Abstract

Standard factor models in asset pricing imply a linear relationship between expected returns on assets and their exposures to one or more risk factors. We show that exploiting this linear relationship leads to statistical gains of up to 31% in the variances of estimated expected returns on individual assets compared to using historical averages. When the factors are weakly correlated with assets, i.e., the $\beta$'s are small, the generalized method of moment estimators of risk premiums still leads to reliable inference, i.e., inference does not suffer from lack of identification. If the factor model is misspecified in the sense of an omitted factor, we show that estimates based on this factor model may be inconsistent. Adding an alpha to the model capturing mispricing only leads to consistent risk premium estimates in case of traded factors. A simulation experiment documents that the more precise estimates of expected returns based on factor models, rather than historical averages, translates into significant improvements in the out–of–sample performance of optimal portfolios.

Keywords: Factor Pricing, Risk–Return Models, Risk Premium, Omitted Factors

JEL: C13, C38, G11

1. Introduction

Estimating expected returns on individual assets or portfolios is perhaps one of the longest standing challenges in finance. The standard approach at hand is to use historical

We thank Torben G. Andersen, Lieven Baele, Jules van Binsbergen, Svetlana Bryzgalova, Joost Driessen, Joachim Grammig, Thijs van der Heijden, Frank de Jong, Frank Kleibergen, Tristan Linke, Andrew Patton, Eric Renault, Enrique Sentana, George Tauchen, Viktor Todorov, Brian Weller and participants at the SoFiE conference 2015 for their helpful comments.

January 30, 2017
averages. However, it is known that these estimates are generally very noisy. Even using daily data does not help much, if at all.

The asset pricing literature provides a wide variety of linear factor models motivating certain risks that explain the cross section of expected returns on individual assets. Examples include Sharpe (1964)’s CAPM, Merton (1973)’s ICAPM, Breeden (1979)’s CCAPM, Ross (1976a,b)’ APT and Lettau and Ludvigson (2001)’s conditional CCAPM, among many others. These models all imply that expected excess returns on individual assets are linear in their exposures to certain risk factors. The coefficients in this linear relationship are the prices of risk for these factors. The literature on factor models mainly concentrates on determining these prices of risk and evaluating the ability of the models in explaining the cross section of expected returns.

In our study, the focus is different: we analyze the estimation of the expected (excess) returns on individual assets or portfolios, i.e., the product of exposures ($\beta$) and risk prices ($\lambda$). The tremendous literature on asset pricing provides potential estimators of the expected returns on individual assets and, as mentioned by Black (1993), these theoretical restrictions can help to improve the estimates of expected returns.

Estimating expected returns using factor models is not a new idea and was, to our knowledge, first suggested by Jorion (1991). In his empirical analysis, he compares CAPM—based estimators with classical sample averages of past returns finding the former outperforming the latter in estimating expected stock returns for his data. Our paper complements this work by providing the first detailed asymptotic efficiency analysis for both estimators. We additionally evaluate the implications of weakly correlated and omitted factors in the estimation of expected (excess) returns.

First, we derive the asymptotic statistical properties of expected excess return, that is risk premium, estimators based on factor models. These limiting distributions are useful for obtaining the standard errors and, accordingly, the confidence bounds of the risk–premium estimators for individual assets or portfolios. We thereby assess the precision gains from using factor–model based risk premium estimators vis–à–vis the historical
averages approach. In particular, we provide closed form asymptotic expressions for these efficiency gains. We show that exploiting the linear relationship implied by linear factor models indeed leads to more precise estimates of expected excess returns as compared to historical averages, see Corollaries 4.1 and 4.2. In an empirical analysis, for instance when estimating risk-premiums on 25 Fama and French (1992) size and book-to-market sorted portfolios, we document reductions in estimated variances of up to 31% for individual portfolios.

Second, we analyze the estimation of risk premiums in the presence of weakly correlated factors. When factors are weakly correlated with assets, i.e. $\beta$'s are small, the standard confidence intervals of the price of risk estimates are known to be erroneous (see, e.g., Kleibergen (2009)). This effect may be severe in empirical research, as these confidence intervals may be unbounded as documented for the case of consumption CAPM of Lettau and Ludvigson (2001)\(^1\). We demonstrate that such issues do not exist if the object of interest are the risk premiums on individual assets rather than the prices of risk attached to the factors. In particular, the limiting variances of the risk premium estimators are not affected by the $\beta$'s being small, see Corollary 5.1.

Third, we consider the issue of estimating risk premiums in the ubiquitous situation of omitted factors in the specification of the linear factor model. After the Capital Asset Pricing Model (CAPM) had been substantially criticized, researchers have come up with new risk factors to help explaining the cross section of expected returns. See, e.g., Fama and French (1993), Lettau and Ludvigson (2001), Lustig and Van Nieuwerburgh (2005), Li, Vassalau and Xing (2006), and Santos and Veronesi (2006), to name just a few. While it is doubtful that “the correct” factors have been found, the literature points to the existence of missing factors. We show that when a model is misspecified, in the sense that a relevant pricing factor is omitted, standard methods will generally not even provide consistent estimates of risk premiums on the individual assets or portfolios (see Theorem 6.1). However, we also show that adding an alpha, capturing the misspecification,

\(^1\)See also Bryzgalova (2014), Burnside (2015) on the role of weakly identified factors in inference about the prices of risk.
leads to a consistent estimator but only in the case the factors are traded. However, then there is no longer an efficiency gain over historical averages. Thus, our paper documents precisely the trade-off any empirical researcher faces: allow for misspecification and loose efficiency or run the risk of misspecification and gain efficiency.

Expected returns are not only interesting in the sense of single quantities for individual assets but they are also crucial inputs for theoretical formulations in various subfields of finance, i.e. calculations of cost of capital or valuation of cash flows. From an asset pricing perspective, the most prominent presence of expected returns is in portfolio allocation problems. We analyze the economic implications of the efficiency gains from using factor model based estimates of expected returns in Markowitz (1952)’s setting.

Implementation of the mean–variance framework of Markowitz (1952) in practice requires the estimation of the first two moments of asset returns. Constructing optimal portfolios with imprecise estimates of expected returns, using historical averages, and the sample covariance matrix generally lead to poor out of sample performance, see, for example, Frost and Savarino (1988), Michaud (1989), Jobson and Korkie (1980, 1981), Best and Grauer (1991), and Litterman (2003). In the far end, this has led to simply abandoning the application of theoretically optimal decisions and using naive techniques such as the 1/N strategy or the global minimum portfolio as these are not subject to estimation risk on expected returns (DeMiguel et al. (2009b)).

The mean–variance optimal portfolio weights can also be constructed with more precise factor–based risk–premium estimates instead of the using historical averages. Accordingly, we investigate the out–of–sample performance of optimal portfolios constructed using factor–model based risk premium estimates in a simulation analysis. We document that the average out–of–sample Sharpe ratios of the optimal portfolios improve strikingly. Specifically, we observe an improvement of up to 64% in the average out–of–sample Sharpe ratios

---

2Several studies provide solutions on improving the covariance matrix estimates (see, e.g., Ledoit and Wolf (2003), DeMiguel et al. (2009a) among others). However, the estimation error in asset return means is more severe than error in covariance estimates (see Merton (1980), Chopra and Ziemba (1993)) and the imprecision in estimates of the expected returns has a much larger impact on the optimal portfolio weights compared to the imprecision in covariance estimates (DeMiguel et al. (2009b)).
if factor–model based estimates are used as estimators of risk premiums instead of historical averages. Moreover, optimal portfolios constructed using the factor–model based estimates perform better than both the global minimum variance portfolio and 1/N strategy portfolio. Lastly, the average out–of–sample Sharpe ratios of the factor–model based optimal portfolios are much more precise and significant compared to the ones based on historical averages.

The rest of the paper is organized as follows. Section 2 introduces our set–up and presents the linear factor model with the assumptions that form the basis of our statistical analysis. Next, we introduce factor–mimicking portfolios and clarify the link between the expected return obtained with non–traded factors and with factor–mimicking portfolios. Section 3 discusses in detail the standard GMM estimators we consider. In particular, we recall the different sets of moment conditions for various cases such as all factors being traded or using factor–mimicking portfolios. Section 4 derives the asymptotic properties of these induced GMM estimators. In particular, we derive the efficiency gains over and above the risk–premium estimator based on historical averages. Section 5 presents the analysis for the small $\beta$s. Section 6 adresses the question of omitted factors. Section 7 presents our simulation analysis for portfolio optimization, and Section 8 concludes. All proofs are gathered in the appendix.

2. Model

It is well known that, in the absence of arbitrage, there exists a stochastic discount factor $M$ such that for any traded asset $i = 1, 2, \ldots, N$ with excess return $R_i^e$

$$
E[M R_i^e] = 0. \quad (2.1)
$$

Linear factor models additionally specify $M = a + b'F$, where $F = (F_1, ..., F_K)'$ is a vector of $K$ factors (see, e.g., Cochrane (2001), p.69). Note that (2.1) can be written in matrix notation using the vector of excess returns $R_e = (R_1^e, ..., R_N^e)'$. Throughout we impose the following.
Assumption 1. The $N$–vector of excess asset returns $R^e$ and the $K$–vector of factors $F$ with $K < N$ satisfy the following conditions:

(a) The covariance matrix of excess returns $\Sigma_{R^eR^e}$ has full rank $N$,
(b) The covariance matrix of factors $\Sigma_{FF}$ has full rank $K$,
(c) The covariance between excess returns and factors, $\text{Cov}[R^e, F']$, has full rank $K$.

Given the linear factor model and Assumption 1, it is classical to show

$$\text{E}[R^e] = \beta \lambda,$$

where

$$\beta = \text{Cov}[R^e, F'] \Sigma_{FF}^{-1},$$
$$\lambda = -\frac{1}{\text{E}[M]} \Sigma_{FF} b.$$  

Thus, (2.2) specifies a linear relationship between risk premiums on individual assets, $\text{E}[R^e]$, and their exposures $\beta$ to the risk factors, $F$. The vector $\lambda$ denoted the so-called prices of risk of the factors.

In empirical work, we need to make assumptions about the time–series behavior of consecutive returns and factors. In this paper, we focus on the simplest, and most used, setting where returns are i.i.d. over time. Express the excess asset returns

$$R^e_t = \alpha + \beta F_t + \varepsilon_t, \quad t = 1, 2, \ldots, T,$$

where $\alpha$ is an $N$–vector of constants, $\varepsilon_t$ is an $N$–vector of idiosyncratic errors and $T$ is the number of time–series observations. We then, additionally, impose the following.

Assumption 2. The disturbance $\varepsilon_t$ and the factors $F_t$, are independently and identically distributed over time with

$$\text{E}[\varepsilon_t | F_t] = 0,$$
$$\text{Var}[\varepsilon_t | F_t] = \Sigma_{\varepsilon\varepsilon},$$

where $\Sigma_{\varepsilon\varepsilon}$ has full rank.

2.1. Factor–Mimicking Portfolios

A large number of studies in the asset pricing literature suggest “macroeconomic” factors that capture systematic risk. Examples include the C-CAPM of Breeden (1979), the I-
CAPM of Merton (1973), and the conditional C-CAPM of Lettau and Ludvigson (2001). In order to assess the validity of macroeconomic risk factors being priced or not, it has been suggested to refer to alternative formulations of such factor models replacing the factors by their projections on the linear span of the excess returns. This is commonly referred to as factor–mimicking portfolios and early references go back to Huberman (1987) (see also, e.g., Fama (1998), Kandel and Stambaugh (1995), and Lamont (2001), Balduzzi and Robotti (2008)). We analyze, in this paper, the effect of such formulations on the estimation of risk premiums and we show, in Section 4, that there are efficiency gains from the information in mimicking portfolios when estimating risk premiums.

