set is the one used by Lettau and Ludvigson (2001). The portfolio returns used by Lettau and Ludvigson (2001) consist of quarterly observations from the third quarter of 1963 to the third quarter of 1998 of the return on twenty-five size and book-to-market sorted portfolios so $N = 25$ and $T = 141$. The second data set results from Jagannathan and Wang (1996) and their portfolio return series consist of the monthly returns on one hundred size and beta sorted portfolios. Their return series begin in July 1963 and end in 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 return 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 portfolio returns. It is clear from these characteristic roots 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.

We use the fraction of the total variation of the portfolio returns that is explained by the three largest principal components as a check for the presence of a factor structure. 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

Table 1 shows that there is compelling evidence for a factor structure in portfolio returns, see also Connor and Koraczyk (1988). Alongside describing portfolio returns using "unobserved factors", as briefly discussed previously, a large literature exists which explains portfolio returns using observed factors which are proxies for the unobserved ones. The observed proxy factors that are used in this literature 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 $N \times m$ 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)   show that the variation in portfolio returns can be explained using a factor model with three observed asset return based factors: the return on a value weighted portfolio, a "small minus big" (SMB) factor which consists of the difference in returns on a portfolio consisting

---

[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.

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

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 of 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 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. Table 2 shows that the three largest characteristic roots of the covariance matrix have decreased substantially after incorporating the three FF factors. The factor structure check shows that the factors associated with the three largest principal components of the residuals explain 47.5% of their total variation while the factors associated with the three largest principal components of the covariance matrix of the portfolio returns explain 94.3% of their total variation. It shows that the three FF factors eradicate the factors associated with the three largest principal components of the portfolio returns. After incorporating the FF factors, there is no unexplained factor structure left over in the residuals since the characteristic roots of their covariance matrix decline gradually and do not have a specific cutoff after which the decline of the characteristic roots becomes much more gradual.

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, $H_0 : B = 0$, against the alternative hypothesis that they are unequal to zero, $H_1 : B \neq 0$, can be specified as

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

with $\hat{V}_{Port}$ and $\hat{V}_{\text{Res}}$ 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 $\lambda_{i,\text{port}}$, $i = 1, \ldots, N$, the characteristic roots of the covariance matrix of the portfolio returns, $\hat{V}_{Port}$, and $\lambda_{i,\text{res}}$, $i = 1, \ldots, N$, the characteristic roots of the covariance matrix of the residuals of the observed factor model, $\hat{V}_{\text{Res}}$.[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 equals 2064 which is strongly significant since its p-value is 0.000 which is in line with our finding that the FF factors eradicate the factors associated with the largest principal components of the covariance matrix of the portfolio returns.[4]

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

---

[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 $\hat{V}_{\text{Res}}$, 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. 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.

	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
6	14.0	10.8	16.2	14.7	14.8	14.8
7	11.6	9.17	14.8	13.9	14.4	14.3
8	10.9	8.61	14.0	13.5	13.7	13.6
9	9.92	7.84	12.6	12.2	12.2	12.2
10	8.18	6.55	12.1	11.7	12.1	12.0
11	7.19	6.43	12.1	11.6	11.8	11.7
12	6.32	5.77	11.4	11.3	11.3	11.2
13	6.17	5.16	11.3	10.7	11.1	11.1
14	5.63	5.09	11.1	10.6	10.8	10.8
15	5.21	4.64	10.8	10.4	10.5	10.5
16	5.02	4.31	10.3	10.3	10.3	10.3
17	4.40	3.88	10.2	9.81	10.2	9.94
18	3.83	3.47	9.9	9.58	9.80	9.71
19	3.43	2.95	9.6	9.38	9.54	9.51
20	2.90	2.83	9.5	9.03	9.18	9.17
21	2.79	2.74	9.2	8.90	9.11	9.11
22	2.75	2.51	9.0	8.88	8.91	8.85
23	2.47	2.19	8.9	8.64	8.70	8.51
24	2.12	1.76	8.7	8.39	8.47	8.44
25	1.77	1.64	8.4	8.15	8.36	8.30
FACCHECK	94.3%	47.5%	86%	23%	57%	57%
LR against raw		2064 0.000		2994 0.000	1586 0.000	1845 0.000
LR against R_{vw}		1285 0.000		1408 0.000		259 0.004
F-stat prin. comp.		13777 0.000		26321 0.000	8777 0.000	8826 0.000
F-stat prin. comp. R_{vw}						16.8 0.010
F-stat HML-SMB						3.51 0.476
pseudo-R^2		0.915		0.823	0.681	0.684
pseudo-R^2 prin. comp				0.96	0.791	0.794
pseudo-R^2 FF				1	0.627	0.63

Table 2: The largest twenty 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 three largest principal components, the FF factors or only the HML-SMB factors on them. The pseudo-R^2 of these regressions are listed further down. FACCHECK equals the percentage of the variation explained by the three largest principal components.

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 equal to 1285 which is highly significant when compared with its 95% critical value that results from a $\chi^2(2N)$ distribution 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 tests the significance of the FF factors in a regression of the three factors that result from the three largest principal components of the portfolio returns on the FF Factors.[5] The F-statistic therefore results from testing $\mathrm{H}_0 : \delta = 0$ in the linear model:

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

with F_t a 3×1 vector that results from the three largest principal components and G_t a 3×1 vector containing the FF factors. The number of elements of δ equals nine. Table 2 shows that the F-statistic is huge which further reflects the close connection between the largest principal components and the FF factors which is also revealed by the LR statistics and the factor structure checks, see also Bai and Ng (2006).

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. Table 2 contains the largest twenty-five characteristic roots of the covariance matrices that result for the Jagannathan and Wang (1996) data set. Besides the characteristic roots of the covariance matrix of the raw portfolio returns, Table 2 contains the characteristic roots of the covariance matrix of the residuals of three observed factor models. The first one of these uses the three FF factors (for which we use the SMB and HML factors from Ken French's website), the second one has the value weighted return, R_{vw}, as the only factor while the third specification uses all three factors from Jagannathan and Wang (1996).

The characteristic roots associated with the Jagannathan and Wang (1996) data show that the FF factors remove the factor structure in the portfolio returns since the characteristic roots of the covariance matrix of the residuals decline gradually so there are no roots that are distinctly larger than all the others. This is further reflected by the value of the factor structure check of 23% and the pseudo-R^2 of 82.3%. The LR statistics indicate that the parameters of all the FF factors are strongly significant in the regression of the portfolio returns on the FF factors. The F-statistic shows that the

---

[5]Since the dependent variable is estimated, we should correct the covariance matrix involved in the F-statistic but we did not do so since it would not alter the overall conclusion that the F-statistic is highly significant.

[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 $\hat{V}_R = \hat{V}_{\mu+\hat{B}G} + \hat{V}_{R-\mu-\hat{B}G}$.

same holds for the regression of the three largest principal components on the FF factors. All these findings are identical to those previously discussed for the quarterly portfolio returns.

If the value weighted return is the only factor in the observed factor model, Table 2 shows that there remains a considerable factor structure in the residuals since the characteristic roots of the covariance matrix of the residuals show that there are (probably) still two unexplained factors present. This is further revealed by the factor structure check of 57% which indicates a factor structure in the residuals. Interestingly the pseudo-R^2 is 68.1% which means that the value weighted return is amongst the FF factors the most important explanator of the portfolio returns which also explains the highly significant values of the LR and F-statistics.

When we add the yield spread and labor income factors to the factor model with only the value weighted return, Table 2 shows that the characteristic roots of the covariance matrix of the residuals hardly change. The factor structure check therefore remains equal to 57% and the pseudo-R^2 still equals 68.4%. This indicates that the yield spread and labor income factors do not capture the two unexplained factors present in the residuals that result from the factor model with only the value weighted return. This is further reflected by the LR statistic that tests the significance of the parameters associated with the yield spread and labor income factors. This LR statistic is equal to 259 and results from an expression identical to (8) but using the characteristic roots of the residual covariance matrices that result from factor models with only the value weighted return and all three factors used by Jagannathan and Wang (1996) for which the largest twenty-five are stated in the sixth and seventh columns of Table 2. Although this LR statistic is significant compared with its 95% critical value from a $\chi^2(2N = 200)$ distribution, its p-value is only 0.004. This indicates that the parameters associated with the yield spread and labor income factors are all rather close to zero.

The small explanatory power of the yield spread and labor income growth is further reflected by the F-statistics that result from regressing the three largest principal components or the SMB and HML factors on the value weighted return, yield spread and labor income growth. The F-statistics labeled "F-stat prin. comp. R_{vw}" and "F-stat HML-SMB" test the significance of the yield spread and labor income factors in regressions of respectively the three largest principal components ("F-stat prin. comp. R_{vw}") and the HMB and SMB factors ("F-stat FF factors no R_{vw}") on the value weighted return, yield spread and labor income growth. The first of these F-statistics equals 16.8 which equals the 99% critical value of the $\chi^2(6)$ distribution while the latter equals 3.51 which is below the 95% critical value of the $\chi^2(4)$ distribution (9.49). It shows that we cannot reject the hypothesis that the yield spread and labor income growth have no explanatory power for the HML and SMB factors at the 95% significance level.

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 ($cay\Delta y$). The characteristic roots of the covariance matrix of the residuals that result from six different observed factor models estimated in Lettau and Ludvigson (2001) are stated in Table 3.

The characteristic roots in Table 3 show that for all specifications of the observed factor model, except the one using the FF factors, the residuals still contain a factor structure. This is reflected by the factor structure check which is above 82% for all these specifications. For the observed factor models that do not contain the value weighted return as one of the factors, the factor structure check is close to the one which results for the portfolio returns itself. Hence, there are still three unexplained factors present in the residuals of these factor models while there are two unexplained factors present

	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
6	13.5	12.2	13.5	8.91	12.1	13.0	11.8
7	12.1	9.33	12.0	8.0	9.20	11.9	9.19
8	9.31	8.51	9.28	7.03	8.31	9.24	8.18
9	8.42	7.26	8.37	6.27	7.06	8.24	6.91
10	7.25	6.03	7.12	5.17	5.75	7.09	5.65
11	6.02	5.53	6.01	4.65	5.02	5.61	4.95
12	5.4	4.91	5.34	4.40	4.88	5.27	4.87
13	4.9	4.38	4.86	4.25	4.36	4.86	4.25
14	4.38	4.26	4.38	3.73	3.99	4.34	3.93
15	4.26	3.97	4.18	3.61	3.85	3.89	3.85
16	3.93	3.54	3.93	3.34	3.40	3.85	3.40
17	3.5	3.39	3.44	3.06	3.38	3.35	3.32
18	3.39	3.03	3.30	2.91	2.99	3.12	2.92
19	3.02	2.77	3.01	2.52	2.77	2.98	2.73
20	2.71	2.51	2.69	2.29	2.49	2.68	2.48
21	2.5	2.33	2.46	2.07	2.30	2.42	2.29
22	2.18	1.99	2.17	1.54	1.99	2.16	1.97
23	1.74	1.47	1.73	1.44	1.42	1.71	1.41
24	1.46	1.44	1.46	0.95	1.33	1.38	1.30
25	0.93	0.92	0.92	0.68	0.90	0.91	0.90
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 prin. comp		7965 $_{0.000}$	6.26 $_{0.100}$	32853 $_{0.000}$	6926 $_{0.000}$	32.01 $_{0.000}$	7820 $_{0.000}$
F-stat prin. comp. R_{vw}				12515 $_{0.000}$	4.10 $_{0.663}$	18.1 $_{0.006}$	13.72 $_{0.319}$
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		0.78	0.016	0.95	0.78	0.10	0.79
pseudo-R^2 prin. comp		0.814	0.016	0.99	0.815	0.10	0.822
pseudo-R^2 FF		0.63	0.01	1	0.63	0.094	0.64

Table 3: The twenty 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 three largest principal components, the FF factors or the HML-SMB factors on them. The pseudo-R^2 of these regressions are listed further down. FACCHECK equals the percentage of the variation explained by the three largest principal components.

in the residuals of the factor models that include the value weighted return (except for the specification that includes all three FF factors).

