On Estimation of Risk Premia in Linear Factor Models

Kewei Hou
Fisher College of Business
Ohio State University*

Robert L. Kimmel
Fisher College of Business
Ohio State University†

This Version: September 14, 2006

Abstract

We examine theoretical and econometric issues in the estimation of risk pemia in a linear factor model, when the model is misspecified. We show that, for a given set of test assets, the risk premium of an unspanned factor is very sensitive to the choice of other factors in the model. However, the risk premium of the projection of the unspanned factor onto the asset space is robust to the choice of other factors. These results highlight the importance of using factor-mimicking portfolios, rather than unspanned factors, in estimation of linear factor models.

---

*Columbus, OH 43210. Phone: (614) 292-0552. E-mail: hou.28@osu.edu.
†Columbus, OH 43210. Phone: (614) 292-1875. E-mail: kimmel.42@osu.edu.

1 Introduction

Linear factor models of expected returns have played a central role in financial economics for decades. Such models are sometimes motivated by theoretical considerations, as with the CAPM of Sharpe (1964) and Lintner (1965), the ICAPM of Merton (1973), the APT of Ross (1976), and the CCAPM of Breeden (1979). All of these models require that the excess returns of financial assets obey a linear relationship with their exposures to various sources of economic risk. However, linear factor models can also be empirically motivated. For example, the three-factor linear model of Fama and French (1996) includes two factors that do not play any role in traditional economic theories, but whose importance in asset pricing has nonetheless been observed in stock price data. Empirical studies of expected excess returns, such as Black, Jensen, and Scholes (1972), Fama and MacBeth (1973), Fama and French (1992), Jagannathan and Wang (1996), Daniel and Titman (1997), and Daniel, Titman, and Wei (2001) generally focus on either or both of the following questions: (1) What are the factors that explain the cross-section of expected returns of financial assets? and (2) What are the risk premia associated with those factors?

We examine this second question, the assignment and estimation of risk premia in linear factor models for expected excess returns, in the presence of possible misspecification, where "misspecification" means that the factors do not fully explain the expected excess returns of all assets. Much of the literature on estimation of risk premia either assumes correct model specification, or focuses on very specific forms of misspecification. The former category includes Shanken (1992), who derives the distribution of risk premia estimates from a two-pass regression method under an assumption that the time series of returns and factor realizations are i.i.d., Kim (1995), who considers the effect of heteroskedasticity on risk premia estimates, Amihud, Christensen, and Mendelson (1992), who offer an alternative to traditional two-pass regression methodologies, and Kan and Zhou (1999), who examine the relative merits of estimation of risk premia (as well as other quantities) by regression methods or GMM methods. In the latter category, Roll and Ross (1994), inspired by the findings of Fama and French (1992), determine when a two-pass regression study is likely to estimate a risk premium of zero for a market factor. Kandel and Stambaugh (1995) consider a misspecified model with a single spanned factor, and find that two-pass OLS regression estimates of risk premia are highly sensitive to economically meaningless repackagings of the assets included in the study; they go on to show that GLS risk premium estimates are invariant to such repackaging. Kan and Zhang (1999b) and Kan and Zhang (1999a) focus on the effect of including "useless" factors, not correlated with any of the asset returns, in regression and GMM estimation procedures, and find that an estimate of the risk premium of the useless factor does not necessarily converge in probability to zero with increasingly long data samples, despite the fact that the factor has nothing to do with asset returns.