It is important to understand that, the prices of risk of (non–traded) factors generally differ from the risk–premiums on their factor–mimicking portfolios. However, using factor–mimicking portfolios leads to identical risk premiums on individual assets. This is shown in Theorem 2.1. To be precise, we project the factors \( F_t \) onto the space of excess asset returns, augmented with a constant. In particular, given Assumption 1, there exists a \( K \)-vector \( \Phi_0 \) and a \( K \times N \) matrix \( \Phi \) of constants and a \( K \)-vector of random variables \( u_t \) satisfying

\[
F_t = \Phi_0 + \Phi R^e_t + u_t, \tag{2.8}
\]

\[
E[u_t] = 0_{K \times 1}, \tag{2.9}
\]

\[
E[u_t R^e_t'] = 0_{K \times N}, \tag{2.10}
\]

We then define the factor–mimicking portfolios by

\[
F^m_t = \Phi R^e_t. \tag{2.11}
\]

Now, we obtain an alternative formulation of the linear factor model by replacing the original factors by their factor–mimicking portfolios

\[
R^e_t = \alpha^m + \beta^m F^m_t + \varepsilon^m_t, \quad t = 1, 2, \ldots, T. \tag{2.12}
\]
Recall that, using the projection results, $\Phi$ and $\beta$ are related by

$$\Phi = \Sigma_{FF}\beta'\Sigma^{-1}_{R_t R_t},$$  \hspace{1cm} (2.13)

while $\beta^m$ and $\beta$ satisfy

$$\beta^m = \beta \left( \beta'\Sigma^{-1}_{R_t R_t}\beta \right)^{-1}\Sigma^{-1}_{FF}.\hspace{1cm} (2.14)$$

The following theorem recalls that, while factor loadings and prices of risk change when using factor mimicking portfolios, expected (excess) returns on individual assets, i.e., their product, are not affected. For completeness we provide a proof in the appendix.

**Theorem 2.1.** *Under Assumptions 1 and 2, we have $\beta\lambda = \beta^m\lambda^m$ with $\lambda^m = E[F_t^m]$.***

Note that since the factor–mimicking portfolio is an excess return, asset pricing theory implies that the price of risk attached to it, $\lambda^m$, equals its expectation. This can be imposed in the estimation of expected (excess) returns and one may hope that the expected (excess) return estimators obtained with factor–mimicking portfolios are more efficient than the expected (excess) return estimators obtained with the non-traded factors themselves. We study this question in Section 4.

3. Estimation

We concentrate on Hansen (1982)’s GMM estimation technique. The GMM approach is particularly useful in our paper as it avoids the use of two-step estimators and the resulting “errors-in-variables” problem when calculating limiting distributions. In addition, we immediately obtain the joint limiting distribution of estimates for $\beta$ and $\lambda$ which is needed as we are interested in their product.

In the following sections, we study the asymptotics of the expected (excess) return estimators by specifying different sets of moment conditions. In Section 3.1, we study a set of moment conditions which generally holds, i.e., both when factors are traded and when they are non-traded. In Section 3.2, we study the case where all factors are traded. We then incorporate the moment condition that factor prices equal expected
factor values. In Section 3.3, we consider expected (excess) return estimates based on factor-mimicking portfolios.

3.1. Moment Conditions - General Case

We first provide the moment conditions for a general case, i.e., where factors may represent excess returns themselves, but not necessarily. In that case, the standard moment conditions to estimate both factor loadings \( \beta \) and factor prices \( \lambda \) are

\[
E[ht(\alpha, \beta, \lambda)] = E \left[ \begin{bmatrix} 1 \\ Ft \end{bmatrix} \otimes [Rt - \alpha - \beta Ft] \right] = 0.
\] (3.1)

The first set of moment conditions identifies \( \alpha \) and \( \beta \) as regression coefficients, while the last set of conditions represents the pricing restrictions. Note that there are \( N(1 + K + 1) \) moment conditions although there are \( N(1 + K) + K \) parameters, which implies that the system is overidentified. Again following Cochrane (2001), we set a linear combination of the given moment conditions to zero, that is, we set \( AE[ht(\alpha, \beta, \lambda)] = 0 \) with

\[
A = \begin{bmatrix} I_{N(1+K)} & 0_{N(1+K) \times N} \\ 0_{K \times N(1+K)} & \Theta_{K \times N} \end{bmatrix}.
\]

Note that the matrix \( A \) specified above combines the last \( N \) moment conditions into \( K \) moment conditions so that the system becomes exactly identified. Following Cochrane (2001), we take \( \Theta = \beta^T \Sigma_{\varepsilon}^{-1} \). The advantage of this particular choice is that the resulting \( \lambda \) estimates coincide with the GLS cross-sectional estimates.

3.2. Moment Conditions - Traded Factor Case

Asset pricing theory provides an additional restriction on the prices of risk when factors are traded, meaning that they are excess returns themselves. If a factor is an excess return, we have \( \lambda = E[F_t] \). For example, the price of market risk is equal to the expected excess market return, and the prices of size and book-to-market risks, as captured by Fama-French’s SMB and HML portfolio movements, are equal to the
expected SMB and HML excess returns. Note that we use the term “excess return” for
difference of gross returns, that is, not only in excess of the risk-free rate. Prices of
excess returns are zero, i.e., excess returns are zero investment portfolios.

The standard two–pass estimation procedure commonly found in the finance liter-
are may not give reliable estimates of risk prices when factors are traded. Hou and
Kimmel (2010) provide an interesting example to point out this issue. They generate
standard two–pass expected (excess) return estimates (both OLS and GLS) in the three
factor Fama–French model by using 25 size and book–to–market portfolios as test assets.
As shown in their Table 1, both OLS and GLS risk price estimates of the market are sig-
nificantly different from the sample average of the excess market return. It is important
to point out that the two–pass procedure ignores the fact that the Fama–French factors
are traded factors and it treats them in the same way as non–traded factors.

Consequently, when factors are traded we usually replace the second set of moment
conditions with the condition that their expectation of the vector of factors equals λ.
Then, the relevant moment conditions are given by

\[
E[h_t(\alpha, \beta, \lambda)] = E \left[ \begin{bmatrix} 1 \\ F_t \end{bmatrix} \otimes \begin{bmatrix} R_t^e - \alpha - \beta F_t \\ F_t^e - \lambda \end{bmatrix} \right] = 0, \tag{3.2}
\]

where \(F_t^e\) is the \(K \times 1\) vector of factor (excess) returns.

In this case, estimates are obtained by an exactly identified system, i.e., the number
of parameters equals the number of moment conditions. Note that if the factor is traded,
but we do not add the moment condition that the factor averages equal \(\lambda\), then the
results are just those of the non-traded case in Section 3.1.

Note that, alternatively, we could incorporate the theoretical restriction on factor
prices into the estimation by adding the factor portfolios as test assets in the linear pricing
equation, \(R_t^e - \beta \lambda\). This set of moment conditions would be similar to the general case,
with the only difference being that the linear pricing restriction incorporates the factors
as test assets in addition to the original set of test assets, i.e., we define $R^F_t = \begin{bmatrix} R^c_t \\ F_t \end{bmatrix}$.

Under this setting, the moment conditions would be given by

$$E[h_t(\alpha, \beta, \lambda)] = E \begin{bmatrix} 1 \\ F_t \end{bmatrix} \otimes [R^c_t - \alpha - \beta F_t] R^F_t - \beta F,R^c_t \lambda = 0,$$

(3.3)

with $\beta_{F,R} = \begin{bmatrix} \beta \\ I_K \end{bmatrix}$. Following the same procedure as in Section 3.1, we specify the $A$ matrix and set $\Theta = \beta_{F,R}^T \Sigma_{R^F}^{-1} R^F$. Because we find that GMM based on (3.3) leads to the same asymptotic variance covariance matrices for risk premiums as GMM based on (3.2), we omit the conditions (3.3) in the rest of the paper.

### 3.3. Moment Conditions - Factor–Mimicking Portfolios

Following Balduzzi and Robotti (2008), we also consider the case where risk prices are equal to expected returns of factor–mimicking portfolios. Then, the moment conditions used are

$$E[h_t(\alpha^m, \beta^m, \Phi_0, \Phi, \lambda^m)] = E \begin{bmatrix} 1 \\ F^m_t \end{bmatrix} \otimes [R^c_t - \alpha^m - \beta^m F^m_t] = 0,$$

(3.4)

with $F^m_t = \Phi R^c_t$. In this case, there are $K(1+N) + N(1+K) + K$ moment conditions and parameters, which makes the system exactly identified.

### 4. Precision of Risk–Premium Estimators

As mentioned in the introduction, our focus is on estimating risk premiums of individual assets or portfolios. Much of the literature on multi–factor asset pricing models has
primarily focused on the issue of a factor being priced or not. Formally, this is a test on (a component of) $\lambda$ being zero or not and, accordingly, the properties of risk–price estimates for $\lambda$ have been studied and compared. Examples include Shanken (1992), Jagannathan (1998), Kleibergen (2009), Lewellen, Nagel and Shanken (2010), Kan and Robotti (2011), and Kan et al. (2013).

In the current paper, since our focus is on analyzing the possible efficiency gains based on linear factor models in estimating expected (excess) returns, we first derive the joint distribution of estimates for $\beta$ and $\lambda$ for the three cases discussed in Sections 3.1 to 3.3. Then, we derive the asymptotic distributions of the implied expected (excess) return estimators given by the product $\hat{\beta}\hat{\lambda}$. Moreover, we illustrate the empirical relevance of our asymptotic results using the Fama–French three factor model with 25 Fama–French size and book–to–market portfolios as test assets. In particular, we provide the (asymptotic) variances of the various risk–premium estimators with empirically reasonable parameter values and evaluate the benefits of using linear factor models in estimating risk premiums, see Table 1 below.

The asset data used in this paper consists of 25 portfolios formed by Fama-French (1992,1993), downloaded from Kenneth French’s website. These portfolios are value–weighted and formed from the intersections of five size and five book–to–market (B/M) portfolios and they include the stocks of the New York Stock Exchange, the American Stock Exchange, and NASDAQ. For details, we refer the reader to Fama and French articles (1992,1993). The factors are the 3 factors of Fama–French (1992) (market, book–to–market, and size). Our analysis is based on monthly data from January 1963 until October 2012, i.e., we have 597 observations for each portfolio.

The following theorem provides the limiting distribution of the historical averages estimator. It’s classical and provided for comparison with the three GMM–based estimators in Sections 3.1–3.3.

**Theorem 4.1.** Given that $R_1, R_2, ..., R_T$ is a sequence of independent and identically distributed excess return vectors, we have $\sqrt{T} \left( \hat{R}_t - E[R_t] \right) \overset{d}{\to} \mathcal{N}(0, \Sigma_{R_t})$. 