The likelihood ratio statistics reported in Table 3 test the significance of the parameters associated with the observed factors. They result from a specification identical to the one in (8) when the characteristic roots are chosen from the appropriate columns in Table 3. The degrees of freedom of their χ^2 large sample distribution equals the number of tested parameters. The likelihood ratio statistics show that only the parameters of the FF factors and the value weighted return are strongly significant. For all the other specifications, the likelihood ratio statistic are always less than twice the number of estimated parameters. This indicates that although the LR statistic might be significant at the 95% significance level, which is not even the case for the consumption growth, the values of the parameters associated with the observed factors are all close to zero.

The F-statistics in Table 3 show that the value weighted return and the HML and SMB factors are the only factors that have strong explanatory power for the three largest principal components. For all other factors, the F-statistics testing their significance are either not significant or just barely so. This is in contrast to the huge F-statistics that result for the value weighted return and the HML and SMB factors. As expected, because of the strong correlation between the three largest principal components and the FF factors, the F-statistics that result from regressions of the FF factors on the factors are similar to those that result from regressing the three largest principal components on them.

Li, Vassalou and Xing (2006) use the 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 observed factor model using the previously discussed quarterly portfolio returns from French's website.

Table 4 contains the estimation results for the observed factor model from Li *et. al.* (2006). It contains two specifications of the observed factor model, one which uses all three factors and one which only uses the FINAN factor. For both specifications, the characteristic roots of the covariance matrix of the residuals are comparable to those of the covariance matrix of the raw portfolio returns. Hence, both specifications do not capture the factor structure of the portfolio returns. This is further reflected by the factor structure check which equals 94.3% for both specifications and coincides with that of the raw portfolio returns. The LR statistics reported in Table 4 therefore show that the parameters of the FINAN factors are not significant at the 95% significance level while the parameters of the HHOLD and NFINCO factors are only significant at the 99% significance level. Hence, the parameters of all three factors are close to zero. The F-statistics that result from regressing the three largest principal components on these factor further confirm this since they are borderline significant (FINAN) or not (all three factors) when we test at the 95% significance level.

Lustig and Van Nieuwerburgh (2005) employ an observed factor model that contains non-durable 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 are reported in Table 4.

Table 4 contains two specifications of the factor model used by Lustig and Van Nieuwerburgh (2005). One of these specifications uses all three factors while the other one has nondurable consumption growth as the only factor (Δc_{nondur} column). The characteristic roots in Table 4 show that both specifications do not capture the factor structure of portfolio returns since Table 4 shows that the residuals of both of these two models still contain three unexplained factors. This is further reflected by the factor structure check which equals 93.9% for the specification with nondurable consumption

	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
6	14.0	12.3	13.9	13.9	13.8	13.8	12.2	11.9
7	11.6	10.9	11.1	11.6	11.3	11.1	10.9	10.8
8	10.9	9.98	10.8	10.8	10.8	10.8	9.90	9.86
9	9.92	8.24	9.89	9.71	9.70	9.76	8.01	8.19
10	8.18	7.22	8.13	8.14	8.12	8.12	6.91	7.22
11	7.19	6.50	7.19	7.16	7.00	7.16	6.33	6.47
12	6.32	6.19	6.24	6.31	6.31	6.23	6.16	6.15
13	6.17	5.70	6.17	6.17	6.16	6.09	5.64	5.66
14	5.63	5.59	5.61	5.62	5.59	5.55	5.57	5.56
15	5.21	5.03	5.20	5.20	5.19	5.19	4.98	5.01
16	5.02	4.62	5.02	5.02	5.00	4.99	4.57	4.51
17	4.40	4.31	4.39	4.29	4.25	4.37	4.31	4.21
18	3.83	3.62	3.81	3.82	3.77	3.79	3.48	3.53
19	3.43	3.43	3.42	3.40	3.32	3.38	3.37	3.41
20	2.90	2.89	3.88	2.84	2.77	2.84	2.87	2.88
21	2.79	2.79	2.78	2.77	2.74	2.75	2.69	2.77
22	2.75	2.71	2.73	2.73	2.68	2.53	2.60	2.67
23	2.47	2.39	2.46	2.47	2.43	2.45	2.27	2.40
24	2.12	2.12	2.10	2.10	2.00	2.09	2.10	2.10
25	1.77	1.62	1.76	1.73	1.72	1.74	1.61	1.62
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 prin. comp		4976 (0.000)	17.8 (0.000)	9.2 (0.027)	16.2 (0.064)	27.5 (0.001)	5135 (0.000)	5065 (0.000)
F-stat prin. comp. R_{vw}								26.6 (0.000)
F-stat FF factors		—	22.4 (0.000)	6.9 (0.074)	11.5 (0.240)	31.2 (0.000)	—	—
pseudo-R^2		0.723	0.065	0.007	0.015	0.083	0.739	0.724
pseudo-R^2 prin. comp		0.584	0.05	0.008	0.017	0.065	0.609	0.587
pseudo-R^2 FF		0.321	0.028	0.015	0.026	0.043	0.374	0.326

Table 4: The twenty 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). All use the quarterly portfolio returns from French's website. 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 three largest principal components or the FF factors on them. The pseudo-R^2 of these regressions are listed further down. FACCHECK equals the percentage of the variation explained by the three largest principal components.

growth as the only factor and 93.8% for the specification that uses all three factors. The likelihood ratio statistics that test the significance of the parameters of the consumption growth and the joint significance of the housing collateral and the interaction factor are only significant at the 98% and 59% significance level while the F-statistic that results from regressing the three largest principal components on all three factors is significant at the 94% level. This all shows that the empirical support for the factors proposed by Lustig and Van Nieuwerburgh (2005) is rather minor.

Santos and Veronesi (2006) use adaptations of the factors from Lettau and Ludvigson (2001). Alongside the value weighted returns, 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 contains the estimation results for two specifications of the observed factor model used by Santos and Veronesi (2006). The first specification has the value weighted return as the only factor while the other specification uses all three factors. The characteristic roots of the covariance matrix of the residuals that result for the first specification show that the value weighted return removes one factor so the residuals contain two unexplained factors. Table 4 shows that if we consider all three observed factors, the residuals still have a factor structure with two unexplained factors. This is further reflected by the factor structure check which equals 81.4% for the specification with the value weighted return as the only factor and 80.5% for the specification with all three factors. Hence, the additional two factors explain little of the variation in the portfolio returns. The likelihood ratio statistic that tests the joint significance of the cay and labor-consumption ratios is equal to 118. Although this is a rather significant value of the LR statistic, we have to consider that it tests fifty parameters so it is only twice the number of tested parameters which shows that all these parameters are relatively close to zero. The LR statistic which tests the significance of the parameters of the value weighted return is obviously highly significant.

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 his specification using the portfolio returns from French's website.

Table 4 contains two specifications of the observed factor model from Yogo (2006). The first one has the value weighted return as the only factor and has been discussed previously. The other specification has all three factors. The characteristic roots of the factor model with all three factors show that the additional two consumption growth factors fail to absorb the two factors left over in the residuals from the model with the value weighted return as the only factor. The characteristic roots of the covariance matrices of the residuals of these two models are almost identical which indicates that there are still two unexplained factors left in the residuals of the three factor model. This is further reflected by the factor structure check which equals 81.4% for the specification with only the value weighted return and 81.5% for the specification with all three factors. The likelihood ratio statistic which tests the joint significance of the parameters associated with the two consumption growth factors is also only significant at the 54% significance level so we cannot rule out that the parameters associated with the two consumption growth factors are all equal to zero.

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 β's, to obtain estimates of the zero-β return λ_0 and the risk premia λ_F:

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

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 therefore hinge on the ability of the observed factor model to capture the factor structure of the portfolio returns. To show this clearly, we consider the (infeasible) linear regression model for the unobserved factors F_t that uses the observed factors G_t as explanatory variables:

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

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 uncorrelated with ε_t since F_t is uncorrelated with ε_t. 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 reflected by the F-statistics in Tables 2-4. They test the hypothesis that δ, or some of its rows, in (14) is equal to zero. Since F_t is unknown, we approximate it by the three largest principal components or by the FF factors. Tables 2-4 show that, when the elements of δ being tested do not concern any of the FF factors, 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 FF factors. 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. *The parameter δ in the (infeasible) linear regression model for the unobserved factors that uses the observed factors as explanatory variables (14) is drifting to zero:*

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

with d a fixed full rank matrix.

Traditional large sample inference requires that both β and δ are full rank matrices, which is as argued before 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 F-statistic testing the significance of δ in (14) converges to a non-central χ^2 distributed random variable with km degrees of freedom and non-centrality parameter $trace(d^{*\prime}d^*)$, $d^* = V_{VV}^{\frac{1}{2}}dV_{GG}^{-\frac{1}{2}}$, $V_{VV} = var(V_t)$.*

Proof. see Appendix A. ∎

Theorem 1 shows that, since the correlation between the observed proxy and unobserved true factors is small, the F-statistic testing the significance of δ in (14) is not proportional to the sample size and therefore quite small. This is in line with the F-statistics reported in Tables 2-4 so Assumption 1 is more realistic than the traditional full rank assumption for δ. It extends in an identical manner to

the LR statistics testing the significance of the elements of B in the observed factor model (7) which are reported in Tables 2-4.[7]

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

Proof. see Appendix A. ∎

Theorem 2 shows that when the observed factors, whose significance is tested using the LR statistic, only minorly correlate with the three largest principal components, or put differently the FF factors, the LR statistic is small and only minorly significant. This is exactly as we observe in Tables 2-4 for all factors except the FF factors.

The large sample properties of the F and LR statistics stated in Theorems 1 and 2 are in line with the realized values of these statistics stated in Tables 2-4 for all factors except the FF ones. Under the traditional assumption of correlation between observed and unobserved factors, all these statistics are proportional to the sample size which they clearly aren't. 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 used 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 two kind of statistics which are commonly used in the FM two pass approach: goodness of fit measures and statistics testing hypotheses on the risk premia of the observed factors.

3.1 The R^2's of the Second Pass Regression

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:*

$$R^2_{OLS} \approx \frac{[\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]' P_{M_{\iota_N}(\beta(d+\psi_{VG}) + \psi_{\varepsilon G})} [\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]}{[\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]' M_{\iota_N}[\beta\lambda_F + \frac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]}, \tag{19}$$

---

[7]Theorems 1 and 2 apply to the tests on all elements of resp. δ in (14) and B in (7). When some of the elements of δ and B refer to factors that are part of the FF factors, the results in Theorems 1 and 2 apply to the tests on those elements of δ and B that do not refer to the FF factors.