These studies focus mainly on econometric issues; by contrast, we find that misspecification makes it impossible even to define the risk premium of an unspanned factor unambiguously, let alone estimate it consistently from a panel of data. A spanned factor has a unique and well-defined risk premium, which can be estimated consistently by any one of several simple econometric techniques. By contrast, the risk premium of an unspanned factor depends on the other factors included in the model, and can be made to take on any arbitrarily specified value by appropriate selection of the other factors. Furthermore, if factors are unspanned because the set of assets included in a study does not span the full universe of assets available, then omission of even seemingly innocuous assets (in terms of the Sharpe ratio offered to investors) can have dramatic effects on the risk premia assigned to unspanned factors. Spanned factors can therefore be assigned a meaningful risk premium, independently of the other factors included in the model and the assets used in the study (provided a change in the set of assets does not cause the factor to cease to be spanned). By contrast, the risk premium assigned to an unspanned factor depends both on the other factors in the model and the assets used in a study, rather than being an intrinsic property of the factor itself. Furthermore, we find that the asymptotic variance of the risk premium estimate for a spanned factor is unaffected by misspecification, whereas the asymptotic variance for the risk premium estimate for an unspanned factor is increasing in the degree of misspecification.

To interpret these results, we note that the risk premium of a spanned factor is based directly on the expected excess returns of the assets which span it. The risk premium of an unspanned factor can be viewed as the sum of the risk premium of its spanned and unspanned components. The spanned component of an unspanned factor (i.e., its projection onto the asset space) is itself a spanned factor; like other spanned factors, its risk premium is the expected excess return of its spanning portfolio. However, there are by definition no assets whose expected excess returns reveal the risk premium of the unspanned component. Rather, this part is determined by extrapolation of the risk premium of the spanned component onto the unspanned component. This extrapolation is highly sensitive to changes in the factor set and in the asset space, and relies on an assumption of correct model specification. The validity of the risk premium of unspanned factors therefore relies on an assumption of model correctness, whereas that of spanned factors does not. This reliance on model correctness can have severe consequences. For example, in a given factor model, a particular factor may have a projection with a risk premium of zero, but a large (positive or negative) risk premium assigned to the unspanned component. If the model is in fact not correctly specified, this large risk premium may turn out to be entirely fictional. Even worse, a factor may have a projection with a positive risk premium, but an unspanned component with an even larger negative imputed risk premium. The net negative risk premium of the factor, based on an assumption of correct model specification, entirely obscures the positive risk premium of the projection, which is robust to misspecification. Empirical studies in general do not distinguish between these two components of factor risk premia, and often interpret the risk premium of an unspanned factor as if it were an intrinsic property of the factor itself, rather than a quantity that is dependent both on the other factors included in the model and on the assets included in the empirical study. We argue that the risk premia of projected factors, being robust to misspecification, may provide more useful information than the usually reported risk premia of the unspanned factors themselves.

The rest of this paper is organized as follows. In Section 2, we examine the assignment of a vector of risk premia to a set of factors, and show that the risk premium of an unspanned factor is highly sensitive to the choice of the other factors included in the model. In Section 3, we examine the effect of asset omission on risk premium assignment, and find that failure to include even seemingly unimportant assets in a study can have dramatic effects on the risk premia assigned to unspanned factors. In Section 4, we offer an interpretation of the results of the previous sections, showing that the assignment of risk premia to unspanned factors effectively assigns shadow expected returns to the unspanned components; the risk premium of a factor then contains a component that is directly observed in the data, and a component that is extrapolated out onto the unspanned components. This latter component, which relies on correct model specification, can dominate the first, which does not. Section 5 considers the problem of estimation of risk premia from data, and finds that misspecification increases the asymptotic variance of estimates of risk premia for unspanned factors, but not for spanned factors. This section also provides an estimator of the risk premia vector for factor projections, and derives its asymptotic variance, which is unaffected by misspecification. Finally, Section 6 concludes.

2 Linear Factor Models and Risk Premia

In this section, we examine the definition of risk premia for both spanned and unspanned factors. Whereas the risk premium of a spanned factor admits a simple, straightforward, and unique definition, the risk premium assigned to an unspanned factor in general depends on the other factors included in the model. Consequently, it is not possible to refer to the risk premium of an unspanned factor without reference to the model in which it occurs. The only exception occurs when the factor in question explains the expected returns of all assets perfectly; such an unspanned factor can be assigned a risk premium unambiguously.