12
Note that Theorem 4.1 assumes no factor structure. We will, next, provide the asymptotic distributions of expected (excess) return estimators given the linear factor structure implied by the Asset Pricing models. Note that the joint distributions of \( \lambda \) and \( \beta \) are different for each set of moment conditions, which leads to different asymptotic distributions for the risk premiums \( \beta \lambda \) as well. Hence, we derive the asymptotic distributions of expected (excess) return estimators for the three set of moment conditions introduced in Sections 3.1, 3.2 and 3.3 separately.

4.1. Precision with General Moment Conditions

The following theorem provides the asymptotic variances of the risk–premium estimators based on the general moment conditions in Section 3.1. Note again that this result is valid for both traded and non-traded factors.

**Theorem 4.2.** Impose Assumptions 1 and 2 and consider the moment conditions (3.1), i.e.,

\[
E[h_t(\alpha, \beta, \lambda)] = E \begin{bmatrix} 1 \\ F_t \end{bmatrix} \otimes [R_t - (\alpha - \beta F_t)] = 0.
\]

Then, the limiting variance of the expected (excess) return estimator \( \hat{\beta} \hat{\lambda} \) is given by

\[
\Sigma_{R^eR^e} = (1 - \lambda' \Sigma^{-1}_F \lambda) \left( \Sigma_{ee} - \beta' \Sigma^{-1}_e \beta \right)^{-1} \beta' \beta.
\] (4.1)

The proof is provided in the appendix. Theorem 4.2 provides the asymptotic covariance matrix of the factor–model based risk–premium estimators with the general moment conditions as in Section 3.1. This formula is useful mainly for two reasons. First, it can be used to compute the standard errors of these risk–premium estimates and, accordingly, the related t–statistics and p-values can be obtained. Second, it allows us to study the precision gains for estimating the risk premiums from incorporating information about the factor model.
In case of a one–factor model and a single test asset, the (asymptotic) variances of both the naive risk–premium estimator and the factor–model based risk–premium estimator are the same. When more assets/portfolios are available, \( N > 1 \), observe that the magnitude of the asymptotic variances of the risk–premium estimators depends on the prices of risk \( \lambda \), the exposures \( \beta \), and the idiosyncratic variances \( \Sigma_{\varepsilon \varepsilon} \). Note that the difference between the asymptotic covariance matrix of the naive estimator, \( \bar{\Sigma}_{\text{Re}} \) and the factor–based risk–premium estimator is 
\[
(1 - \lambda' \Sigma^{-1}_{FF} \lambda) \left( \Sigma_{\varepsilon \varepsilon} - \beta (\beta' \Sigma^{-1}_{\varepsilon \varepsilon} \beta)^{-1} \beta' \right).
\]
The following corollary formalizes the relation between the asymptotic covariance matrices of the naive estimator \( \bar{\Sigma}_{\text{Re}} \) and the general factor–model based risk–premium estimator of Section 3.1.

**Corollary 4.1.** Impose Assumptions 1 and 2 and consider the moment conditions (3.1).

If \( \lambda' \Sigma^{-1}_{FF} \lambda < 1 \), then the limiting variance of the expected (excess) return estimator \( \hat{\beta} \hat{\lambda} \) is at most \( \Sigma_{\text{Re}, \text{Re}} \).

Corollary 4.1 shows that there may be precision gains for estimating risk premiums from the added information about the factor model if \( \lambda' \Sigma^{-1}_{FF} \lambda \) is smaller than one. Note that although \( \lambda' \Sigma^{-1}_{FF} \lambda \) can be larger than one mathematically, it is typically smaller than one given the parameters found in empirical research. Observe that in the one–factor case with a traded factor, \( \lambda' \Sigma^{-1}_{FF} \lambda \) is the squared Sharpe ratio of that factor. This squared Sharpe ratio is, for stocks and stock portfolios, generally much smaller than 1. Moreover, plugging in the estimates from the Fama–French three factor model (based on GMM with moment conditions (3.1)) gives \( \lambda' \Sigma^{-1}_{FF} \lambda = 0.058 \). Note that the smaller the value for \( \lambda' \Sigma^{-1}_{FF} \lambda \), the larger the efficiency gains from imposing a factor model.

We calculate the (asymptotic) variances of the factor–model based risk–premium estimates for all 25 FF portfolios by plugging the parameter estimates into (4.1). Table 1 presents the results. Comparing the asymptotic variances of the factor–model based risk–premium estimators to those of the naive estimators, we see that the factor–model based risk–premium estimators are more precise than the naive estimators for all 25 Fama–
French portfolios. In particular, using the 3-factor model in estimating risk premiums of 25 FF portfolios leads to striking gains in variances of up to 25%. Note that a 25% gain in variances means that the same (statistical) precision can be obtained with 25% less observations.

4.2. Precision with Moment Conditions for Traded Factors

When the risk factors are traded, meaning that the factor itself is an excess return, additional restrictions on the prices of risk can be incorporated into the estimation. With the availability of such information, one could again expect efficiency gains in estimating both the prices of risk and the expected (excess) returns. In this section, we consider this case and the following theorem gives the asymptotic variances of the expected (excess) return estimators with the moment conditions (3.2) for the case all factors are traded.

**Theorem 4.3.** Suppose that all factors are excess returns. Under Assumptions 1 and 2, consider the moment conditions (3.2)

\[ E[h_t(\alpha, \beta, \lambda)] = E \left[ \begin{array}{c} 1 \\ F_t^- \end{array} \right] \otimes [R_t^e - \alpha - \beta F_t^- - \lambda] = 0. \]

Then, the limiting variance of the expected (excess) return estimator \( \hat{\beta} \hat{\lambda} \) is given by

\[ \Sigma_{R^e R^e} - (1 - \lambda' \Sigma_F^{-1} \lambda) \Sigma_{\varepsilon\varepsilon}. \]  

(4.2)

The theorem above shows that when the factors are traded, the asymptotic covariance matrices of the factor–based risk–premium estimators changes. This is because we incorporate, the additional information that prices of risk associated with the factors equal their expected returns.

Theorem 4.3 allows us to study the efficiency gains for estimating risk premiums from a model where the factors are traded compared to historical averages. Comparing the
asymptotic covariance matrix of the factor–based risk–premium estimators from GM-M (3.2) to the one of the naive estimator, we observe that the difference is given by 

\[(1 - \lambda'\Sigma_{FF}^{-1}\lambda) \Sigma_{\varepsilon\varepsilon}.\] 

Again, the factor \(1 - \lambda'\Sigma_{FF}^{-1}\lambda\) is empirically generally found to be positive implying an efficiency gain. Moreover, observe that asymptotic covariance matrix of the risk–premium estimator based on GMM with (3.2) can be different from the ones of the risk–premium estimator based on GMM with (3.1). This indicates that there may be efficiency gains even within the GMM framework from that the information that the factors are traded. The following corollary formalizes these issues.

**Corollary 4.2.** Suppose that all factors are traded. Under Assumption 1 and 2, consider the GMM estimator based on the moment conditions (3.2). Then, we have the following.

1. If \(\lambda'\Sigma_{FF}^{-1}\lambda < 1\), then the limiting variance of the expected (excess) return estimator \(\hat{\beta}_{\lambda}\) is at most \(\Sigma_{R^eR^e}\).

2. The limiting variance of this expected (excess) return estimator is at most the limiting variance of the estimator based on the moment conditions (3.1).

Plugging in the parameter estimates from the analysis of the Fama–French model now gives \(\lambda'\Sigma_{FF}^{-1}\lambda < 1 = 0.052^3\). Comparing the variances of the risk–premium estimates based on GMM with (3.2) to those of the naive estimators (in Table 1, we see that the risk–premium estimates based on GMM with (3.2) have smaller asymptotic variances than the naive estimators. In particular, the magnitude of efficiency gains goes up to 31%. Moreover, consistent with Theorem 4.2, the asymptotic variances of the risk–premium estimates based on GMM with (3.1) exceed those of the risk premium estimators based on GMM with (3.2). Specifically, the risk–premium estimates based on GMM with (3.1) have up to 7.6% larger variances than the risk–premium estimates based on GMM with (3.2). Overall, these sizeable precision gains from estimating risk premiums based on factor models stem from two sources. First, the linear relation implied by asset pricing models

---

\[^3\text{Note that } \lambda'\Sigma_{FF}^{-1}\lambda < 1 \text{ is equal to } 0.058 \text{ in the general case based on GMM 3.1. This happens because estimation based on GMM with the set of moment conditions (3.1) leads to } \lambda \text{ estimates which are different than } \lambda \text{ estimates obtained with GMM with (3.2).} \]
is valuable information in the estimation of risk premiums. Second, when the factors are traded, the additional information that the prices of risk equal the expected factor valued increases the precision of risk–premium estimates.

4.3. Precision with Moment Conditions Using Factor–Mimicking Portfolios

One may hope that replacing factors by factor–mimicking portfolios may also bring efficiency gains compared to (4.1) since the additional restriction on the price of the factor risk can be incorporated into the estimation. In this subsection, we derive the asymptotic variances of expected (excess) return estimators obtained with factor–mimicking portfolios.

**Theorem 4.4.** Under Assumption 1 and 2, consider the GMM estimator based on the moment conditions (3.4)

\[
E[h_t(\alpha^m, \beta^m, \Phi_0, \Phi, \lambda^m)] = E \left[ \begin{bmatrix} 1 \\ F_t^m \end{bmatrix} \otimes \left[ F_t - \Phi_0 - \Phi R_t^c \right] \right] = 0.
\]

Then, the limiting variance of the expected (excess) return estimator, \( \hat{\beta}^m \hat{\lambda}^m \), is given by

\[
\Sigma_{R^c R^c} = \left( \mu_{R^c} \right)^\prime \left\{ \Sigma_{R^c R^c}^{-1} - \Sigma_{R^c R}^{-1} \beta \left( \beta^\prime \Sigma_{R^c R}^{-1} \beta \right)^{-1} \beta^\prime \Sigma_{R^c R}^{-1} \right\} \mu_{R^c} \\
\times \left( \Sigma_{R^c R^c}^{-1} - \beta \left( \beta^\prime \Sigma_{R^c R}^{-1} \beta \right)^{-1} \Sigma_{F^c F^c} \left( \beta \Sigma_{R^c R}^{-1} \beta \right)^{-1} \beta^\prime \right) \\
- \left( 1 - \mu_{R^c} \Sigma_{R^c R^c}^{-1} \mu_{R^c} \right) \left( \Sigma_{R^c R^c}^{-1} - \beta \left( \beta^\prime \Sigma_{R^c R}^{-1} \beta \right)^{-1} \beta^\prime \right),
\]

with \( \mu_{R^c} = E[R_t^c] \).