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 A. ∎

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_{OLS}^2 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, R_{OLS}^2 converges to one when the sample size increases.*

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_{OLS}^2 is approximately equal 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 to

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

when the sample size goes to infinity.

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 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_{OLS}^2 so we understand where its properties result from.

The first important property Theorem 3 shows is that, under Assumption 1, R_{OLS}^2 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 elements do not affect the large sample behavior of R_{OLS}^2 so R_{OLS}^2 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_{OLS}^2 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_{OLS}^2 associated with the strongly correlated observed factors converges to its population value while the remaining part converges to a random variable. In total, R_{OLS}^2 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, the large sample behavior of R^2_{OLS} is characterized by*

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

(22)

where we use that $P_{(A\,:\,B)} = P_A + P_{M_A B}$ and d is such that $d = (d_1 \,\vdots\, d_2)$, with d_1 much larger than d_2 and $\psi_{VG,1}$, $\psi_{VG} = (\psi_{VG,1} \,\vdots\, \psi_{VG,2})$, $\psi_{\varepsilon V} = (\psi_{\varepsilon V,1} \,\vdots\, \psi_{\varepsilon V,2})$ and d_1, $\psi_{VG,1} : k \times m_1$, d_2, $\psi_{VG,2} : k \times m_2$, $\psi_{\varepsilon V,1} : N \times m_1$, $\psi_{\varepsilon V,2} : N \times m_2$, $m_1 + m_2 = m$.

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 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, 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 $M_{\iota_N}\beta(\lambda_F + \tfrac{1}{\sqrt{T}}\psi_{\iota F})$ 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 result 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 prin. comp and pseudo-R^2 FF[8], 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 either by the three largest principal components or FF factors, 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

---

[8]These pseudo-R^2's result from the expression in (10) but then applied to the regression of the three largest principal components on the observed proxy factors (prin. comp) or the regression of the FF factors on the observed proxy factors (FF).

averages of these components using $\frac{1}{\sqrt{T}}\bar{G}'_t\left(\frac{1}{T}\sum_{t=1}^T\bar{G}_t\bar{G}'_t\right)^{-1}$, $\bar{G}_t = G_t - \bar{G}$, $\bar{G} = \frac{1}{T}\sum_{t=1}^T G_t$, as weight converge to ψ_{VG} and $\psi_{\varepsilon G}$. The small values of the pseudo-R^2's then 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 results from the minorly correlated observed factors and equals a ratio whose numerator results from the projection of

$$M_{(\iota_N\,:\,\beta d_1)}[\beta\lambda_F + \tfrac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})] \qquad \text{on} \qquad M_{(\iota_N\,:\,\beta d_1)}\left(\beta\left(d_2 + \psi_{VG,2}\right) + \psi_{\varepsilon G,2}\right)$$

and its denominator equals $[\beta\lambda_F + \tfrac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]'M_{\iota_N}[\beta\lambda_F + \tfrac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]$. This second component has a similar expression as R^2_{OLS} in case 1. Hence, it is big when $M_{(\iota_N\,:\,\beta d_1)}\beta\left(d_2 + \psi_{VG,2}\right)$ is large compared to $M_{(\iota_N\,:\,\beta d_1)}\psi_{\varepsilon G,2}$. This occurs when either d_2 is large, so all observed factors are strongly correlated with the unobserved ones which case is covered by Corollary 1, or when d_2 is small and $\beta\psi_{VG,2}$ is large compared to $\psi_{\varepsilon G,2}$. The latter case occurs for the same reason as discussed previously for case 1. Hence, when d_2 is small and the number of strongly correlated observed factors is less than the number of unobserved factors, an unexplained factor structure in the residuals of (15) remains. This is shown in Tables 2-4 when we use only the value weighted return alongside some factors other than the FF factors. Tables 2-4 show that the residuals of the observed factor model exhibit a strong factor structure so βV_t is large compared to ε_t. The two components $\beta\psi_{VG,2}$ and $\psi_{\varepsilon G,2}$ are the limits of the weighted averages of βV_t and ε_t when we use $\frac{1}{\sqrt{T}}G^{*\prime}_{2,t}\left(\frac{1}{T}\sum_{t=1}^T G^*_{2,t}G^{*\prime}_{2,t}\right)^{-1}$, with $G^*_{2,t} = \bar{G}_{2,t} - \hat{\gamma}\bar{G}_{1,t}$, $\hat{\gamma} = \sum_{t=1}^T \bar{G}_{2,t}\bar{G}'_{1,t}\left(\frac{1}{T}\sum_{t=1}^T \bar{G}_{1,t}\bar{G}'_{1,t}\right)^{-1}$, $\bar{G}_t = \binom{\bar{G}_{1,t}}{\bar{G}_{2,t}}$, $\bar{G}_{1,t}: m_1 \times 1$, $\bar{G}_{2,t}: m_2 \times 1$, as weights. This implies that since βV_t is large compared to ε_t also $\beta\psi_{VG,2}$ is large compared to $\psi_{\varepsilon G,2}$. It is therefore the estimation error in $\hat{B}$ which results from the unexplained factor structure in the residuals of (15) that leads to possibly large values of R^2_{OLS}.

The above shows that it is the unexplained factor structure in the residuals of (15) that leads to the large values of R^2_{OLS} when the observed proxy and unobserved true factors are minorly correlated. Since these large values of R^2_{OLS} then result from the estimation error in the estimator of the β's of the observed proxy factors, they are not indicative of the presence of a relationship between the expected portfolio returns and the observed proxy factors. An easy diagnostic to assess the adequacy of R^2_{OLS}, as a measure for such a relationship, is therefore to use a factor structure statistic, like, for example, the one in (6).

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 the data from Lettau and Ludvigson (2001). We use the FM two pass procedure to estimate the risk premia on the three FF factors using 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 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 and distribution 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 and distribution function of R^2_{OLS} that result 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. Figures 1.1 and 1.2 also show 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 reports the density and distribution functions of our factor structure check. Figures 1.3-1.4 show 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.

Panel 2 shows the density and distribution 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 and distribution functions of R^2_{OLS} in Figures 2.1 and 2.2 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 and distribution functions of the factor structure check. As expected, its density and distribution 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 distribution of the factor structure check which shows that the unexplained factor structure in Panel 3 is less pronounced than in Panel 2.

---

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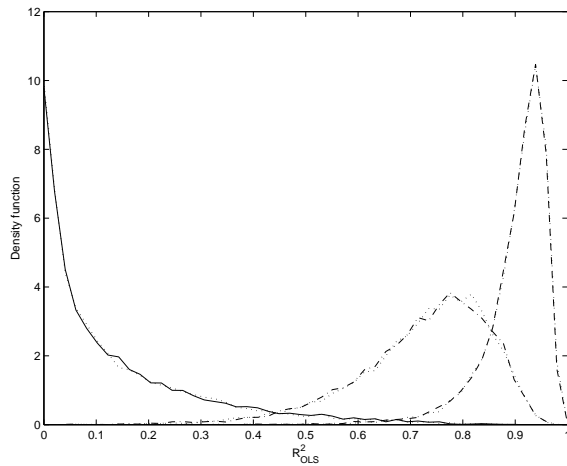
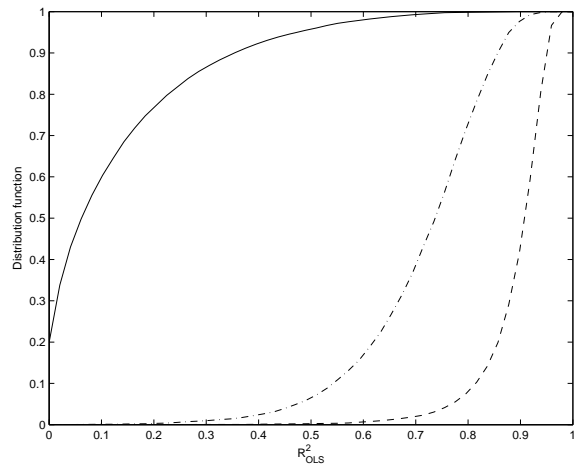Figure 1.1. Density functions of R^2_{OLS}

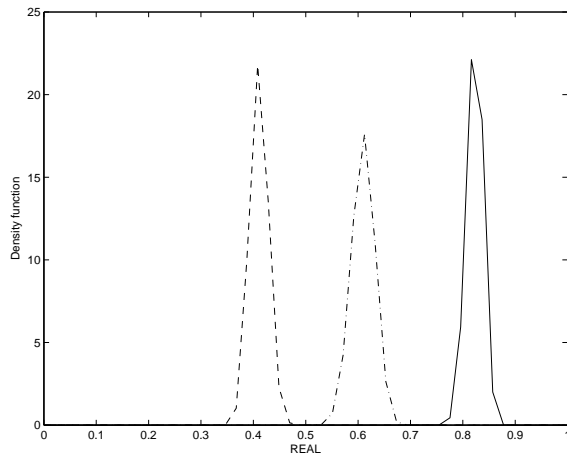Figure 1.2. Distribution functions of R^2_{OLS}