Throughout, our interest is in the relation between a set of factor realizations and a set of asset payoffs, with each asset having an initial cost of zero. We usually assume that these factor realizations and asset payoffs have finite variances and that their variances and covariances satisfy certain rank conditions.

Definition 1. *An M-vector of asset payoffs Z and an N-vector of factor realizations F satisfy the full-rank assumptions if all of the following conditions are satisfied:*

1. *Each asset payoff and each factor realization has finite variance.*
2. *The $M \times M$ covariance matrix of the asset payoffs, Σ_{ZZ}, has full rank.*
3. *The $N \times N$ covariance matrix of the factor realizations, Σ_{FF}, has full rank.*
4. *The $N \times M$ covariance matrix between the factor realizations and asset payoffs, Σ_{FZ}, has rank N (which in turn requires $N \leq M$).*

We denote the expected values of the factor realizations and the asset payoffs by μ_F and μ_Z, respectively; the means are necessarily finite if the full rank assumptions are satisfied. The transpose of Σ_{FZ} is denoted by Σ_{ZF}. When a set of M asset payoffs Z and N factor realizations F satisfy the full-rank assumptions, there exist an $N \times M$ matrix of constants β and an M-vector of random variables ε such that the following conditions hold:

$$(Z - \mu_Z) = \beta^T (F - \mu_F) + \varepsilon \tag{2.1}$$

$$E[\varepsilon] = 0_{M\times 1} \tag{2.2}$$

$$Cov[F, \varepsilon] = 0_{N\times M} \tag{2.3}$$

Since the covariance matrix Σ_{FF} is of full rank, the matrix β is unique:

$$\beta = \Sigma_{FF}^{-1}\Sigma_{FZ} \tag{2.4}$$

Note that the rank assumptions on Σ_{FF} and Σ_{FZ} ensure that β is also of full rank.

Definition 2. *A set of factor realizations F is called a linear factor model with respect to a set of asset payoffs Z if F and Z satisfy the full-rank assumptions, and if:*

$$\mu_Z = \beta^T\gamma = \Sigma_{ZF}\Sigma_{FF}^{-1}\gamma \tag{2.5}$$

for some N-vector γ. The elements of γ are called the risk premia of the corresponding elements of F.

Equation 2.5 has the familiar interpretation that each asset has an expected excess return proportional to the exposure of that asset to various sources of economic risk. Linear factor models play a fundamental role in asset pricing theory; for example, the CAPM model of Sharpe (1964) and Lintner (1965) predicts that the market portfolio is a linear factor model for any set of assets; the ICAPM of Merton (1973) and the APT of Ross (1976) are linear factor models with potentially more than one factor. Mathematically, a linear factor model always exists, provided Z has a finite and full rank covariance matrix. For example, for a given Z, we can construct a model with only a single factor $F = \mu_Z^T\Sigma_{ZZ}^{-1}Z$ which tautologically constitutes a linear factor model for Z, with risk premium $\gamma = \mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z$. However, this purely mechanical construction offers no economic insight or intuition; much research in financial economics focuses instead on constructing linear factor models in which the factors have simple intuitive interpretations. For example, Fama and French (1996) construct a three factor model in which the factors relate to firm size, book-to-market ratios, and comovement with the market return.

We use the notation $\gamma(F, Z)$ to denote the risk premia vector when either the set of factors or the set of assets under consideration is not clear from the context. The vector of risk premia can be expressed in terms of the moments of the asset payoffs and factor realizations, as per the following Lemma.

Lemma 1. *If a set of factor realizations F is a linear factor model for a set of asset payoffs Z, then the risk premia vector γ is unique and is given by:*

$$\gamma = \gamma(F, Z) = \Sigma_{FF}\left(\Sigma_{FZ}\Sigma_{ZZ}^{-1}\Sigma_{ZF}\right)^{-1}\Sigma_{FZ}\Sigma_{ZZ}^{-1}\mu_Z \tag{2.6}$$

Proof: See Appendix.