Theorem 4.4 enables us to study the efficiency gains in risk premiums using factor–mimicking portfolios. Observe that the difference between the asymptotic covariance matrices of the naive estimator and the factor–model based GMM risk–premium estima-
tor with (3.4) is given by

\[
\mu'_{R^e} \left\{ \Sigma^{-1}_{R^e R^e} - \Sigma^{-1}_{R^e R_m} \beta \left( \beta' \Sigma^{-1}_{R^e R^e} \beta \right)^{-1} \beta' \Sigma^{-1}_{R^e R^e} \right\} \mu_{R^e} \\
\times \left( \Sigma_{R^e R^e} - \beta \left( \beta' \Sigma^{-1}_{R^e R^e} \beta \right)^{-1} \Sigma^{-1}_{FF} \left( \beta' \Sigma^{-1}_{R^e R^e} \beta \right)^{-1} \beta' \right) \\
+ \left(1 - \mu'_{R^e} \Sigma^{-1}_{R^e R^e} \mu_{R^e}\right) \left( \Sigma_{R^e R^e} - \beta \left( \beta' \Sigma^{-1}_{R^e R^e} \beta \right)^{-1} \beta' \right).
\]  

(4.4)

Efficiency gains with respect to the historical averages estimator are dependent on Eqn. (4.4) being positive semi–definite or not. Although, we were not able to prove this formally, the results from our empirical analysis with FF-3 factor model illustrates that there are considerable efficiency gains over the naive estimation for all 25 Fama–French 25 portfolios (see Table 1). In particular, estimating risk premiums with GMM (3.4) leads to, of up to 31%, smaller variances than estimating them with the naive estimator. Moreover, we find that estimating risk premiums by making use of the mimicking portfolios leads to small efficiency losses over the estimation based on the general case, i.e, GMM (3.1) for all assets, ranging between 0.1% and 1.5%.

Note that one important difference between Theorem 4.4 and Theorem 4.2 may potentially come from the estimation of the mimicking portfolio weights. The estimation of the weights of the factor–mimicking portfolio potentially leads to different (intuitively higher) asymptotic variances for the betas of the mimicking factors as well as for the mimicking factor prices of risk, and the risk premiums, which are essentially a multiplication of \( \beta^m \) and \( \lambda^m \). Such issue is similar to errors–in–variables type of corrections in two step Fama–Macbeth estimation, i.e., the Shanken (1992) correction in asymptotic variances for generated regressors. We should recall here that GMM standard errors automatically account for such effects as the system of moment conditions is solved simultaneously.

In particular, in our setting with moments conditions (3.4), GMM treat the moments producing \( \Phi \) simultaneously with the moments generating \( \beta^m \) and \( \lambda^m \). Hence, the long run covariance matrix captures the effects of estimation of \( \Phi \) on the standard errors of the \( \beta_m \) and \( \lambda_m \), hence the risk premiums.

If we consider the Fama–French three factor model with the 25 FF–portfolios, we
can also intuitively gain insights about the difference between the inferences about risk premiums based on GMM with the two sets of moment conditions (3.2) and (3.4). In fact, since the factors are traded, meaning that they are excess returns themselves, we can estimate the risk premiums via the second set of moment conditions (3.2). Moreover, we can also estimate such system via the third set of moment conditions (3.4), which has the additional burden of estimating the coefficients for the construction of the mimicking portfolio. Accordingly, GMM estimation via the second set and the third set of moment conditions may lead to different precisions for the risk premium estimates. The last column in Table 1 documents the efficiency comparisons in estimating risk premiums of 25 FF portfolios employing factor mimicking portfolios over risk premium estimation with moment conditions (3.2). Efficiency losses are present for all 25 Fama–French portfolios, meaning that risk premium estimates employing factor mimicking portfolios, i.e., based on (3.4), are less precise than risk premium estimates based on (3.2). Small efficiency losses range between 1% and 8.4% across portfolios.

5. Inference about Risk Premiums when the $\beta$’s are small

A number of papers in the literature documents inference issues regarding the prices of risk when the factors are weakly correlated with the asset returns (see, e.g., Kleibergen (2009), Burnside (2015), Bryzgalova (2014), and Kleibergen and Zhan (2015)). When $\beta$’s are close to zero and/or when the $\beta$ matrix is almost of reduced rank, the confidence intervals of the prices of risk estimates, $\hat{\lambda}$, are erroneous, which leads to unreliable statistical inference. The effects may be severe in empirical research, as the confidence intervals of the risk price estimates may even be unbounded as documented, for the case of the conditional consumption CAPM of Lettau and Ludvigson (2001), in Kleibergen (2009). Accordingly, Kleibergen (2009) provides identification-robust statistics and confidence sets for the risk price estimates when the $\beta$’s are small.

\footnote{Kleibergen (2009) documents that \%95 percent confidence bounds of the prices of risk on the scaled consumption growth coincides with the whole real line.}
However, once our interest is in estimating risk premiums on individual assets or portfolios, a natural question is whether similar issues exist in the presence of weakly correlated factors. The reader should remember that our focus of interest is on the product of $\beta$ and $\lambda$, rather than on $\lambda$ only. In this section, we provide some results to shed light on this issue.

In the rest of this section, we will focus on the specification where $\beta$ has small but non-zero values. Following the literature on weak instruments (see, e.g., Staiger and Stock (1997), and Kleibergen (2009), we consider a sequence of $\beta$’s getting smaller as the sample size increases.

**Corollary 5.1.** Suppose Assumption 1 (a), (b) and Assumption 2 hold and consider the small $\beta$ case where $\beta = \frac{1}{\sqrt{T}}B$ for a fixed full rank $N \times K$ matrix $B$. Then

1. Consider the GMM estimator based on the moment conditions (3.1). The limiting variance of the risk price estimator, $(1 + \lambda'\Sigma^{-1}_{FF,\lambda})(\beta'\Sigma_{\varepsilon\varepsilon}\beta^{-1})^{-1} + \Sigma_{FF}$, is unbounded.

2. The limiting variance of the expected (excess) return estimator based on the moment conditions (3.1), i.e., the variance in (4.1), is bounded and depends on the space spanned by the columns of $B$ only.

3. Suppose that all factors are traded and consider the GMM estimator based on the moment conditions (3.2). The limiting variance of the risk price estimator, $\Sigma_{FF}$, is not affected by the value of $\beta$.

4. The limiting variance of the expected (excess) return estimator based on the moment conditions (3.2), i.e., the variance in (4.2), is not affected by the value of $\beta$, and, thus, bounded.

5. Consider the GMM estimator based on the moment conditions (3.4). The limiting variance of the expected (excess) return estimator, i.e., the variance in (4.3), depends on the space spanned by the columns of $B$ only, and, thus, is bounded.

Corollary 5.1 documents several important findings of our analysis regarding the issue of small but non-zero factor loadings. First, if the parameters of the linear factor
model are estimated with GMM based on the moment conditions (3.1), then the limiting variances of the risk premium estimators do not suffer from either lack of identification or unboundedness when $\beta = B/\sqrt{T}$, see Corollary 5.1-2. In particular, the $\frac{1}{\sqrt{T}}$ term cancels out in the limiting variance (4.1), and, hence, the limiting variance only depends on the space spanned by the columns of $B$. However, the first point in Corollary 5.1 documents that this is not the case if one is interested in the prices of risk, $\lambda$. Specifically, it highlights that the limiting variances of the risk price estimators blow up when $\beta = B/\sqrt{T}$. This result is in line with the literature documenting unreliable statistical inference about the prices of risk based on the Fama-Macbeth and GLS two-pass estimation and their unbounded confidence sets.

Corollary 5.1-3 and Corollary 5.1-4 shed light on the issue of small $\beta$’s when all factors in the linear factor model of interest are traded. In this case, if one estimates the parameters of the model with GMM based on (3.2), then the limiting variances of the risk premium estimators are not affected by the $\beta$ having a weak value or not. This is a straightforward result in the sense that the asymptotic variance (4.2) is independent of the value of $\beta$. Lastly, Corollary 5.1-5 considers the small $\beta$ issue for the estimation based on factor–mimicking portfolios. The finding is consistent with the previous two cases and the limiting variances of the risk premium estimators does not suffer from either lack of identification or the unboundedness when the $\beta$ has a weak value, i.e., $\beta = B/\sqrt{T}$.

The bottomline of this section is the following: if one is interested in making statistical inference about the prices of risk $\lambda$, small but non-zero $\beta$’s may have detrimental effects on this inference in line with Kleibergen (2009). However, once the interest is in estimating risk premiums, i.e., expected (excess) returns on individual assets or portfolios, the estimators based on GMM with (3.1), (3.2) or (3.4) do not suffer from lack of identification or unbounded limiting variances when the $\beta$’s are weak.
6. Risk Premium Estimation with Omitted Factors

The asymptotic results in the previous section are based on the assumption that the pricing model is correctly specified. The researcher is assumed to know the true factor model that explains expected excess returns on the assets. In that case, the risk-premium estimators are consistent certainly under our maintained assumption of independently and identically distributed returns. However, the pricing model may be misspecified and this might induce inconsistent risk-premium estimates. We investigate this issue and its solution in the present section.

We consider model misspecification due to omitted factors. An example of such type of misspecification would be to use Fama–French three factor model if the true pricing model is the four factor Fama–French–Carhart Model. Formally, assume that excess returns are generated by a factor model with two different sets of distinct factors, $F$ and $G$ such that

$$R_e = \alpha^* + \beta^* F + \delta^* G + \varepsilon^*, \quad (6.1)$$

where $\varepsilon^*$ is a vector of residuals with mean zero and $E[F\varepsilon^*'] = 0$ and $E[G\varepsilon^*'] = 0$. Note that the sets of factors $F$ and $G$ perfectly explain the expected excess returns of the test assets, i.e., $E[R_e] = \beta^* \lambda_F + \delta^* \lambda_G$.

However, a researcher may be ignorant about the presence of the factors $G$ and thus estimates the model only with factors $F$. Then, the estimated model is

$$R_e = \alpha + \beta F + \varepsilon, \quad (6.2)$$

with zero-mean $\varepsilon$, and $E[F\varepsilon'] = 0$. As the researcher might not know the underlying factor model exactly, she allows for misspecification by adding an $N$-vector of constant terms, $\alpha$, in the estimation as in Fama and French (1993).

The bias in the parameter estimates for, $\alpha$, $\beta$ and $\lambda$ are presented in the following theorem:

**Theorem 6.1.** Assume that returns are generated by (6.1) but $\alpha$, $\beta$, and $\lambda$ are estimated
from (6.2) with GMM (3.1). Then,

1. \( \hat{\alpha} \) converges to \( \alpha^* + (\beta^* - \beta)E[F] + \delta^* E[G] \),
2. \( \hat{\beta} \) converges to \( \beta^* + \delta^* \text{Cov}[G, F^T] \Sigma_{FF}^{-1} \),
3. \( \hat{\lambda} \) converges to \( \lambda_F + (\beta' \Sigma_{xx}^{-1} \beta)^{-1} \beta' \Sigma_{xx}^{-1} \left[ (\beta^* - \beta) \lambda_F + \delta^* \lambda_G \right] \),

in probability.

Lemma 6.1 shows that, if a researcher ignores some risk factors \( G \), then the risk price estimators associated with the factors \( F \) are inconsistent if and only if

\[ \beta' \Sigma_{xx}^{-1} \left[ (\beta^* - \beta) \lambda_F + \delta^* \lambda_G \right] \neq 0. \]

It is important to note that the inconsistency of the estimates of risk prices may be caused not only by the risk prices \( \lambda \) of the omitted factors but also the bias in betas of the factors \( F \). This result has an important implication: even if the ignored factors have zero price of risk, the cross-sectional estimates of the prices of risk on the true factors included in the estimation (\( F \)) can still be asymptotically biased. This happens in case \( F \) and \( G \) are correlated.

Next, we analyse the asymptotic bias in the parameter estimates for \( \alpha \), \( \beta \) and \( \lambda \) in case the factors are traded and the estimation is based on GMM with the moment conditions (3.2) of Section 3.2.