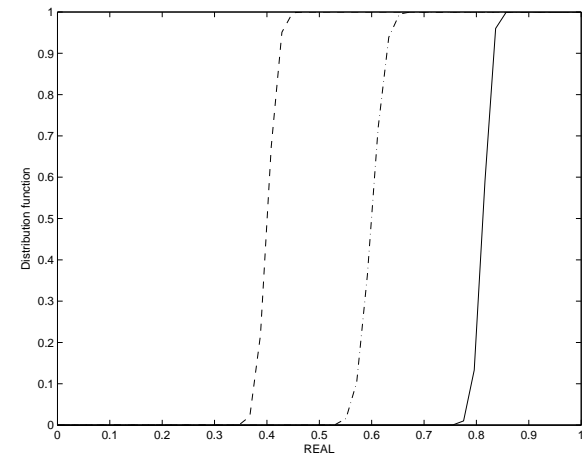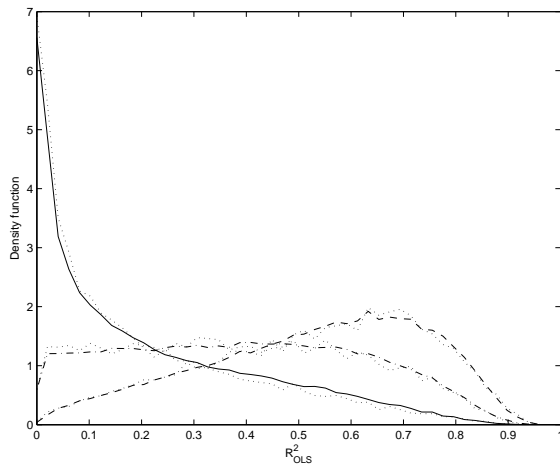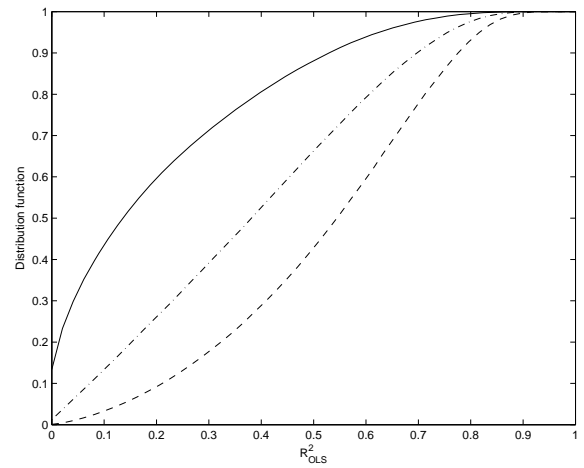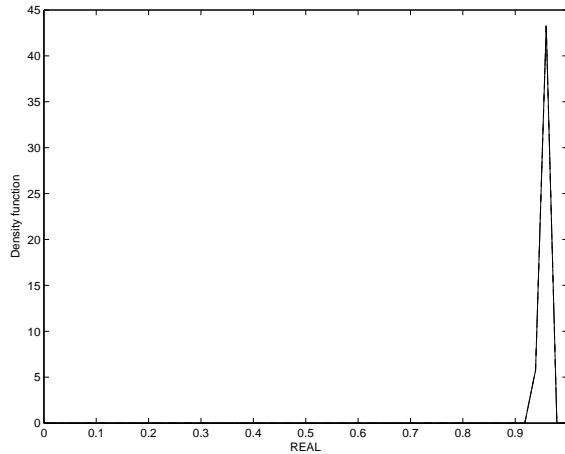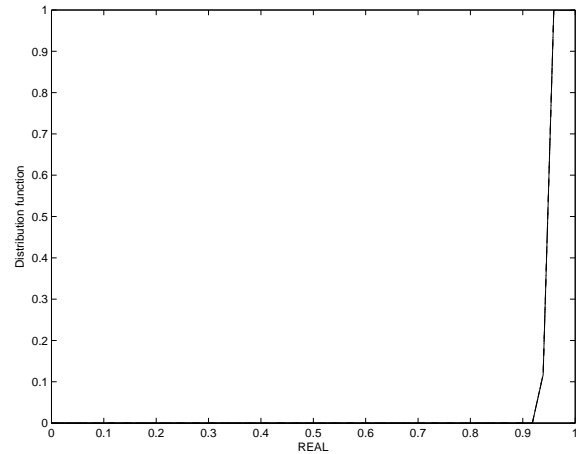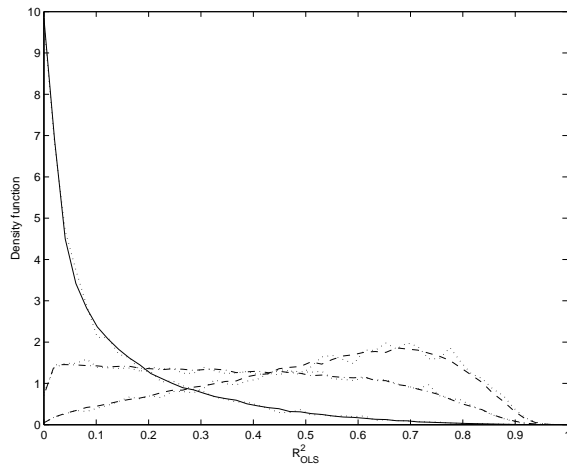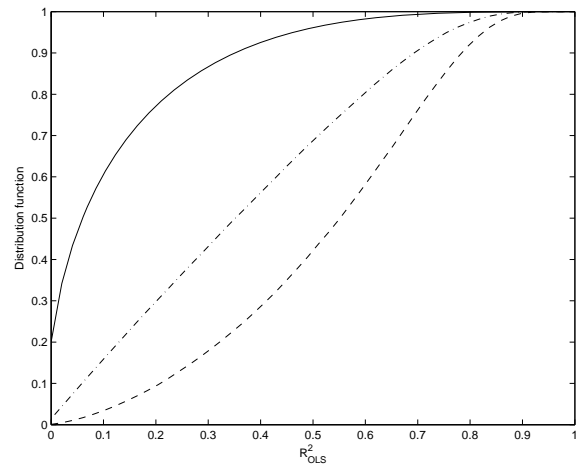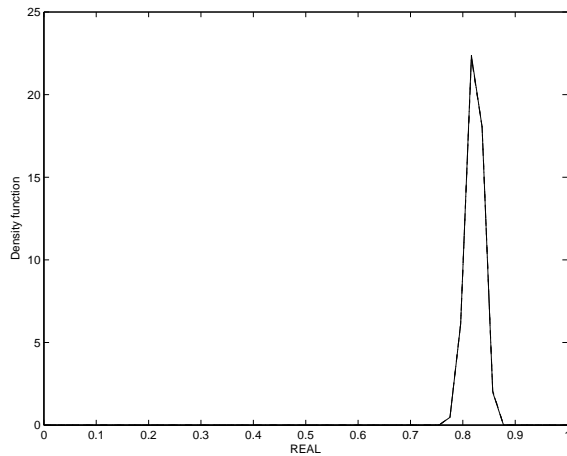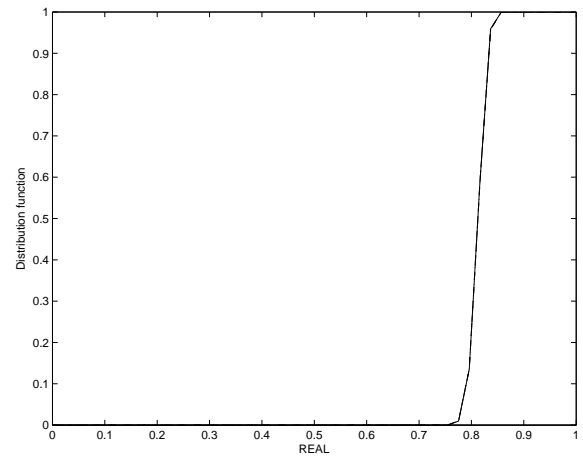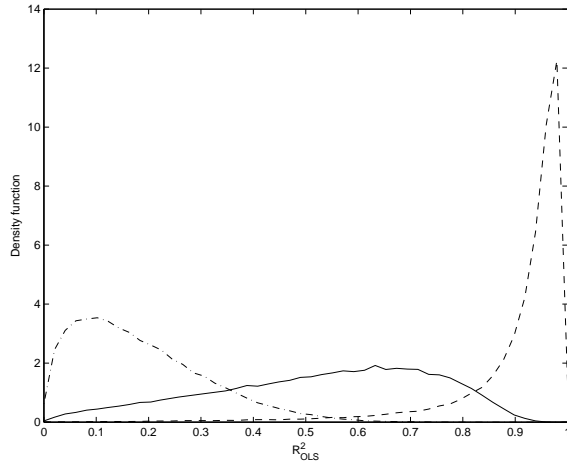
Figure 1.3. Density functions FACCHECK

Figure 1.4. Distribution functions FACCHECK

Panel 1. Density and distribution 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).

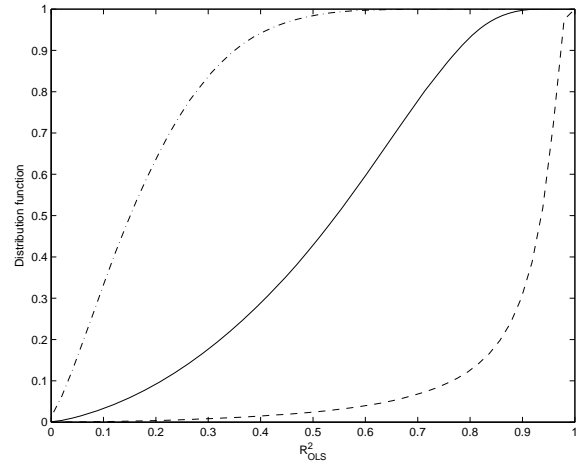Figure 2.1. Density functions of R^2_{OLS}

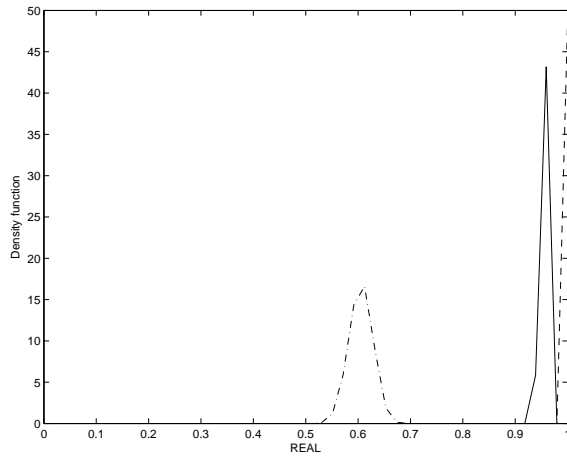Figure 2.2. Distribution functions of R^2_{OLS}

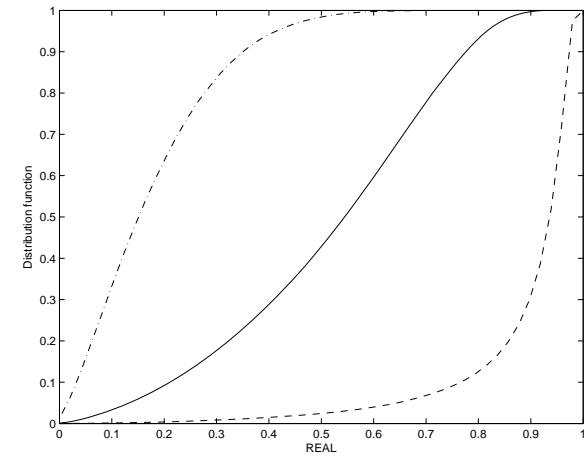Figure 2.3. Density functions FACCHECK

Figure 2.4. Distribution functions FACCHECK

Panel 2. Density and distribution 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. Distribution functions of R^2_{OLS}

Figure 3.3. Density functions FACCHECK

Figure 3.4. Distribution functions FACCHECK

Panel 3. Density and distribution 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).

24

Figure 4.1. Density functions of the R^2

Figure 4.2. Distribution functions of the R^2

Figure 4.3. Density functions FACCHECK

Figure 4.4. Distribution functions FACCHECK

Panel 4. Density and distribution 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).

The expression of the large distribution of R^2_{OLS} in Theorem 3 states the importance of the unexplained factor structure for R^2_{OLS}. This is again shown by the simulation results in Panels 1-3. It all shows that R^2_{OLS} cannot be interpreted appropriately without some diagnostic statistics that report 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. Figures 4.1 and 4.2 show that for the same irrelevant explanatory power of the observed factor model, R^2_{OLS} varies greatly. Figures 4.3 and 4.4 show that for the observed factor models where R^2_{OLS} is high in Figures 4.1 and 4.2 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%

Table 5: R-squared of the second pass regression of the FM two pass procedure and FACCHECK (the percentage of the variation explained by the three largest principal components) 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%

Table 6: R-squared of the second pass regression of the FM two pass procedure and FACCHECK (the percentage of the variation explained by the three largest principal components) 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} and our factor structure check (6) for the specifications in Tables 3 and 4. Many of the specifications stated in Table 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 R^2_{OLS}'s reported in these articles 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 $\bar{M} = V_{RR}^{-1} - V_{RR}^{-1}\iota_N(\iota_N'V_{RR}^{-1}\iota_N)^{-1}\iota_N'V_{RR}^{-1}$.