It is a brief exercise to verify that if F is a linear factor model for Z, it is also a linear factor model for any other set of assets Y that spans a subspace of the space spanned by Z (provided Y and F satisfy the full-rank assumptions), with $\gamma(F, Y) = \gamma(F, Z)$. It is also straightforward to verify that if F is a linear factor model for Z and G is a set of factors that spans the same space as F, then G is also a linear factor model for Z.

The β matrix describes the projection of the assets Z onto the space of factors F; we shall often wish to refer to the reverse projection, of the factors F onto the space of assets Z. Under the full-rank assumptions, there exist an $M \times N$ matrix of constants Γ and an N-vector of random variables η, such that the following conditions hold:

$$\begin{aligned} (F-\mu_F) &= \Gamma^T (Z-\mu_Z)+\eta && (2.7)\\ E[\eta] &= 0_{Nx1} && (2.8)\\ Cov[Z,\eta] &= 0_{MxN} && (2.9)\end{aligned}$$

The matrices β and Γ are related by the following identities:

$$\begin{aligned} \beta &= \Sigma_{FF}^{-1}\Gamma^T\Sigma_{ZZ} && (2.10)\\ \Gamma &= \Sigma_{ZZ}^{-1}\beta^T\Sigma_{FF} && (2.11)\end{aligned}$$

from which it follows that Γ is of rank N. We usually denote by P the projection of the factors onto the asset space:

$$P=\Gamma^T Z \tag{2.12}$$

and denote by Σ_{PP} and Σ_{PZ} the covariance matrix of the projection P, and the covariance between the projection P and the assets Z. We refer to a factor F_i, with $1 \leq i \leq N$, as "spanned" if $Var[\eta_i]=0$ (i.e., a spanned factor may differ from its projection only by a constant). Similarly, a linear combination of factors $w^T F$ is considered spanned if $Var\left(w^T\eta\right)=0$.

It follows immediately from the definition of the projection P that:

$$\begin{aligned} \Sigma_{PZ} &= \Sigma_{FZ} && (2.13)\\ \Sigma_{PP} &= \Sigma_{FZ}\Sigma_{ZZ}^{-1}\Sigma_{ZF} && (2.14)\end{aligned}$$

The risk premia vector $\gamma(F,Z)$ from Equation 2.6 can now be rewritten as:

$$\gamma(F,Z)=\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z \tag{2.15}$$

We do not necessarily assume that the covariance matrix of η, denoted by $\Sigma_{\eta\eta}$, is of full rank. Since Z and η are uncorrelated:

$$\Sigma_{FF}=\Sigma_{PP}+\Sigma_{\eta\eta} \tag{2.16}$$

Equation 2.15 can then also be written as:

$$\gamma(F,Z)=\Gamma^T\mu_Z+\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Gamma^T\mu_Z \tag{2.17}$$

Although the definition of the vector of risk premia from Equations 2.6 and 2.15 is motivated by linear factor models (i.e., models that explain the expected excess returns of all assets correctly), it is possible to extend the definition to apply to any set of factors and any set of assets that jointly satisfy the full-rank assumptions.

Definition 3. *Let F be a set of factors, and let Z be a set of assets such that F and Z satisfy the full-rank assumptions. Then the vector of risk premia of F with respect to Z is defined by Equation 2.15, whether or not F is a linear factor model for Z.*

When F is a linear factor model for Z, the motivation for defining the risk premia vector in this way is clear; with this choice, the factors describe the expected excess returns of all assets perfectly. When F is not a linear factor model for Z, simply applying the definition from Equation 2.15 may seem somewhat arbitrary. However, a researcher will not necessarily know a priori whether a particular set of factors is a linear factor model for a given set of assets, and statistical tests to that point may well prove inconclusive. It is therefore useful to have a definition that can be applied in the face of uncertainty, but that is consistent with Equation 2.15 when F is a linear factor model for Z. Furthermore, as shown in Section 5, this definition is equivalent to that obtained by a two-pass GLS regression technique. The risk premia vector γ minimizes the distance between the predicted and actual vectors of expected excess returns, if distance is measured with respect to the matrix Σ_{ZZ}^{-1}:

$$\gamma = \operatorname*{argmin}_{\gamma_0 \in \mathbb{R}^N} \left(\mu_Z - \beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1} \left(\mu_Z - \beta^T \gamma_0\right) \tag{2.18}$$