**Theorem 6.2.** Assume that returns are generated by (6.1) but \( \alpha \), \( \beta \), and \( \lambda \) are estimated from (6.2) with GMM (3.2). Then,

1. \( \hat{\alpha} \) converges to \( \alpha^* + (\beta^* - \beta)\lambda_F + \delta^* \lambda_G \),
2. \( \hat{\beta} \) converges to \( \beta^* + \delta^* \text{Cov}[G, F^T] \Sigma_{FF}^{-1} \),
3. \( \hat{\lambda} \) converges to \( \lambda_F \),

in probability.

Theorem 6.2 illustrates that, even if the researcher forgets some risk factors, risk price estimators will still be asymptotically unbiased. Notice that this is in contrast with the
estimator based on GMM with moment conditions (3.1) of Section 3.1. It is important to note that, if the forgotten factors, \( G \), are uncorrelated with the factors, then the bias in \( \beta \) disappears. Moreover, if the ignored factors are associated with zero prices of risk and are uncorrelated with \( F \), then \( \hat{\alpha} \) will converge to zero.

What happens to the risk-premium estimators on individual assets or portfolios if some true factors are ignored? The following corollary provides consistency conditions for risk-premium estimators of individual assets or portfolios.

**Corollary 6.1.** If the returns are generated by (6.1) and

- the model (6.2) is estimated with GMM (3.1), then the vector of resulting risk-premium estimators \( \hat{\beta} \hat{\lambda} \) converges to
  
  \[
  E[R^c] \text{ if and only if } [I_N - \beta(\beta'\Sigma_{cc}^{-1}\beta)^{-1}\beta'\Sigma_{cc}^{-1}]E[R^c] = 0.
  \]

- all factors are traded. If the model (6.2) is estimated with GMM (3.2), then the vector of resulting risk-premium estimators \( \hat{\beta} \hat{\lambda} \) converges to \( E[R^c] \) if and only if
  
  \[
  (\beta^* - \beta)\lambda_F + \delta^*\lambda_G = 0.
  \]

In the view of the theorem above, if the model (6.2) is estimated with GMM (3.1), the consistency of the risk-premium estimators is dependent on a specific condition that may not be satisfied. Moreover, if the factors are traded and the estimation is via GMM with moment conditions (3.2), then the risk-premium estimator obtained may be biased.

In order to capture misspecification, it is a common approach to add an \( N \)-vector of constant terms, \( \alpha \), to the model as in (6.2). In the following theorem, we will show that in case of traded factors, it is possible to achieve the consistency for estimating risk premiums, however, this comes at the cost of loosing all efficiency gains.

**Theorem 6.3.** Assume that all factors in \( F \) are traded. If the returns are generated by (6.1) but the model (6.2) is estimated with GMM (3.2) where the risk price estimates are given by the factor averages, then the estimator \( \hat{\alpha} + \hat{\beta} \hat{\lambda} \) is consistent for \( E[R^c] \). However, the asymptotic variance of such estimator equals \( \Sigma_{R^c R^c} \).
It is important to note that adding the $\hat{\alpha}$ to $\hat{\beta}\hat{\lambda}$ does not solve the inconsistency problem if the system is estimated via GMM with (3.1). If some factors are non-traded and the parameters are estimated via GMM with (3.1), adding the $\hat{\alpha}$ capturing the misspecification to $\hat{\beta}\hat{\lambda}$ doesn’t lead to consistent estimates of $E[R^e]$. In particular, $\hat{\alpha} + \hat{\beta}\hat{\lambda}$ converges to $E[R^e] - \beta(\lambda - E[F])$ and $\lambda - E[F]$ is not necessarily zero.

7. Application: Portfolio Choice with Parameter Uncertainty

In the previous sections, we provided an asymptotic analysis of the three factor–model based risk premium estimators and analyzed the efficiency gains with respect to naive historical averages. In this section, we analyze the economic significance of these gains in a portfolio allocation problems à la Markowitz (1952).

The implementation of the mean–variance framework of Markowitz (1952) requires the estimation of first two moments of the asset returns. Although in the setting of Markowitz (1952), optimal portfolios are supposed to achieve the best performance, in practice, the estimation error in expected returns via the historical averages leads to large deterioration of the out–of–sample performance of the optimal portfolios (see, e.g., DeMiguel et al. (2009b)). At the extreme, this has led to simply abandoning the application of theoretically optimal decisions and using the naive techniques such as the $1/N$ portfolio or the global minimum variance portfolios as these are not subject to estimation risk of expected returns. In this section, we analyze the out–of–sample performances of the optimal portfolios based on factor–based risk–premium estimates as well as the historical averages, $1/N$ portfolio and global minimum variance portfolio in a simulation analysis.

**Optimization Problem:** Suppose a risk–free asset exists and $w$ is the vector of relative portfolio allocations of wealth to $N$ risky assets. The investor has preferences that are fully characterized by the expected return and variance of his selected portfolio. The investor maximizes her expected utility, by choosing the vector of portfolio weights $w$ such that
\[ E[U] = w' \mu^e - \frac{\gamma}{2} w' \Sigma_{RR} w, \]  
(7.3)
is maximize, where \( \gamma \) measures the investor’s risk aversion level and \( \mu^e \) and \( \Sigma_{RR} \) denote the expected excess returns on the assets and covariance matrix of returns. The solution to the maximization problem above is given by
\[ w_{opt} = \frac{1}{\gamma} \Sigma_{RR} \mu^e. \]  
(7.4)

In the optimization problem above, since the true risk premium vector, \( \mu^e \), and the true covariance matrix of asset returns, \( \Sigma_{RR} \), are unknown, in empirical work, one needs to estimate them. Following the classical “plug in” approach, the moments of the excess return distribution, \( \mu^e \) and \( \Sigma_{RR} \), are replaced by their estimates.

**Portfolios Considered:** We consider four portfolios constructed with different risk–premium estimators: the optimal portfolio constructed with historical averages, the optimal portfolios constructed with the three factor model–based GMM risk premium estimates with moment conditions (3.1), (3.2) and (3.4). Note that the covariance matrix is estimated using the traditional sample counterpart, \( \frac{1}{T-1} \sum_{t=1}^{T} (R_t - \bar{R}_t)(R_t - \bar{R}_t)' \), where \( \bar{R}_t \) is the sample average of returns. We also consider the global minimum variance (GMV thereafter) portfolio\(^5\) to which we compare the performance of the portfolios based on the risk–premium estimates. Note that the implementation of this portfolio only requires estimation of the covariance matrix, for which we again use the sample counterpart, and completely ignores the estimation of expected returns. Moreover, we analyze the performance of the \( 1/N \) portfolio.

**Performance Evaluation Criterion and Methodology:** We compare performances of the portfolios by using their out-of-sample Sharpe Ratios\(^6\). We provide results
\(^5\)This portfolio is obtained by minimizing the portfolio variance with respect to the weights with the only constraint that weights sum to 1 and the \( N \)-vector of portfolio weights is given by \( w_{gmv} = \frac{\Sigma_{RR}^{-1}}{\Sigma_{RR}^{-1} \Sigma_{RR}^{-1}} \)
\(^6\)See Peñaranda and Sentana (2011) for an analysis examining the improvements in the estimation of in–sample mean variance frontiers based on asset pricing model restrictions, tangency or spanning constraints.
both for “enlarging windows” and “rolling windows”.

- **Enlarging Windows**: We set an initial window length over which we estimate the mean vector of excess returns and covariance matrix, and obtain the various portfolio weights. For our analysis, the initial window length is of 120 data points, corresponding to 10 years of data. We then calculate the one-period ahead returns, \( \hat{w}_t R_{t+1} \), of the estimated portfolios. Next, we re-estimate the portfolio weights by including the next period’s return and use this to calculate the portfolio return for the subsequent period. We continue doing this and obtain the time series of out-of-sample excess returns for each portfolio considered, from which we calculate the out-of-sample Sharpe ratios.

- **Rolling Windows**: We start with an initial window length of 120 observations over which we estimate the mean vector of excess returns, and obtain the various portfolio weights. We then calculate the one-period ahead returns, \( \hat{w}_t R_{t+1} \), of the estimated portfolios. Next, we re-estimate the portfolio weights by including the next period’s return and dropping the first period’s return, and use this to calculate the subsequent period’s portfolio return. We obtain a time series out-of-sample excess returns for all the portfolios considered, and obtain the out-of-sample Sharpe ratios.

**Simulation Setting**: We use the following return-generating process:

\[
R_t^e = \alpha + \beta F_t + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{7.5}
\]

with \( F_t \) and \( \varepsilon_t \) drawn from multivariate normal distributions under the null of \( \text{E}[R_t^e] = \beta \lambda \). To make our simulations realistic, we calibrate the parameters for the return-generating process by using the monthly data from January 1963 until October 2012 on twenty-five Fama-French (1992) portfolios sorted by size and book-to-market as risky assets and the nominal 1-month Treasury bill rate as a proxy for risk-free rate, and the 3 Fama-French (1992) portfolios (market, book-to-market and size factors) as
the risk factors. Specifically, we estimate $\alpha, \beta, \mu, \Sigma_{FF}, \Sigma_{ee}, \lambda$ and take them to be the truth in the simulation exercise to generate samples of 597 observations. We simulate independent sets of $Z = 10000$ return samples and for each set of simulated sample, we calculate the out–of–sample Sharpe ratios for the various portfolios.

Table 2 provides the simulation results for the out–of–sample Sharpe ratios of different portfolios. In particular, we provide results on the optimal portfolios based on different risk–premium estimates, GMV and $1/N$ portfolios. Moreover, we provide the true Sharpe ratio of the optimal portfolio, which we refer to as theoretical. For each portfolio, we present the average estimate over simulations, $\overline{SR}$ (first line), the bias as the percentage of the population Sharpe ratios, $(\overline{SR} - SR)/SR$ (second line) and the root–mean–square error (RMSE) in parentheses, the square root of $\sum_{s=1}^{Z}(\hat{SR}_s - SR)/Z$, (third line), where $Z = 10000$.

In order to isolate the effect of the error in risk–premium estimates, we present our results with true and estimated $\Sigma_{RR}$. Firstly, note that the true Sharpe ratio of the optimal portfolio is superior to the portfolios based on estimated risk–premiums or co–variance matrix of asset returns. Comparing the average Sharpe ratio of the optimal portfolio based on historical averages to the true Sharpe ratio of optimal portfolio for enlarging samples (rolling samples), we see that the bias is striking and negative with $-41.6\% \ (-58.5\%)$ and $-44.4\% \ (-62.3\%)$, depending on whether the covariance matrix of asset returns is the true one or the estimated one. However, using factor–models to estimate risk–premiums reduces the bias in Sharpe ratios substantially to a level ranging from $-12.5\% \ (-28.2\%)$ to $-11.2\% \ (-26.1\%)$ and ranging from $-9.1\% \ (-22.2\%)$ to $-7.7\% \ (-19.7\%)$ depending on true or estimated covariance matrices. In particular, with GMM–Gen estimates, average Sharpe ratio of the optimal portfolio is $0.188 \ (0.154)$ in case of true covariance matrix (with an improvement of 50% over the average Sharpe ratios with the historical averages) and $0.195 \ (0.167)$ in case of an estimated covariance matrix (with an improvement of 64% over the average Sharpe ratios with the historical averages). Among the optimal portfolios constructed with factor–model based risk–premium
estimates, the one based on GMM–Tr estimates perform the best with 0.198 (0.172). However, the differences in biases are minimal for all optimal portfolios constructed with factor–model based risk–premium estimators.