Theorem 3 shows that the large sample distribution of R^2_{OLS} depends on the explanatory power of the observed proxy factors for the true unobserved factors. In order to obtain an expression for the large sample distribution of R^2_{GLS} which applies to a large range of settings, we therefore also assume that the explanatory power of the observed proxy factors drifts with the sample size as stated in Assumption 1.

Alongside the explanatory power of the observed proxy factors, the large sample distribution of R^2_{GLS} crucially depends on the scaled risk premia of the unobserved true factors:

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

When we use the FF factors, the relative size of these size premia is small and proportional to the realization of a standard normal random variable. This is shown in Table 7 for the three data sets considered previously. Alongside the risk premia λ_F estimated using the FM two step estimator, Table 7 contains estimates of the relative risk premia in (24) and their quadratic form $\lambda_F'\left(\frac{V_{FF}}{T}\right)^{-1}\lambda_F$.

To construct the large sample distribution of R^2_{GLS}, we assume that the relative risk premia in (24) do not change with the sample size which is in line with their relatively small values reported in Table 7.

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*

$$
\left(\frac{V_{FF}}{T}\right)^{-\frac{1}{2}}\lambda_F = l,
\tag{25}
$$

with l a k dimensional fixed vector, for different values of 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) in large samples is characterized by:*

$$R^2_{GLS} \approx \cfrac{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' P_{M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}dV_{GG}^{\frac{1}{2}} \\ 0 \end{pmatrix}+\varphi^*\right\}}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}, \tag{26}$$

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

$$\left[(\beta'\beta)^{\frac{1}{2}}{}'V_{FF}(\beta'\beta)^{\frac{1}{2}} + (\beta'\beta)^{-\frac{1}{2}}{}'\beta'V_{\varepsilon\varepsilon}\beta(\beta'\beta)^{-\frac{1}{2}}\right] \tag{27}$$

and

$$M_{V_{RR}^{-\frac{1}{2}}\iota_N} \approx I_N - \begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta'\iota_N \\ (\beta'_\perp V_{\varepsilon\varepsilon}\beta_\perp)^{-\frac{1}{2}}\beta'_\perp\iota_N \end{pmatrix}$$
$$\{\iota'_N[\beta(\beta'\beta)^{-1}V_{FF}^{-1}(\beta'\beta)^{-1}\beta' + \beta_\perp(\beta'_\perp V_{\varepsilon\varepsilon}\beta_\perp)^{-1}\beta'_\perp]\iota_N\}^{-1} \begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta'\iota_N \\ (\beta'_\perp V_{\varepsilon\varepsilon}\beta_\perp)^{-\frac{1}{2}}\beta'_\perp\iota_N \end{pmatrix}', \tag{28}$$

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 A. ∎

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, the large sample behavior of R^2_{GLS} is characterized by*

$$R^2_{GLS} \approx \cfrac{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' P_{M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}V_{FG}V_{GG}^{-\frac{1}{2}} \\ 0 \end{pmatrix}+\frac{1}{\sqrt{T}}V_{RR}^{-\frac{1}{2}}V_{\varepsilon\varepsilon}^{\frac{1}{2}}\varphi^*\right\}}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' M_{V_{RR}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}, \tag{29}$$

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*

$$R^2_{GLS} \approx \frac{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}' P_{M_{V_{RR}^{-\frac{1}{2}} \iota_N}} \left\{\begin{pmatrix} W'\begin{pmatrix} I_m \\ \phi^*_{VG}/\sqrt{T} \\ 0 \end{pmatrix} \end{pmatrix} + \frac{1}{\sqrt{T}} V_{RR}^{-\frac{1}{2}} V_{\varepsilon\varepsilon}^{\frac{1}{2}} \varphi^*\right\} \left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}' M_{V_{RR}^{-\frac{1}{2}} \iota_N} \left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix} + \psi^*\right\}}, \tag{30}$$

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*

$$R^2_{GLS} \approx \frac{\begin{pmatrix} W'l \\ 0 \end{pmatrix}' P_{M_{V_{RR}^{-\frac{1}{2}} \iota_N}\left\{\begin{pmatrix} W'\begin{pmatrix} I_m \\ 0 \end{pmatrix} \\ 0 \end{pmatrix}\right\}} \begin{pmatrix} W'l \\ 0 \end{pmatrix}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}\right\}' M_{V_{RR}^{-\frac{1}{2}} \iota_N} \left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}\right\}}. \tag{31}$$

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.
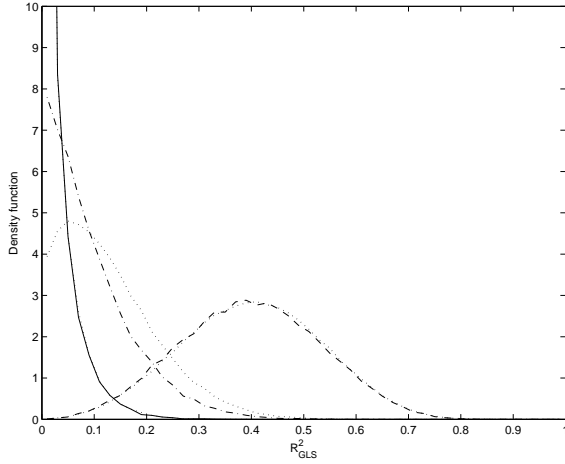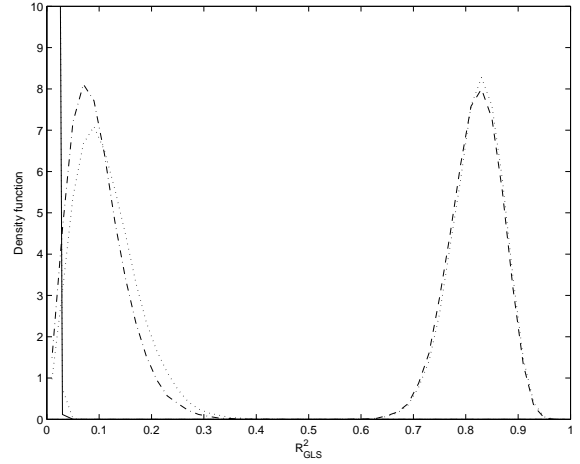
Figure 5.1. True factors, standard premia
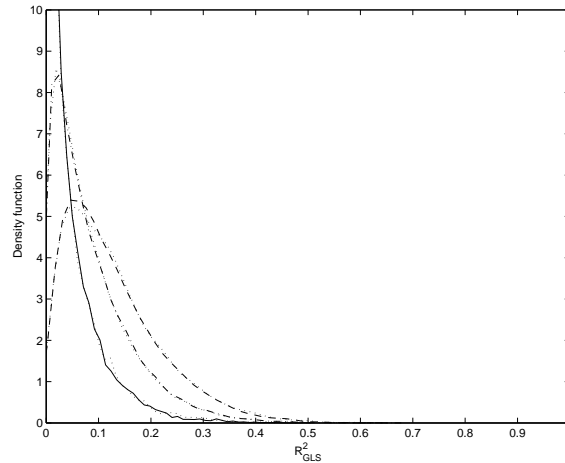
Figure 5.2. True factors, large premia
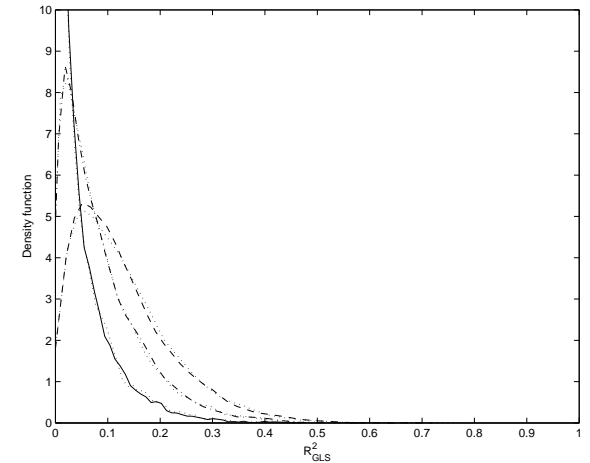
Figure 5.3. Irrelevant factors, standard premia

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.

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. For all three cases, the resulting density functions are almost indistinguishable. 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 lies close to one. Also its variance has decreased which applies to the density when we use two factors as well. This density has slightly shifted to the right.

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 tenfold the estimated ones. Unlike when we use the true factors, the larger risk premia have no effect on the density of R^2_{GLS}.

The density of R^2_{GLS} in case 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 expressions of the large sample distributions of R^2_{OLS} and R^2_{GLS} in Theorems 3 and 4 depend on the parameters d and l. Unless the observed factors explain the true unobserved factors well and the risk premia are large, we cannot estimate these parameters consistently. The latter is needed in order to be able to use the bootstrap to approximate the distribution of R^2_{OLS} or R^2_{GLS}. Hence, the bootstrap cannot be used to conduct inference on either R^2_{OLS} or R^2_{GLS}, when the observed factors do not explain the unobserved factors well.

3.2 Tests on the Risk Premia

We just showed that the large sample distributions of R^2_{OLS} or R^2_{GLS} are such that we can only unequivocally interpret them when there is no unexplained factor structure left in the first pass residuals. Alongside R^2_{OLS} and R^2_{GLS}, this argument applies to many other second pass statistics that are commonly used as well. The most prominent of these is the two pass t-statistic.

The FM two pass t-statistic testing the hypothesis that the i-th risk premia is equal to $\lambda_{i,0}$, $H_0 : \lambda_i = \lambda_{i,0}$, reads

$$t(\lambda_i) = \frac{\hat{\lambda}_i - \lambda_{i,0}}{\sqrt{var(\hat{\lambda}_i)}}, \tag{32}$$

with $\hat{\lambda}_F = (\hat{\lambda}_1 \dots \hat{\lambda}_m)'$ resulting from (13) and $var(\hat{\lambda}_i)$ the variance of the i-th element of $\hat{\lambda}_F$. The variance of $(\hat{\lambda}_0, \hat{\lambda}_F)$ equals

$$var(\hat{\lambda}_0, \ \hat{\lambda}_F) = \ \frac{1}{T} \left[(\iota_N \vdots \hat{B})'(\iota_N \vdots \hat{B}) \right]^{-1} (\iota_N \vdots \hat{B})' \hat{\Theta} (\iota_N \vdots \hat{B}) \left[(\iota_N \vdots \hat{B})'(\iota_N \vdots \hat{B}) \right]^{-1}, \tag{33}$$

with $\hat{\Theta} = \hat{V}_{\varepsilon\varepsilon}(1 + \hat{\lambda}'_F \hat{V}_{GG}^{-1} \hat{\lambda}_F)$, and the variance of the i-th element of $\hat{\lambda}_F$ equals the $(i+1)$-th diagonal element of this covariance matrix, see Shanken (1992). The FM two pass t-statistic has a standard normal distribution in large samples when the null hypothesis holds and the estimand of $\hat{B}$, $\beta\delta$, has a full rank value. This implies that both β and δ have to be sizeable full rank matrices. When the full rank assumption fails, for example, since Assumption 1 holds, the large sample distribution of the FM two pass t-statistic is not normal and we cannot use the FM two pass t-statistic to conduct inference, see Kleibergen (2009) and Kan and Zhang (1999). We have just shown that Assumption 1 applies to many empirically relevant settings so we then cannot use the FM two pass t-statistic to conduct inference.