This definition is invariant to the linear transformations of the asset space considered by Kandel and Stambaugh (1995); for choices of the distance matrix other than Σ_{ZZ}^{-1}, this is not necessarily the case. Finally, if the vector μ_Z is estimated from a time series of i.i.d. observations of Z, its covariance matrix is given by $\frac{1}{n}\Sigma_{ZZ}$, with n equal to the number of observations. The inverse of this matrix is therefore a reasonable choice of distance matrix. Both Shanken (1992) and (in the case of a single spanned factor) Kandel and Stambaugh (1995) point out the desirable econometric features of using Σ_{ZZ}^{-1} as a GLS weighting matrix in the second pass when estimating γ by two-pass linear regression.

With a risk premia vector defined whether or not a set of factors is a linear factor model for a given set of assets, it is convenient to define the deviations of the actual expected excess returns of a set of assets from the fitted expected excess returns:

$$\varepsilon\left(F, Z\right) = \mu_Z - \beta^T \gamma\left(F, Z\right) \tag{2.19}$$

A set of factors F is thus a linear factor model for a set of M assets Z if and only if F and Z satisfy the full-rank assumptions and $\varepsilon\left(F, Z\right) = 0_{M \times 1}$.

Although we have defined risk premia for any set of factors and any set of assets, as long as they satisfy the full-rank assumptions, we will be interested in determining when the risk premium assigned to an individual factor is invariant to the additional or removal of other factors, and also when this risk premium is invariant to changes in the asset space. When a factor is spanned by the assets, Definition 3 always assigns the same risk premium to that factor, regardless of the other factors included. This result is an immediate consequence of an argument presented in Shanken (1992). However, it is possible to prove a more general result, as per the following theorem.

Theorem 1. *Let F and G be two sets of factor realizations, and let Z be a set of asset payoffs, such that F*

and Z satisfy the full-rank assumptions, and G and Z also satisfy the full-rank assumptions. For any vectors x, y, and z such that:

$$x^T (F - \mu_F) = y^T (G - \mu_G) = z^T (Z - \mu_Z) \tag{2.20}$$

the risk premia vectors $\gamma(F, Z)$ and $\gamma(G, Z)$ satisfy:

$$x^T \gamma(F, Z) = y^T \gamma(G, Z) = z^T \mu_Z = x^T \Gamma_F^T \mu_Z = y^T \Gamma_G^T \mu_Z \tag{2.21}$$

Proof: See Appendix.

Theorem 1 states that any spanned factor, or any spanned linear combination of factors, must have the same risk premium, regardless of the other factors included in the model, and regardless of whether the full set of factors is a linear factor model. Furthermore, the risk premium of a spanned factor (or spanned linear combination of factors) is equal to the expected excess return of its projection onto the asset space. Since this value does not depend on the other factors included in a factor set, it is meaningful to speak of the risk premium of a spanned factor, independently of the other factors needed to form a linear factor model.

We also note that, in an economy free of arbitrage opportunities, the risk premium of a spanned factor is invariant to changes in the asset space (provided the factor does not cease to be spanned as a result of the change). If the same factor is spanned by two different sets of assets, its projections onto the two asset spaces must have the same expected excess return, and therefore the same risk premium as specified by Theorem 1. The risk premium of a spanned factor is therefore not only invariant to the other factors included in a model, but also to the particular choice of spanning assets. Furthermore, if a factor is itself a traded asset [e.g., the excess return of the market portfolio in the CAPM of Sharpe (1964) and Lintner (1965)], since it spans itself, its risk premium is simply equal to its expected excess return.[1] The well-known finding of Fama and French (1992), that the risk premium of the market portfolio is approximately zero or slightly negative during a sample period when the market portfolio experienced large positive returns, can thus be called into question purely on theoretical grounds; the risk premium of the market portfolio in any well-specified linear factor model must be equal to its expected excess return. This phenomenon is further studied by Roll and Ross (1994).