Next, we analyse the RMSEs of the various portfolios. Out–of–sample Sharpe ratio of the optimal portfolios based on historical averages is extremely volatile across simulations. That is, for the case of enlarging samples, it has a RMSE of 0.108 (given the average estimate 0.119) if the covariance matrix is estimated. The situation gets worse if the optimization is based on rolling samples, with a RMSE of 0.142 (given the average estimate 0.081). However, using factor-model based risk–premium estimators decreases the RMSEs substantially. Among the optimal portfolios based on factor–model based risk–premium estimators, GMM–Tr performs the best with a RMSE of 0.052 (given the average estimate of 0.198), as expected from the asymptotic analyses of risk–premium estimators in previous sections. However, the differences in RMSEs are minor among the portfolios with factor–based risk–premium estimates.

Comparing the average Sharpe ratios of the optimal portfolios the factor model–based risk premium estimates with GMV and 1/N, we see that optimal portfolios based on the naive estimator performs worse than 1/N strategy and slightly better than the GMV portfolio when the optimization is based on the enlarging samples, and performs considerably worse than both the GMV portfolios and 1/N strategy in case of rolling samples. Moreover, both GMV and 1/N have substantially lower RMSEs. This result is consistent with the findings in the literature that GMV portfolio as well as 1/N strategy has better out–of–sample performance than the optimal portfolios based on sample moments (See, e.g., Jagannathan and Ma (2003), De Miguel et al (2009) and Jorion (1985, 1986, 1991)). However, the average Sharpe ratios for all optimal portfolios based on factor model–based risk premium estimates are considerably larger than both the GMV and 1/N portfolios, with an improvement ranging from 13% to 46% for the case of enlarging samples. Moreover, their out of Sharpe ratios across simulations are almost as stable as the GMV portfolio as well as the 1/N strategy.
Overall, using the factor–model based risk–premium estimators improves the performance of optimal portfolios substantially over the optimal portfolios based on the plug in estimates of historical averages in terms of both bias and RMSEs. Moreover, in contrast to the optimal portfolios with historical averages, these portfolios perform considerably better than the global minimum variance portfolio.

8. Conclusions

It has been the standard technique in the literature to use average historical returns as estimates of expected excess returns, that is risk premiums, on individual assets or portfolios. These estimators are very noisy. This translates into the need for very large, in practice, mostly infeasible, samples of data in order to gain some precision. However, the finance literature provides a wide variety of risk–return models which imply a linear relationship between the expected excess returns and their exposures.

In this paper, we show that, when correctly specified, such parametric specifications on the functional form of risk premiums lead to significant inference gains for estimating expected (excess) returns. In the standard Fama–French three factor model (MKT, SMB, HML) setting with 25 FF portfolios, the efficiency gains are sizable and go up to 31% for individual portfolios. For real life applications, this translates into the benefit of using only 69% of the data with factor–model based risk–premium estimates to obtain the same precision as with the historical averages estimator. Moreover, we show that the presence of weakly identified factors, the confidence bounds of factor model based risk premium estimators are not affected, whereas the confidence bounds of the risk price estimators may be unbounded. We also show that using a misspecified asset pricing model in the sense that some factors are forgotten generally leads to inconsistent estimates. However, in case the factors are traded, then adding an alpha to the model capturing mispricing leads to consistent estimators. Out of sample performance of optimal portfolios significantly improves if factor–model based estimates of risk premium are used in portfolio weights instead of the classical historical averages.
A. Proofs

In the rest of the paper, the covariance matrix of the factor–mimicking portfolios is denoted by $\Sigma_{FmFm}$. 

A.1. Equivalence of factor pricing using mimicking portfolios

Proof of Theorem 2.1. Define $M^m$ as the projection of $M$ onto the augmented span of excess returns,

$$M^m = \mathcal{P}(M|1, R^e) \quad (A.1)$$

so that

$$\mathbb{E}[M] = \mathbb{E}[M^m], \quad (A.2)$$

$$\text{Cov}[M, R^e] = \text{Cov}[M^m, R^e]. \quad (A.3)$$

Thus, we have

$$\beta \lambda = \text{Cov}[R^e, F^\prime] \Sigma_{FF}^{-1} \left( -\frac{1}{\mathbb{E}[M]} \Sigma_{FF} b \right) \quad (A.4)$$

$$= -\frac{1}{\mathbb{E}[M]} \text{Cov}[R^e, F^\prime] b$$

$$= -\frac{1}{\mathbb{E}[M^m]} \text{Cov}[R^e, F^{m\prime}] b$$

$$= -\frac{1}{\mathbb{E}[M^m]} \text{Cov}[R^e, F^{m\prime}] \Sigma_{FmFm}^{-1} \Sigma_{FmFm} b$$

which completes the proof. \qed

A.2. Precision of Parameter Estimators Given a Factor Model

This section provides the proofs for asymptotic properties of the parameter estimators under the specified linear factor model. The lemma A.1 below illustrates the asymptotic distribution of the GMM estimators with a given set of moment conditions provided that a pre–specified matrix $A$, that essentially determines the weights of the overidentifying moments, is introduced. Thereafter, these results will be used to calculate the variance covariance matrix for the moment conditions (3.1), (3.2) and (3.4), respectively.

Under appropriate regularity conditions, see, e.g., Hall (2005), Chapter 3.4, we have the following result.
Lemma A.1. Let $\theta \in \mathbb{R}^p$ be a vector of parameters and the moment conditions are given by $E[h_t(\theta)] = 0$ where $h_t(\theta) \in \mathbb{R}^q$, independently and identically distributed over time. Given a prespecified matrix $A \in \mathbb{R}^{p \times q}$, its consistent estimator $\hat{A}$ and $\hat{A}' \sum_{t=1}^{T} h_t(\hat{\theta}) = 0$, 

$$\sqrt{T}(\hat{\theta} - \theta) \overset{d}{\rightarrow} \mathcal{N}(0, [AJ]^{-1}ASA'[J'A']^{-1}),$$

(A.5)

where,

$$J = E\left[\frac{\partial h_t(\theta)}{\partial \theta'}\right],$$

(A.6)

$$S = E[h_t(\theta)h_t(\theta)'].$$

(A.7)

The above lemma presents the asymptotic distribution of the parameters in a general GMM context. In the subsequent lemmas, limiting distributions for the expected (excess) return estimators based on the moment conditions (3.1), (3.2) and (3.4), respectively.

Lemma A.2. Under Assumptions 1, 2 and the moment conditions (3.1) with parameter vector $\theta = (\alpha', \text{vec}(\beta)', \lambda)'$, we have

$$\sqrt{T}(\hat{\theta} - \theta) \overset{d}{\rightarrow} \mathcal{N}(0, V),$$

(A.8)

with

$$V = \begin{bmatrix}
1 + \mu_F' \Sigma_{FF}^{-1} \mu_F & -\mu_F' \Sigma_{FF}^{-1} \\
-\Sigma_{FF}^{-1} \mu_F & \Sigma_{FF}^{-1}
\end{bmatrix} \otimes \Sigma_{\epsilon \epsilon} V_c$$

$$V_c' = (1 + \lambda' \Sigma_{FF}^{-1} \lambda)(\beta' \Sigma_{\epsilon \epsilon}^{-1} \beta)^{-1} + \Sigma_{FF}$$

where $\mu_F = E[F_t]$ and $V_c = \begin{bmatrix} 1 + \mu_F' \Sigma_{FF}^{-1} \lambda & \beta' \Sigma_{\epsilon \epsilon}^{-1} \beta \end{bmatrix} \otimes (\beta' \Sigma_{\epsilon \epsilon}^{-1} \beta)^{-1}$.

Proof. The proof follows from plugging the appropriate matrices for the moment condi-
tions provided in Section 3.1 into the variance covariance formula in (A.5) and performing the matrix multiplications. Below, we provide the limiting variance covariance matrix \( S \) and the Jacobian \( J \) for this specific set of moment conditions,

\[
S = \begin{bmatrix}
    \Sigma_{ee} & \mu'_F \otimes \Sigma_{ee} & \Sigma_{ee} \\
    \mu_F \otimes \Sigma_{ee} & [\Sigma_{FF} + \mu_F \mu'_F] \otimes \Sigma_{ee} & \mu_F \otimes \Sigma_{ee} \\
    \Sigma_{ee} & \mu'_F \otimes \Sigma_{ee} & \beta \Sigma_{FF} \beta' + \Sigma_{ee}
\end{bmatrix}.
\]

\[
J(\theta) = E \left[ \frac{\partial h_t(\theta)}{\partial \theta'} \right] = \begin{bmatrix}
    -1 & \mu'_F \\
    \mu_F & \Sigma_{FF} + \mu_F \mu'_F \\
    0_{N \times N} & -\lambda' \otimes I_N
\end{bmatrix} \otimes I_N 0_{N(K+1) \times K}.
\]

Furthermore

\[
A = \begin{bmatrix}
    I_{N(K+1)} & 0_{N(K+1) \times N} \\
    0_{K \times N(K+1)} & \beta' \Sigma_{ee}^{-1}
\end{bmatrix}.
\]

so that the limiting variance of GMM estimator for \( \theta \) is obtained by performing the matrix multiplications \([AJ]^{-1}ASA'[JA']^{-1}\).

\[\square\]

**Lemma A.3.** Suppose that all factors are traded. Then, under Assumptions 1, 2 and the moment conditions (3.2) with parameter vector \( \theta = (\alpha', \text{vec}(\beta)'', \lambda)' \), we have

\[
\sqrt{T}(\hat{\theta} - \theta) \overset{d}{\rightarrow} N(0, V), \quad \text{(A.9)}
\]

with

\[
V = \begin{bmatrix}
    1 + \mu'_F \Sigma_{FF}^{-1} \mu_F & -\mu'_F \Sigma_{FF}^{-1} \\
    -\Sigma_{FF}^{-1} \mu_F & \Sigma_{FF}^{-1} \\
    0_{K \times N(K+1)} & \Sigma_{FF}
\end{bmatrix} \otimes \Sigma_{ee} 0_{N(K+1) \times K}.
\]

**Proof.** The proof follows from plugging the appropriate matrices for the moment con-
ditions (3.2) into the variance covariance formula in (A.5) and performing the matrix multiplications. Below, we provide the limiting variance covariance matrix \( S \), Jacobian \( J \) for this specific set of moment conditions. In this case,

\[
S = \begin{bmatrix}
\Sigma_{ee} & \mu_F \otimes \Sigma_{ee} & 0_{N \times K} \\
\mu_F \otimes \Sigma_{ee} & [\Sigma_{FF} + \mu_F \mu_F'] \otimes \Sigma_{ee} & 0_{NK \times K} \\
0_{K \times N} & 0_{K \times NK} & \Sigma_{FF}
\end{bmatrix},
\]

and

\[
J(\theta) = \begin{bmatrix}
1 & \mu_F \\
\mu_F & \Sigma_{FF} + \mu_F \mu_F' \\
0_{K \times N(K+1)} & I_K
\end{bmatrix} \otimes I_{N \times (K+1)}.
\]

Thus, the limiting variance of the GMM estimator for \( \theta \) is obtained by performing the matrix multiplications \( J^{-1}S[J']^{-1} \) since \( A = I_{N(K+1)+K} \).

The next lemma provides the asymptotic properties of the GMM estimator with factor–mimicking portfolios.