Assumption 1 is similar to the weak instrument assumption made in linear instrumental variables regression, see $e.g.$ Staiger and Stock (1997). There is an extensive literature on weak instruments in econometrics which has led to the development of statistics whose large sample distributions are not affected by the strength of the instruments, see $e.g.$ Anderson and Rubin (1949), Kleibergen (2002), Moreira (2003) and Kleibergen and Mavroeidis (2009). Still, the traditional t-statistic is commonly used. Stock and Yogo (2005) have therefore analyzed the sensitivity of the large sample distribution of the t-statistic to the value of a pre-test for weak instruments that results from the first stage regression,

i.e. the first stage F-statistic. Stock and Yogo (2005) show that the large sample distribution of the t-statistic is approximately standard normal when the first stage F-statistic exceeds ten and differs from a normal distribution when the first stage F-statistic is less than ten.

The first stage F-statistic tests if the instruments used for the instrumental variables regression are jointly significant in the first stage regression. With respect to the FM two pass regression, this is analogous to testing if the observed proxy factors in the first pass regression are jointly significant. The LR statistics in Tables 2-4 test the significance of the observed proxy factors in the first pass regression. Since the first stage F-statistic is defined as an F-statistic, so it results from dividing by the number of tested parameters, a value of the first stage F-statistic of around ten is comparable with a value of the LR statistic that equals ten times the number of tested parameters. This means that the cutoffs for the LR statistic that allow us to use the FM t-statistic are two hundred and fifty in case of twenty five parameters, five hundred in case of fifty parameters, seven hundred and fifty in case of seventy five parameters etc.. The only two settings in Tables 2-4 for which the LR statistics exceed these thresholds are when there is one observed proxy factor which equals the value weighted return or when the observed proxy factors equal the FF factors. For all other specifications, either the LR statistic that tests the parameters of all observed proxy factors or the LR statistic that tests the value of the parameters of the observed proxy factors that are incremental to the value weighted return are below the threshold value. These specifications are obviously also those for which the observed factors leave an unexplained factor structure in the first pass residuals. For all these specifications, we should use the analogs of the weak instrument robust statistics for factor models developed in Kleibergen (2009). We use four of these identification robust factor statistics which are stated in Appendix B: the factor Anderson-Rubin (FAR) statistic, the factor extension of Kleibergen's (2002, 2005) Lagrange multiplier statistic (FKLM), the factor extension of Kleibergen's (2005) J-statistic (FJKLM) and the factor extension of Moreira's (2003) conditional likelihood ratio statistic (FCLR).

The four identificaton robust factor statistics that we use test a hypothesis on one element of λ_F, say $H_0 : \lambda_1 = \lambda_{1,0}$, in a different manner. The FAR statistic tests the joint hypothesis of factor pricing stated in (11) and H_0. The FJKLM statistic tests the hypothesis of factor pricing given that H_0 holds. The FKLM and FCLR statistics both just test H_0. We can use each of these four statistics to construct a 95% confidence set for λ_1 by specifying a grid of s different values for $\lambda_{1,0}$, $(\lambda_{1,0}^1 \ldots \lambda_{1,0}^s)$. We then compute the statistics for each different value of $\lambda_{1,0}$ in the grid. The 95% confidence set consists of all values of $\lambda_{1,0}$ for which the statistic is below its 95% critical value. These confidence sets are trustworthy even when Assumption 1 applies and show if the observed series are informative about the risk premia. If the latter is not the case, these confidence sets are unbounded which shows that the observed series do not contain (much) information about the pricing of the associated factors, see *e.g.* Dufour (1997). In the sequel, we discuss these confidence sets for two settings of the observed proxy factors which are representative for the results at large.

Lettau-Ludvigson (2001): consumption based and value weighted return factors We use two different specifications from Lettau and Ludvigson (2001). The first specification has three consumption based factors: consumption growth, the consumption-wealth ratio (*cay*) and the interaction between consumption growth and the consumption-wealth ratio. The second specification has the value weighted return as its single factor. The specifications correspond with the "*cay*, Δc, *cay*Δc" and "R_{vw}" columns in Table 3. Table 8 contains the FM two pass and maximum likelihood (ML) estimates, see Gibbons (1982).[10] We use the four identification robust factor statistics, the FM two pass t-statistic and the t-statistic based on the ML estimator to construct the confidence sets for the

---

[10] As discussed in Appendix B, the four identification robust statistics all have the ML estimator as their minimizer so we do not report separate estimators for these statistics.

risk premia. Figures 6.1-6.4 in Panel 6 each contain six (one minus the) p-value plots, one for every statistic. These figures also contain a straight line at 0.95 which enables us to construct the 95% confidence set using the intersection of the p-value plot with the line at 0.95.

For the first specification with the consumption based factors, Figures 6.1-6.3 show the p-value plots of the risk premia using the different statistics. The FM t-statistic is the statistic that is almost solely used in the literature. Figures 6.1-6.3 show that according to the FM t-statistic the risk premia on the consumption growth factor and consumption-wealth ratio are not significant at the 95% level while the risk premium on the interaction between consumption growth and the consumption-wealth ratio is. This can be concluded from the intersection of the p-value plot of the FM t-statistic and the line at 0.95. It also results from the risk premia estimates and standard errors in Table 8. Lettau and Ludvigson (2001) use the significant risk premium on the interaction between consumption growth and the consumption-wealth ratio to advocate an extension of the consumption capital asset pricing model in which the consumption growth's risk premium changes over the business cycle. These findings, however, do not correspond with those obtained from the identification robust factor statistics. For all three risk premia, the 95% confidence sets that result from any of the identification robust factor statistics are unbounded. The p-value plots of these statistics in Figures 6.1-6.3 do not intersect with the line at 0.95 or lie below it at large values of the risk premia. Hence, the 95% confidence sets that result from the identification robust factor statistics are unbounded. It shows that we cannot determine the risk premia. This results since the consumption based factors are not able to eradicate the factor structure as shown in Table 3. Because the consumption based factors leave a strong unexplained factor structure in the first pass residuals, δ is rather small as shown by the LR and F statistics in Table 3.[11] The risk premia can then not be identified which leads to a large range of plausible values as indicated by the unbounded confidence sets.

Figure 6.4 contains the (one minus the) p-value plots for the risk premium on the value weighted return when it is the only factor. They are also computed using the statistics discussed before. The results in Figure 6.4 correspond with the "R_{vw}" column in Table 3. Table 3 shows that the value weighted return eradicates one of the factors in the portfolio returns. The values of its δ and/or β parameters are therefore sizeable as revealed by the huge value of the likelihood ratio statistic testing the significance of these parameters, 765. Because of these large values, the p-value plots that result from the identification robust factor statistics and the FM and MLE t-statistics are almost identical and all indicate that the risk premium on the value weighted return is not significant at the 95% level.

The differences in the p-value plots of the identification robust factor statistics in Figures 6.1-6.4 result since they do not test the same hypothesis. The FKLM and FCLR statistics just test if the respective risk premium is equal to a specific value which explains why they are rather similar and equal to zero at the ML estimate. The FAR and FJKLM statistics also or just, in case of the FJKLM statistic, test the hypothesis of factor pricing which is stated in (11). This explains why the p-value plots associated with them are never equal to zero because the number of moment equations in (11) exceeds the number of risk premia.

In Figure 6.4, the p-value plots of the identification robust FAR and FJKLM statistics are not visible. This results because they are significant for all values of the risk premium. It shows that the hypothesis of factor pricing is rejected when we use the value weighted return as the only factor. Hence, the factor pricing moment equation in (11) does not hold, so the average portfolio returns are not just spanned by the β's of the value weighted return.

---

[11] For example, Table 3 shows that the p-value for testing the significance of consumption growth is only 6%.
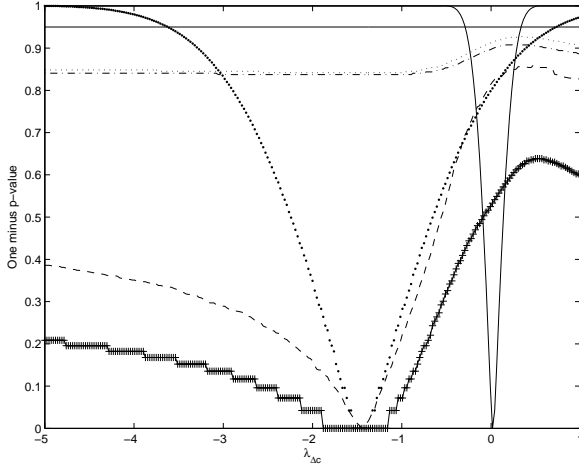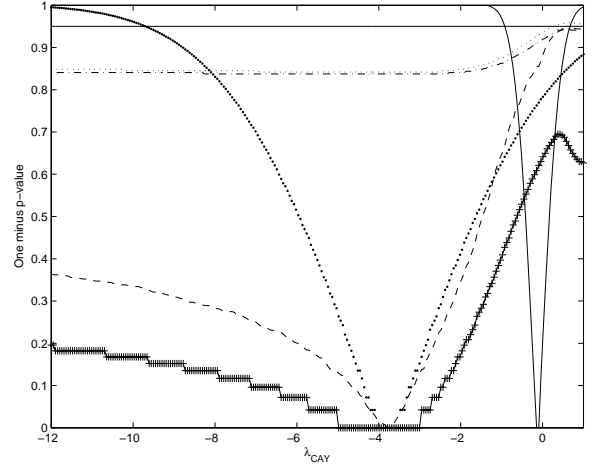
Figure 6.1: p-value plot $\lambda_{\Delta c}$
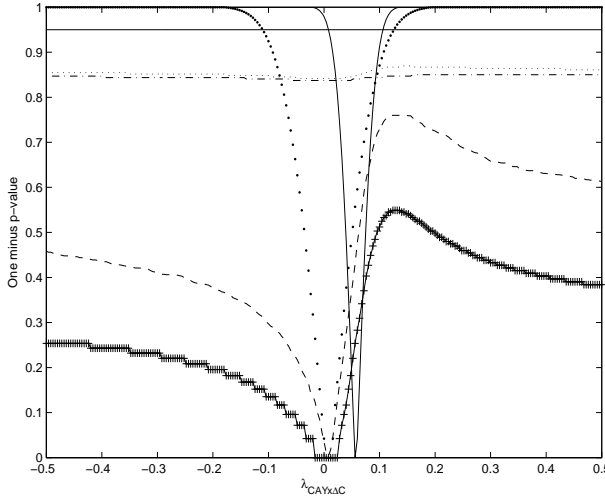

Figure 6.2: p-value plot λ_{cay}

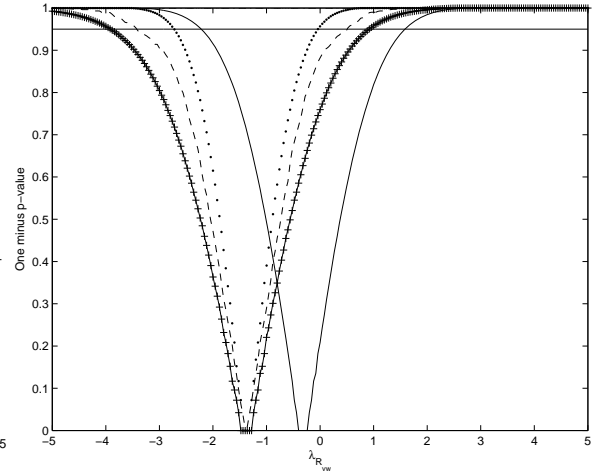Figure 6.3: p-value plot $\lambda_{\Delta c \times cay}$

Figure 6.4: p-value plot $\lambda_{r_{vw}}$

Panel 6. p-value plots for the risk premia that result from different specifications used in Lettau and Ludvigson (2001). Figures 6.1-6.3: three factor model. Figure 6.4 single factor model. FM two pass t-statistic (solid line), MLE t-statistic (points), FKLM (solid-plusses), FCLR (dashed), FJKLM (dash-dotted), FAR (dotted).

Figures 6.1-6.4 show the different conclusions reached when we use either the FM t-statistic (or ML t-statistic) or one of the identification robust factor statistics in case the explanatory power of the observed proxy factors is small. When the explanatory strength of the observed proxy factors is large, there is hardly any difference between the FM t-statistic and the identification robust factor statistics. This results since the standard normal large sample distribution of the FM t-statistic is then valid which is not the case when the explanatory power is minor. The large sample distributions of the identification robust factor statistics are valid in either case so the conclusions drawn from these statistics are valid for all possible explanatory strengths of the observed proxy factors.

We note that the ML estimates for the risk premia on the consumption based factors in Table 8 can be large and quite different from the FM two pass estimates. Table 3 shows that the β and/or δ parameters associated with the consumption based factors are all quite small. In the second pass,

		LL01		
	R_{vw}	Δc	cay	$cay\Delta c$
FM		0.022 0.16	−0.13 0.40	0.057 0.025
MLE		−1.44 1.10	−3.78 3.00	0.0077 0.06
FM	−0.31 0.96			
MLE	−1.38 0.68			

Table 8: FM two pass and ML estimates of the risk premia for different specifications used in Lettau and Ludvigson (2001). Standard errors (with Shanken correction for FM standard errors) are listed below the estimates.

we kind of divide by the estimated parameters from the first pass so their proximity to zero explains the large values of the ML estimates[12]. This also explains why the FM and ML estimates of the risk premium on the value weighted return in Table 8 are quite similar because its β and/or δ parameters are sizeable.

Remaining factor specifications The other specifications of the observed proxy factors can be classified along the lines of the two specifications discussed previously. First there are the specifications for which, according to Tables 2-4, the β's of some or all of the observed proxy factors are rather small and second are the specifications for which the β's of all observed proxy factors are large. The 95% confidence sets of the risk premia that result from the identification robust factor statistics are unbounded for the first setting. According to the two pass t-statistic many of these risk premia are, however, significant so the identification robust factor statistics again lead to different albeit but more credible conclusions. For the specifications where the β's of all observed proxy factors are large, the 95% confidence sets that result from both procedures are similar.

4 Conclusions

Portfolio returns exhibit an (unobserved) factor structure. This factor structure has to be captured appropriately by the observed proxy factors for the traditional statistical criteria used in the FM two pass procedure, like, for example, the OLS and GLS R^2 and second pass t-statistics, to be straightforward to interpret. When the observed proxy factors do not capture the factor structure appropriately, the OLS R^2 can still be large. This large value is, however, not indicative of the relationship between the expected portfolio returns and observed proxy factors since it results from the unexplained factor structure in the first pass residuals. For the GLS R^2, we show that its distribution is confined to rather small values both when the observed proxy factors are strong or minorly correlated with the true unobserved factors. This makes the GLS R^2 also not straightforward to interpret. For both of these goodness of fit measures, we therefore suggest to analyze them based on a measure of the unexplained factor structure in the first pass residuals. When this criteria indicates that there is no unexplained factor structure in the first pass residuals, both the OLS and GLS R^2 are straightforward to interpret. When there is, however, a considerable unexplained factor structure, we have to interpret

---

[12]For the same reason, the distribution of the ML estimator, just like the distribution of the ML estimator in the linear instrumental variables regression model, has no moments so the mean and variance of the distribution of the ML estimator are infinite. The same holds to a lesser extent for the FM two pass risk premia estimator for which we can show that its moments exist up to the order of the number of portfolios minus the number of factors which is around 25. Its distribution therefore has a finite mean and variance so extreme values of the FM risk premium estimator are less likely than for the ML estimator, see Kleibergen (2009).

them with caution.

Similar arguments apply to the second pass t-statistic whose large sample distribution differs from the standard normal one when the β's of the observed proxy factors are small. When these β's are small, an unexplained factor structure is present so it implies that both the second pass R^2's and t-statistics have to be interpreted with caution. The large sample distributions of the identification robust factor statistics of Kleibergen (2009) are not sensitive to the β's of the observed proxy factors so we recommend these to be used instead of the second pass t-statistic.

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 can reduce the factor structure in the portfolio returns, a sizeable factor structure typically remains present.

Appendix A

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 $\hat{V}_{VV} \underset{p}{\to} V_{VV}$, $\frac{1}{T} \sum_{t=1}^{T} \bar{G}_t \bar{G}'_t \underset{p}{\to} V_{GG}$, we have that

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

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

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

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

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

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 Likelihood ratio statistic equals

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

where we used that the restricted covariance matrix estimator,

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

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

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

since $\frac{1}{T}\sum_{t=1}^{T}\bar{G}_t\bar{G}'_t \underset{p}{\to} V_{GG}$, $\hat{\Sigma} \underset{p}{\to} V_{RR} = \beta V_{FF}\beta' + V_{\varepsilon\varepsilon}$, 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, $\text{vec}(A)$ is the column vectorization of the matrix A.

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 $\hat{B}$. The large sample behavior of $\hat{B}$ under Assumption 1 is stated in the proof of Theorem 2 while 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

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

so

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

where $\tfrac{1}{T}\sum_{t=1}^{T} F_t - \mu_{Ft} \underset{d}{\to} \psi_F$ and $\tfrac{1}{T}\sum_{t=1}^{T}\varepsilon_t \underset{d}{\to} \psi_\varepsilon$ with ψ_F and ψ_ε 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

$$
M_{\iota_N}\bar{R} \approx M_{\iota_N}\beta\lambda_F + \tfrac{1}{\sqrt{T}}(M_{\iota_N}\beta\psi_{\iota F} + M_{\iota_N}\psi_{\iota\varepsilon}).
$$

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:

$$
R^2_{OLS} \approx \frac{[\beta\lambda_F + \tfrac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]' P_{M_{\iota_N}(\beta(d+\psi_{VG})+\psi_{\varepsilon G})}[\beta\lambda_F + \tfrac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]}{[\beta\lambda_F + \tfrac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]' M_{\iota_N}[\beta\lambda_F + \tfrac{1}{\sqrt{T}}(\beta\psi_{\iota F} + \psi_{\iota\varepsilon})]}.
$$

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

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

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,

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

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

$$
Q = (\beta'\beta)^{-\frac{1}{2}} W
$$

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

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 $V_{RR}^{-\frac{1}{2}} V_{RR} V_{RR}^{-\frac{1}{2}'} = I_N$:

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

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

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

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

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

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

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

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

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

so

$$M_{V_{R\bar{R}}^{-\frac{1}{2}}\iota_N} \approx I_N - \begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta'\iota_N \\ (\beta'_\perp V_{\varepsilon\varepsilon}\beta_\perp)^{-\frac{1}{2}}\beta'_\perp \iota_N \end{pmatrix}$$
$$\{\iota'_N[\beta(\beta'\beta)^{-1}V_{FF}^{-1}(\beta'\beta)^{-1}\beta' + \beta_\perp(\beta'_\perp V_{\varepsilon\varepsilon}\beta_\perp)^{-1}\beta'_\perp]\iota_N\}^{-1} \begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}(\beta'\beta)^{-1}\beta'\iota_N \\ (\beta'_\perp V_{\varepsilon\varepsilon}\beta_\perp)^{-\frac{1}{2}}\beta'_\perp \iota_N \end{pmatrix}'.$$

The specification of GLS R^2 reads

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

We proceed with constructing expressions for the large sample behavior of the components of the GLS R^2 : $V_{R\bar{R}}^{-\frac{1}{2}}\bar{R}$ and $V_{R\bar{R}}^{-\frac{1}{2}}\hat{B}$ 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:

$$\bar{R} \approx \iota_N \lambda_0 + \beta(\lambda_F + \tfrac{1}{\sqrt{T}}\psi_F) + \tfrac{1}{\sqrt{T}}\psi_\varepsilon = \iota_N \lambda_0 + \beta V_{FF}(V_{FF}^{-\frac{1}{2}}\lambda_F) + \tfrac{1}{\sqrt{T}}(\beta\psi_F + \psi_\varepsilon).$$

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:

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

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

$\mathbf{V}_{RR}^{-\frac{1}{2}}\hat{\mathbf{B}}.$ 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:

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

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:

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

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:

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

with $\varphi^* = V_{RR}^{-\frac{1}{2}}(\beta\psi_{VG} + \psi_{\varepsilon G})V_{GG}^{\frac{1}{2}}$ 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 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:

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

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

Weak factors:

$$
R^2_{GLS} \approx \frac{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' P_{M_{V_{R\bar{R}}^{-\frac{1}{2}}\iota_N}}\left\{\begin{pmatrix} W'V_{FF}^{-\frac{1}{2}}dV_{G\bar{G}}^{\frac{1}{2}} \\ 0 \end{pmatrix}+\varphi^*\right\}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}{\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}' M_{V_{R\bar{R}}^{-\frac{1}{2}}\iota_N}\left\{\begin{pmatrix} W'l \\ 0 \end{pmatrix}+\psi^*\right\}}.
$$

Appendix B. Identification robust factor statistics

To test a hypothesis on one element of λ_F, say λ_1, the identification robust factor statistics first remove the zero-β return λ_0 by taking all portfolio returns in deviation from a baseline portfolio return, say $r_{t,N}$:

$$
\mathcal{R}_t = R_{1t} - \iota_{N-1}r_{Nt}
$$

with $R_{1t} = (r_{1t} \ldots r_{N-1t})'$ and ι_{N-1} is an $(N-1)$-dimensional vector of ones. The identification robust factor statistics are invariant with respect to the choice of the baseline portfolio and also with respect to other transformations.