**Lemma A.4.** Given that Assumption 1, 2 are satisfied and that (2.8)–(2.10) hold, then under the moment conditions (3.4), for \( \theta = (\text{vec}(\beta^m)', \lambda^m)' \), we have

\[
\sqrt{T}(\hat{\theta} - \theta) \overset{d}{\rightarrow} \mathcal{N}(0, V),
\]

(A.10)

with

\[
V = \begin{bmatrix}
\Sigma_{FF}^{-1} \Phi \Sigma_{RR}^{-1} \Phi' \Sigma_{FF}^{-1} \otimes \beta^m \Sigma_{uu} \beta^m' + \Sigma_{FF}^{-1} \otimes \Sigma_{ee} \\
-\mu_F \Sigma_{FF}^{-1} \otimes \Sigma_{uu} \\
-\mu_F \Sigma_{FF}^{-1} \otimes \Sigma_{uu} \\
0_{K \times N(K+1)} & \Sigma_{FF}^{-1} \mu_F \otimes \beta^m \Sigma_{uu}
\end{bmatrix}.
\]

**Proof.** The proof follows again from plugging the appropriate matrices for the moment conditions (3.4) into the variance covariance formula in (A.5) and performing the matrix multiplications.
multiplications. Now, observe that from (A.7), we have
\[
S = \begin{bmatrix}
1 & \mu_R' \\
\mu_R' & \Sigma_{RR} + \mu_R' \mu_R
\end{bmatrix} \otimes \Sigma_{uu} = \begin{bmatrix}
0_{K(1+N) \times N(K+1)} & 0_{K(1+N) \times K}
\end{bmatrix},
\]
and from (A.6), we have
\[
J(\theta) = E \begin{bmatrix}
1 & \mu_F' \\
\mu_F' & \Sigma_{FF} + \mu_F' \mu_F
\end{bmatrix} \otimes \Sigma_{e\times e} = \begin{bmatrix}
0_{N(K+1) \times K} & 0_{N(K+1) \times K}
\end{bmatrix},
\]
with \( A = I_{K(1+N)+N(K+1)+K} \). Thus, the limiting variance of the GMM estimator for \( \theta = (\text{vec}(\beta^m)', \lambda^m)' \) is obtained by performing the matrix multiplications \( J^{-1} S J^{-1} \).

Here, it is worth stressing that the limiting variance covariance matrix obtained by performing the matrix multiplications corresponds to the parameter vector
\[
(\Phi_0', \text{vec}(\Phi)', \alpha^m, \text{vec}(\beta^m)', \lambda^m)'.
\]

Therefore, the asymptotic variance covariance matrix for \( \theta = (\text{vec}(\beta^m)', \lambda^m)' \) is the lower-right \( KN + K \) by \( KN + K \) sub-matrix of the larger variance covariance matrix.

Lemmas A.2–A.4 allow us to study the asymptotic properties of the obtained risk premium estimators. It is worth mentioning that the lower–left \( NK + K \) dimensional square matrices of the variance covariance matrices in Lemma A.2 and A.3 give the variance covariance matrices corresponding to parameters \( (\text{vec}(\beta)', \lambda') \). We will use
these results to derive the variance covariance matrices of risk premium estimators in the following section.

Proof of Theorem 4.1. This follows from a direct application of the Central Limit Theorem.

Proofs of Theorems 4.2 and 4.3. We are interested in the asymptotic distribution of \( g(\beta, \lambda) = \beta \lambda \). Given

\[
(\text{vec} (\hat{\beta})', \hat{\lambda}') - (\text{vec} (\beta)'', \lambda'') \xrightarrow{d} \mathcal{N}(0, V_{\beta, \lambda}),
\]

we have, by applying the delta method, that

\[
\sqrt{T} \left( g(\hat{\beta}, \hat{\lambda}) - g(\beta, \lambda) \right) \xrightarrow{d} \mathcal{N}(0, \dot{g}' V_{\beta, \lambda} \dot{g}),
\]

with

\[
\dot{g} = \left[ \lambda' \otimes I_N - \beta \right].
\]

Remember that Lemma A.2 and A.3 give the asymptotic distributions of \( \sqrt{T}(\hat{\theta} - \theta) \) where \( \theta = (\alpha', \text{vec} (\beta)', \lambda')' \) for the moment conditions (3.1) and (3.2). Observe that \( V_{\beta, \lambda} \) is the lower \( NK + K \) block diagonal matrix of the variance covariance matrices provided in Lemma A.2 and A.3. Hence, the asymptotic variances of the risk premium estimators in Theorems 4.2 and 4.3 follow from plugging in the limiting variance covariance matrices of (vec (\beta)', \lambda')' and calculating \( \dot{g}' V_{\beta, \lambda} \dot{g} \).

Proof of Theorem 4.4. We are interested in \( g(\beta^m, \lambda^m) = \beta^m \lambda^m \). Given

\[
(\text{vec} (\hat{\beta}^m)', \hat{\lambda}^m)' - (\text{vec} (\beta^m)', \lambda^m)' \xrightarrow{d} \mathcal{N}(0, V_{\beta^m, \lambda^m}),
\]

Then, by applying the delta method, we have

\[
\sqrt{T}(g(\hat{\beta}^m, \hat{\lambda}^m) - g(\beta^m, \lambda^m)) \xrightarrow{d} \mathcal{N}(0, \dot{g}' V_{\beta^m, \lambda^m} \dot{g})
\]

and note that here

\[
\dot{g} = \left[ \lambda^m' \otimes I_N - \beta^m \right].
\]

Then, we have
\[ y' \Sigma_{F_F} y = \lambda (y' \Sigma_{F_F}^{-1} y + \beta' \Sigma_{F_F} \beta) \]

The result follows from plugging the \( \beta \) and \( \Phi \) respectively into the above equation via (2.14) and (2.13).

**Lemma A.5.** Let

\[
K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}
\]

be a symmetric matrix and assume that \( K_{22}^{-1} \) exists. Then \( K \geq 0 \) is equivalent to \( K_{22} \geq 0 \) and \( K_{11} - K_{12} K_{22}^{-1} K_{21} \geq 0 \).

**Proof of Corollary 4.1.** Suppose \( \lambda' \Sigma_{F_F}^{-1} \lambda < 1 \). We need to study the difference between the limiting variance of the historical averages and the limiting variance of the expected (excess) return estimator based on (3.1). In particular, we need to study

\[
\Sigma_{R^*R^*} - (\Sigma_{R^*R^*} - (1 - \lambda' \Sigma_{F_F}^{-1} \lambda) [\Sigma_{ee} - \beta (\beta' \Sigma_{ee}^{-1} \beta)^{-1} \beta']) = (1 - \lambda' \Sigma_{F_F}^{-1} \lambda) [\Sigma_{ee} - \beta (\beta' \Sigma_{ee}^{-1} \beta)^{-1} \beta'].
\]

In order to show that \( \Sigma_{ee} - \beta (\beta' \Sigma_{ee}^{-1} \beta)^{-1} \beta' \) is positive semi-definite, we will use Lemma A.5. Now, let \( K_1 = \Sigma_{ee}^{1/2} \) and \( K_2 = \beta' \Sigma_{ee}^{-1/2} \). Then,

\[
K = \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} \begin{bmatrix} K_1' & K_2' \\ \end{bmatrix} = \begin{bmatrix} K_1 K_1' & K_1 K_2' \\ K_2 K_1' & K_2 K_2' \end{bmatrix}
\]

so that

\[
K = \begin{bmatrix} \Sigma_{ee} & \beta \\ \beta' & \beta' \Sigma_{ee}^{-1} \beta \end{bmatrix}.
\]
Then, Lemma A.5 yields that
\[
\Sigma_{ee} - \beta (\beta' \Sigma_{ee}^{-1} \beta)^{-1} \beta' \geq 0
\]

**Proof of Corollary 4.2.** Suppose \(\lambda' \Sigma_{FF}^{-1} \lambda < 1\).

In order to prove Corollary 4.2–1, we need to study the difference between the limiting variance of the historical averages and the limiting variance of the expected (excess) return estimator based on (3.2). In particular, we need to show that
\[
\Sigma_{RR} - (\Sigma_{RR} - (1 - \lambda' \Sigma_{FF}^{-1} \lambda) \Sigma_{ee})
= (1 - \lambda' \Sigma_{FF}^{-1} \lambda) \Sigma_{ee}
\]
is positive semi–definite. Since \(\Sigma_{ee}\) is positive semi–definite, Corollary 4.2–1 follows.

In order to prove Corollary 4.2–2, we need to study the difference between the limiting variance of the expected (excess) return estimator based on (3.1) and the limiting variance of the expected (excess) return estimator based on (3.2). In particular, we need to show that
\[
(\Sigma_{RR} - (1 - \lambda' \Sigma_{FF}^{-1} \lambda) \Sigma_{ee} - \beta (\beta' \Sigma_{ee}^{-1} \beta)^{-1} \beta')
= (1 - \lambda' \Sigma_{FF}^{-1} \lambda) (\beta' \Sigma_{ee}^{-1} \beta)^{-1} \beta'
\]
is positive semi–definite. This follows immediately from \(\Sigma_{ee}\) being positive semi–definite.

**Proof of Theorem 6.1.** Note that \(\hat{\beta}\) converges to \(\beta\) and \(\hat{\alpha}\) converges to \(\alpha\) in probability, where
\[
\beta = \text{Cov} \left[R^*, F^T\right] \Sigma_{FF}^{-1}, \\
= \text{Cov} \left[R^* = \alpha^* + \beta^* F + \delta^* G + \epsilon^*, F^T\right] \Sigma_{FF}^{-1}, \\
= \beta^* + \delta^* \text{Cov} \left[G, F^T\right] \Sigma_{FF}^{-1}.
\]

and

\[
\alpha = \mathbb{E}[R^*] - \beta \mathbb{E}[F], \\
= \alpha^* + \beta^* \mathbb{E}[F] + \delta^* \mathbb{E}[G] - \beta \mathbb{E}[F], \\
= \alpha^* + (\beta^* - \beta) \mathbb{E}[F] + \delta^* \mathbb{E}[G].
\]

Furthermore, for \(\hat{\lambda}\), first notice that

\[
\hat{\lambda} = \left(\beta^\prime \Sigma_{H}^{-1} \beta\right)^{-1} \beta^\prime \Sigma_{H}^{-1} \hat{R}^e.
\]

The probability limit of \(\hat{\lambda}\) from GMM (3.1) is thus given by

\[
\lambda = \left(\beta^\prime \Sigma_{H}^{-1} \beta\right)^{-1} \beta^\prime \Sigma_{H}^{-1} \left[\beta^* \lambda_F + \delta^* \lambda_G\right] \\
= \lambda_F + \left(\beta^\prime \Sigma_{H}^{-1} \beta\right)^{-1} \beta^\prime \Sigma_{H}^{-1} \left[(\beta^* - \beta) \lambda_F + \delta^* \lambda_G\right].
\]

\[\Box\]