If λ_F has more than one element and we want to test the hypothesis, $H_0 : \lambda_1 = \lambda_{1,0}$, we first compute the maximum likelihood estimates under H_0 of the other elements of $\lambda_F : \lambda_2 \ldots \lambda_m$, see Gibbons (1982). These result from the characteristic vectors (eigenvectors) of a characteristic polynomial for which we use the residuals of three linear regressions:

$$
\begin{aligned}
\mathcal{R}_t &= \alpha_1(\bar{G}_{1t} + \lambda_{1,0}) + \tilde{\mathcal{R}}_t \\
1 &= \alpha_2(\bar{G}_{1t} + \lambda_{1,0}) + \tilde{1}_t \\
\bar{G}_{2t} &= \alpha_3(G_{2t} + \lambda_{1,0}) + \tilde{G}_{2t}
\end{aligned} \tag{34}
$$

where $G_t = \begin{pmatrix} G_{1t} \\ G_{2t} \end{pmatrix}$ with G_{1t} a scalar and G_{2t} an $(m-1)$-dimensional vector, G_t is the vector with observed factors in (7). When we estimate α_1, α_2 and α_3 using least squares and $\tilde{\mathcal{R}}_t$, $\tilde{1}_t$ and $\tilde{G}_{2t}$ are the residuals of these least squares regressions, the characteristic polynomial from which we obtain the maximum likelihood estimators reads

$$
\left| \theta \left[\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\tilde{1}_t \\ \tilde{G}_{2t}\end{pmatrix}\begin{pmatrix}\tilde{1}_t \\ \tilde{G}_{2t}\end{pmatrix}'\right] - \left[\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\tilde{1}_t \\ \tilde{G}_{2t}\end{pmatrix}\tilde{\mathcal{R}}_t'\right]\Sigma^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathcal{R}}_t\begin{pmatrix}\tilde{1}_t \\ \tilde{G}_{2t}\end{pmatrix}'\right]\right| = 0, \tag{35}
$$

with $\Sigma = \frac{1}{T}\sum_{t=1}^{T}\tilde{\mathcal{R}}_t\tilde{\mathcal{R}}_t'$. The maximum likelihood estimator of $\lambda_{F,2} = (\lambda_2 \ldots \lambda_m)'$ then equals

$$\tilde{\lambda}_{F,2} = W_2^{-1\prime}w_1', \tag{36}$$

with w_1 an $m-1$ dimensional row vector and W_2 an $(m-1)\times(m-1)$ dimensional matrix which are such that $\binom{w_1}{W_2}$ contains the $m-1$ characteristic vectors that are associated with the $m-1$ largest roots of the characteristic polynomial in (35), see Kleibergen (2009) and Shanken and Zhou (2007). We then proceed with computing the least squares estimator of B under $\mathrm{H_0}$[13] :

$$\hat{B} = \sum_{t=1}^{T}\mathcal{R}_t\binom{\bar{G}_{1t}+\lambda_{1,0}}{\bar{G}_{2t}+\tilde{\lambda}_{F,2}}' \left[\sum_{t=1}^{T}\binom{\bar{G}_{1t}+\lambda_{1,0}}{\bar{G}_{2t}+\tilde{\lambda}_{F,2}}\binom{\bar{G}_{1t}+\lambda_{1,0}}{\bar{G}_{2t}+\tilde{\lambda}_{F,2}}'\right]^{-1}, \tag{37}$$

which provides the base for our four identification robust factor statistics, see Kleibergen (2009):

1. The factor Anderson-Rubin (FAR) statistic which is based on the Anderson-Rubin statistic, see Anderson and Rubin (1949):

$$\mathrm{FAR}(\lambda_{1,0}) = \frac{T}{1-\binom{\lambda_{1,0}}{\tilde{\lambda}_{F,2}}'\hat{Q}_{GG}^{-1}\binom{\lambda_{1,0}}{\tilde{\lambda}_{F,2}}}\left[\bar{\mathcal{R}} - \hat{B}\binom{\lambda_{1,0}}{\tilde{\lambda}_{F,2}}\right]'\tilde{\Sigma}^{-1}\left[\bar{\mathcal{R}} - \hat{B}\binom{\lambda_{1,0}}{\tilde{\lambda}_{F,2}}\right], \tag{38}$$

with $\bar{\mathcal{R}} = \frac{1}{T}\sum_{t=1}^{T}\mathcal{R}_t$, $\hat{Q}_{GG} = \frac{1}{T}\sum_{t=1}^{T}\binom{\bar{G}_{1t}+\lambda_{1,0}}{\bar{G}_{2t}+\tilde{\lambda}_{F,2}}\binom{\bar{G}_{1t}+\lambda_{1,0}}{\bar{G}_{2t}+\tilde{\lambda}_{F,2}}'$ and $\tilde{\Sigma} = \frac{1}{T-k}\sum_{t=1}^{T}\left[\bar{\mathcal{R}}_t - \hat{B}\binom{\bar{G}_{1t}+\lambda_{1,0}}{\bar{G}_{2t}+\tilde{\lambda}_{F,2}}\right]$ $\left[\bar{\mathcal{R}}_t - \hat{B}\binom{\bar{G}_{1t}+\lambda_{1,0}}{\bar{G}_{2t}+\tilde{\lambda}_{F,2}}\right]'$ and whose large sample distribution is bounded by a $\chi^2(N-m)$ distributed random variable when the sample size T gets large.

2. The factor extension of Kleibergen's (2002, 2005) Lagrange Multiplier statistic:

$$\mathrm{FKLM}(\lambda_{1,0}) = \frac{T}{1-\binom{\lambda_{1,0}}{\tilde{\lambda}_{F,2}}'\hat{Q}_{GG}^{-1}\binom{\lambda_{1,0}}{\tilde{\lambda}_{F,2}}}\left[\bar{\mathcal{R}} - \hat{B}\binom{\lambda_{1,0}}{\tilde{\lambda}_{F,2}}\right]'\tilde{\Sigma}^{-1}\hat{B}(\hat{B}'\tilde{\Sigma}^{-1}\hat{B})^{-1}\hat{B}'\tilde{\Sigma}^{-1}\left[\bar{\mathcal{R}} - \hat{B}\binom{\lambda_{1,0}}{\tilde{\lambda}_{F,2}}\right],$$
$$\tag{39}$$

whose large sample distribution is bounded by a $\chi^2(1)$ distributed random variable when the sample size gets large.

3. The factor extension of Kleibergen's (2005) J-statistic:

$$\mathrm{FJKLM}(\lambda_{1,0}) = \mathrm{FAR}(\lambda_{1,0}) - \mathrm{FKLM}(\lambda_{1,0}), \tag{40}$$

whose large sample distribution is bounded by a $\chi^2(N-m-1)$ distributed random variable when the sample size gets large. This $\chi^2(N-m-1)$ distributed random variable is independent of the $\chi^2(1)$ distributed random variable which bounds the large sample distribution of the FKLM statistic.

4. The factor extension of Moreira's (2003) conditional likelihood ratio statistic:

$$\mathrm{FCLR}(\lambda_{1,0}) = \tfrac{1}{2}\left[\mathrm{FKLM}(\lambda_{1,0}) + \mathrm{FJKLM}(\lambda_{1,0}) - \mathrm{r}(\lambda_{1,0}) + \sqrt{(\mathrm{FKLM}(\lambda_{1,0}) + \mathrm{FJKLM}(\lambda_{1,0}) + \mathrm{r}(\lambda_{1,0}))^2 - 4\mathrm{r}(\lambda_{1,0})\mathrm{FJKLM}(\lambda_{1,0})}\right], \tag{41}$$

with $\mathrm{r}(\lambda_{1,0})$ the smallest root of the characteristic polynomial:

$$\left|\mu\hat{Q}_{GG}^{-1} - \hat{B}'\tilde{\Sigma}^{-1}\hat{B}\right| = 0. \tag{42}$$

---

[13]We are essentially estimating $B_1 - \iota_{N-1}b_N$ if $B = \binom{B_1}{b_N}$, with B_1 an $(N-1)\times m$ dimensional matrix and b_N a $1\times m$ dimensional row vector.

In large samples the distribution of the FCLR statistic is bounded by a random variable whose distribution is conditional on the value of r($\lambda_{1,0}$). The (bounding) critical values of the FCLR statistic are therefore a function of r($\lambda_{1,0}$) and the independent $\chi^2(1)$ and $\chi^2(N-m-1)$ large sample distributions of the FKLM and FJKLM statistics. We obtain these critical values by fixing r($\lambda_{1,0}$) and simulating the FKLM and FJKLM statistics from their independent large sample distributions to compute the FCLR statistic using the simulated values of the FKLM and FJKLM statistics, see Kleibergen (2009).

The critical values that result from the bounding distributions for each of the four identification robust factor statistics are such that these statistics are size correct when ($\lambda_2 \ldots \lambda_m$) are well identified so the associated values of β and V_{FG} constitute sizeable full rank matrices. When ($\lambda_2 \ldots \lambda_m$) are not well identified because of small or zero values of the associated values of β and V_{FG}, the critical values are such that the rejection frequencies of the identification robust factor statistics are smaller than the size of the test. The maximal rejection frequencies of the identification robust factor statistics are therefore equal to the size of the test which makes them size correct, see Kleibergen and Mavroeidis (2009).

The identification robust factor statistics test different hypotheses. The FAR statistic tests the joint hypothesis of factor pricing and $\lambda_1 = \lambda_{1,0}$: $\text{H}_{\text{FAR}} : E(\mathcal{R}_t) = B\binom{\lambda_{1,0}}{\lambda_{F,2}}$ where $\lambda_{F,2}$ is to be estimated. The hypothesis of factor pricing given that $\lambda_1 = \lambda_{1,0}$ is tested using the FJKLM statistic while the FKLM and FCLR statistics both test $\text{H}_0 : \lambda_1 = \lambda_{1,0}$.

We can use each of the four different statistics to construct a 95% confidence set for λ_1 by specifying a grid of s different values for $\lambda_{1,0}$, $(\lambda_{1,0}^1 \ldots \lambda_{1,0}^s)$. We then compute the statistics for each different value of $\lambda_{1,0}$ in the grid. The 95% confidence set consists of all values of $\lambda_{1,0}$ for which the statistic is below its 95% (conditional) critical value. In case that one of the elements of λ_F is not well identified, the 95% confidence sets is unbounded, see *e.g.* Dufour (1997).

The above discussion deals with testing one element of λ_F so it assumes that λ_F has more than one element, $m > 1$. When λ_F just consists of one element, $\lambda_{F,2}$ is not present so we do not conduct the computations in (34)-(36). The expressions of the identification robust factor statistics then extend to the case that there is just one observed factor when we remove $\tilde{\lambda}_{F,2}$ and $\bar{G}_{2t} + \tilde{\lambda}_{F,2}$ from (37)-(42).

The performance of the identification robust FKLM and FCLR statistics is similar to that of t-statistics based on FM two pass or maximum likelihood estimators when the β's and V_{FG} are sizeable. The latter two, however, become unreliable when the β's and/or V_{FG} are quite small. The identification robust factor statistics remain reliable in these cases. It is therefore important to use them for computing confidence sets on the risk premia when the β's and/or V_{FG} are quite small.

References

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

[2] Anderson, T.W. and H. Rubin. Estimation of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics*, **21**:570–582, 1949.

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

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

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

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

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

[8] 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.

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

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

[11] Dufour, J.-M. Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, **65**:1365–388, 1997.

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

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

[14] 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.

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

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

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

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

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

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

[21] Kleibergen, F. Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, **70**:1781–1803, 2002.

[22] Kleibergen, F. Testing Parameters in GMM without assuming that they are identified. *Econometrica*, **73**:1103–1124, 2005.

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

[24] Kleibergen, F. and S. Mavroeidis. Weak instrument robust tests in GMM and the new Keynesian Phillips curve. *Journal of Business and Economic Statistics*, **27**:293–311, 2009.

[25] 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.

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

[27] 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.

[28] 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.

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

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

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

[32] Moreira, M.J.,. A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, **71**:1027–1048, 2003.

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

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

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

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

[37] Shanken, J. and G. Zhou. Estimating and testing beta pricing models: Alternative Methods and their performance in simulations. *Journal of Financial Economics*, **84**:40–86, 2007.

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

[39] 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.

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