Proof of Theorem 6.2. Note that \(\hat{\beta}\) converges to \(\beta\) and \(\hat{\alpha}\) converges to \(\alpha\) in probability, with

\[
\beta = \text{Cov} \left[R^*, F^T\right] \Sigma_{FF}^{-1}, \\
= \text{Cov} \left[R^* = \alpha^* + \beta^* F + \delta^* G + \epsilon^*, F^T\right] \Sigma_{FF}^{-1}, \\
= \beta^* + \delta^* \text{Cov} \left[G, F^T\right] \Sigma_{FF}^{-1}.
\]
\[
\begin{align*}
\alpha &= E[R^e] - \beta E[F] \\
&= \alpha^* + \beta^* \lambda_F + \delta^* \lambda_G - \beta \lambda_F \\
&= \alpha^* + (\beta^* - \beta) \lambda_F + \delta^* \lambda_G.
\end{align*}
\]

Furthermore, for \( \hat{\lambda}_F \), notice that \( \hat{\lambda}_F = \bar{F} \), which converges to \( \lambda_F = E[F] \) in probability.

\[\Box\]

Proof of Corollary 6.1. For the first part of the corollary, note that

\[\beta \lambda_F = \beta (\beta' \Sigma_{ee}^{-1} \beta)^{-1} \beta' \Sigma_{ee}^{-1} E[R^e].\]  

(A.18)

Hence, \( \hat{\beta} \hat{\lambda} \) is consistent for \( E[R^e] \) if and only if \( E[R^e] = \beta (\beta' \Sigma_{ee}^{-1} \beta)^{-1} \beta' \Sigma_{ee}^{-1} E[R^e]. \)

This, in turn, equivalent to

\[\left[ I_N - \beta (\beta' \Sigma_{ee}^{-1} \beta)^{-1} \beta' \Sigma_{ee}^{-1} \right] E[R^e] = 0.\]  

(A.19)

To prove the second part of the corollary, note that \( \hat{\beta} \hat{\lambda} \) converges to \( \beta \lambda \). Using (A.16) and \( \lambda_F = E[F] \), we have

\[\beta \lambda_F = (\beta^* + \delta^* \text{Cov} \left[ G, F^T \right] \Sigma_{FF}^{-1}) \lambda_F,\]  

(A.20)

\[= E[R^e] - ((\beta^* - \beta) \lambda_F + \delta^* \lambda_G).\]

\[\Box\]

Proof of Theorem 6.3. Consistency of \( \hat{\alpha} + \beta \lambda_F \) is straightforward. The asymptotic variance is given by the delta method using \( g(\alpha, \beta, \lambda_F) = \alpha + \beta \lambda_F \). The asymptotic covariance matrix of \( \alpha, \beta, \) and \( \gamma \) is given in Lemma A.3 (denoted by \( V \)).

Thus,

\[\sqrt{T} \left( g(\hat{\alpha}, \hat{\beta}, \hat{\lambda}) - g(\alpha, \beta, \lambda) \right) \xrightarrow{d} N(0, \tilde{V}_{\alpha,\beta,\lambda}).\]  

(A.21)
with

\[ \hat{g} = \begin{bmatrix} 1 & \lambda' \otimes I_N & \beta \end{bmatrix}. \]

Matrix multiplication of calculating \( \hat{g}V_{\alpha,\beta,\lambda}\hat{g} \) gives \( \Sigma_{R^eR^e} \).


Table 1: Improvements in Efficiency for the 25 Fama–French Portfolios (in percentage)

This table illustrates the gains in variances (in percentage) for the various risk–premium estimates for the 25 portfolios formed by Fama and French (1992, 1993). The factors are the three factors from Fama and French (1992): market, size and book-to-market. The results are based on monthly data from January 1963 until October 2012, i.e., 597 observations for each portfolio. The first column ($RP_{GMM} – Gen \over Naive$) presents the improvements for the factor–model based risk–premium estimates based on GMM with (3.1) over the naive estimate of historical averages. The second ($RP_{GMM} – Tr \over Naive$) and the third ($RP_{GMM} – Mim \over Naive$) columns present the gains of factor–model based risk–premium estimates based on GMM with (3.2) or (3.4) over naive estimates, respectively. The fourth column ($RP_{GMM} – Tr \over RP_{GMM} – Gen$) corresponds to the precision gains from estimating the risk premiums based on GMM using the moment conditions (3.2) over the case based on GMM with (3.1). The last column ($RP_{GMM} – Mim \over RP_{GMM} – Gen$) presents the gains from making use of mimicking portfolios using (3.4) over estimation based on (3.1).

<table>
<thead>
<tr>
<th>Assets</th>
<th>$RP_{GMM} – Gen \over Naive$</th>
<th>$RP_{GMM} – Tr \over Naive$</th>
<th>$RP_{GMM} – Mim \over Naive$</th>
<th>$RP_{GMM} – Tr \over RP_{GMM} – Gen$</th>
<th>$RP_{GMM} – Mim \over RP_{GMM} – Gen$</th>
<th>$RP_{GMM} – Mim \over RP_{GMM} – Tr$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>28</td>
<td>17</td>
<td>7.6</td>
<td>-1.5</td>
<td>-8.4</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>25</td>
<td>17</td>
<td>5.6</td>
<td>-1.1</td>
<td>-6.3</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>16</td>
<td>11</td>
<td>4.0</td>
<td>-0.8</td>
<td>-4.5</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>13</td>
<td>10</td>
<td>3.0</td>
<td>-0.6</td>
<td>-3.5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>9</td>
<td>5</td>
<td>2.9</td>
<td>-0.6</td>
<td>-3.4</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>15</td>
<td>6</td>
<td>6.8</td>
<td>-1.4</td>
<td>-7.7</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>11</td>
<td>6</td>
<td>4.3</td>
<td>-0.9</td>
<td>-4.9</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>3.0</td>
<td>-0.6</td>
<td>-3.5</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>1.5</td>
<td>-0.2</td>
<td>-1.8</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>1.3</td>
<td>-0.2</td>
<td>-1.5</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>13</td>
<td>7</td>
<td>5.2</td>
<td>-1.1</td>
<td>-5.9</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>2.3</td>
<td>-0.4</td>
<td>-2.6</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>1.1</td>
<td>-0.1</td>
<td>-1.2</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>1.0</td>
<td>-0.1</td>
<td>-1.0</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>1.0</td>
<td>-0.1</td>
<td>-1.1</td>
</tr>
<tr>
<td>16</td>
<td>10</td>
<td>13</td>
<td>10</td>
<td>2.2</td>
<td>-0.4</td>
<td>-2.6</td>
</tr>
<tr>
<td>17</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>0.9</td>
<td>-0.1</td>
<td>-1.0</td>
</tr>
<tr>
<td>18</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>1.0</td>
<td>-0.1</td>
<td>-1.1</td>
</tr>
<tr>
<td>19</td>
<td>11</td>
<td>12</td>
<td>10</td>
<td>1.6</td>
<td>-0.2</td>
<td>-1.8</td>
</tr>
<tr>
<td>20</td>
<td>14</td>
<td>15</td>
<td>14</td>
<td>1.5</td>
<td>-0.2</td>
<td>-1.6</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>0.9</td>
<td>-0.1</td>
<td>-1.0</td>
</tr>
<tr>
<td>22</td>
<td>11</td>
<td>13</td>
<td>11</td>
<td>1.6</td>
<td>-0.2</td>
<td>-1.7</td>
</tr>
<tr>
<td>23</td>
<td>18</td>
<td>22</td>
<td>18</td>
<td>2.9</td>
<td>-0.5</td>
<td>-3.2</td>
</tr>
<tr>
<td>24</td>
<td>15</td>
<td>21</td>
<td>14</td>
<td>5.3</td>
<td>-1.0</td>
<td>-6.0</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>31</td>
<td>24</td>
<td>4.7</td>
<td>-0.8</td>
<td>-5.3</td>
</tr>
</tbody>
</table>
Table 2: Out-of-Sample Sharpe Ratios based on various risk-premium estimates

This table provides the average out-of-sample Sharpe ratio (first line), its percentage error compared to the true Sharpe ratio (second line) and the root-mean-squared errors (third line) over 10000 simulated data sets for optimal portfolios constructed with various risk premium estimates. The data generating process is

\[ R_t = \alpha + \beta F_t + \varepsilon_t, \quad t = 1, 2, \ldots, T, \]

with normally distributed \( F_t \) and \( \varepsilon_t \) and the asset pricing restriction \( E[R_t] = \beta \lambda \). \( R^e \) is the N-vector of asset returns at period \( t \), \( F_t \) is K-vector of factors and \( T \) is the number of periods. The moments of factors and residuals and the parameters of data generating process are obtained from a calibration of Fama-French 3 factor model from January 1963 to October 2012. The risk premium estimates are based on naive, GMM with moment conditions (3.1)-\( RP_{GMM-Gen} \), (3.2)-\( RP_{GMM-Tr} \) and (3.4)-\( RP_{GMM-Mim} \). The variance-covariance matrix is estimated by the sample variance covariance matrix. \( T \) is assumed to be 597 and risk aversion is 5. The upper panel presents the results for enlarging samples and the lower panel presents the results for the rolling windows based on window sizes of 120 observations.

<table>
<thead>
<tr>
<th>( \gamma = 5 )</th>
<th>True ( \mu )</th>
<th>Naive</th>
<th>RP(_{GMM-Gen})</th>
<th>RP(_{GMM-Tr})</th>
<th>RP(_{GMM-Mim})</th>
<th>GMV</th>
<th>1/N</th>
<th>Theoretical</th>
</tr>
</thead>
<tbody>
<tr>
<td>True ( \Sigma_{RR} )</td>
<td>0.218</td>
<td>0.125</td>
<td>0.188</td>
<td>0.190</td>
<td>0.188</td>
<td>0.111</td>
<td>0.136</td>
<td>0.214</td>
</tr>
<tr>
<td>0.019</td>
<td>-0.416</td>
<td>-0.125</td>
<td>-0.112</td>
<td>-0.125</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.004</td>
<td>0.103</td>
<td>0.056</td>
<td>0.055</td>
<td>0.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.208</td>
<td>0.119</td>
<td>0.195</td>
<td>0.198</td>
<td>0.195</td>
<td>0.111</td>
<td>0.136</td>
<td>0.214</td>
<td></td>
</tr>
<tr>
<td>( \hat{\Sigma}_{RR} )</td>
<td>-0.030</td>
<td>-0.444</td>
<td>-0.091</td>
<td>-0.077</td>
<td>-0.091</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.006</td>
<td>0.108</td>
<td>0.053</td>
<td>0.052</td>
<td>0.053</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \text{Rolling-sample Window Length=120} )</th>
<th>True ( \Sigma_{RR} )</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.218</td>
<td>0.089</td>
<td>0.154</td>
<td>0.158</td>
<td>0.154</td>
<td>0.111</td>
<td>0.136</td>
<td>0.214</td>
<td></td>
</tr>
<tr>
<td>0.019</td>
<td>-0.585</td>
<td>-0.282</td>
<td>-0.261</td>
<td>-0.282</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.004</td>
<td>0.135</td>
<td>0.080</td>
<td>0.076</td>
<td>0.080</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.193</td>
<td>0.081</td>
<td>0.167</td>
<td>0.172</td>
<td>0.167</td>
<td>0.103</td>
<td>0.136</td>
<td>0.214</td>
<td></td>
</tr>
<tr>
<td>( \hat{\Sigma}_{RR} )</td>
<td>-0.101</td>
<td>-0.623</td>
<td>-0.222</td>
<td>-0.197</td>
<td>-0.222</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.022</td>
<td>0.142</td>
<td>0.071</td>
<td>0.067</td>
<td>0.071</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>