Assignment of risk premia to unspanned factors without reference to the other factors in a model is a much trickier business. The matrices Σ_{FZ} and Σ_{FF} can be expressed as:

$$\Sigma_{FZ} = \Gamma^T \Sigma_{ZZ} \tag{2.22}$$

$$\Sigma_{FF} = \Gamma^T \Sigma_{ZZ} \Gamma + \Sigma_{\eta\eta} \tag{2.23}$$

The covariances between the factors F and the assets Z are identical to the covariances between P and Z,

---

[1] This point is made by Shanken (1992), who in two-pass regression estimation constrains the risk premium of a spanned factor (but not necessarily spanned linear combinations of factors) to be equal to the expected return of its spanning portfolio. As we shall see, this constraint is unnecessary provided an appropriate GLS weighting matrix is used; in this case, the results of the estimation procedure automatically satisfy the constraint. See, for example, Kandel and Stambaugh (1995), who consider two-pass regression estimation of a model with a single spanned factor.

but the two sets of factors have different covariance matrices:

$$\begin{aligned} \Sigma_{PZ} &= \Sigma_{FZ} && (2.24) \\ \Sigma_{PP} &= \Gamma^T \Sigma_{ZZ} \Gamma = \Sigma_{FF} - \Sigma_{\eta\eta} && (2.25) \end{aligned}$$

The β coefficients of the assets on F and P (denoted by β_F and β_P, respectively) are related as follows:

$$\beta_F = \Sigma_{FF}^{-1}\Sigma_{FZ} = \left(\Sigma_{PP} + \Sigma_{\eta\eta}\right)^{-1}\Sigma_{FZ} = \left(I + \Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right)^{-1}\Sigma_{PP}^{-1}\Sigma_{PZ} = \left(I + \Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right)^{-1}\beta_P \tag{2.26}$$

Whether or not F and P are linear factor models for Z, the risk premia $\gamma(F, Z)$ and $\gamma(F, P)$ are related by:

$$\gamma(F, Z) = \Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z = \Sigma_{FF}\Sigma_{PP}^{-1}\gamma(P, Z) = \gamma(P, Z) + \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\gamma(P, Z) \tag{2.27}$$

It follows that F and P predict the same expected excess returns for all assets, whether or not F and P are linear factor models:

$$\beta_F^T\gamma(F, Z) = \Sigma_{ZF}\Sigma_{FF}^{-1}\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z = \Sigma_{ZP}\Sigma_{PP}^{-1}\Gamma^T\mu_Z = \beta_P^T\gamma(P, Z) \tag{2.28}$$

It is also readily apparent that F constitutes a linear factor model if and only if its projection P onto the asset space is a linear factor model.

From Equation 2.27, we see that two linear factor models with the same projection onto the asset space do not necessarily have the same vector of risk premia. The following theorem shows how much the vector of risk premia can vary for different models with the same projection.

Theorem 2. *Let Z be a vector of M asset payoffs, and let F be a set of N factors such that F is a linear factor model for Z with projection P onto the asset space. The vector of risk premia $\gamma(F, Z)$ is either within the open half-space of $\mathbb{R}^N$ defined by:*

$$\gamma(P, Z)^T\Sigma_{PP}^{-1}\gamma(F, Z) > \gamma(P, Z)^T\Sigma_{PP}^{-1}\gamma(P, Z) \tag{2.29}$$

or equal to the point $\gamma(P, Z)$ (which lies on the boundary of the half-space). Conversely, let $P = \Gamma^T Z$ be a set of N factors such that P is a linear factor model for Z. Given any N-vector γ that satisfies either:

$$\gamma(P, Z)^T\Sigma_{PP}^{-1}\gamma > \gamma(P, Z)^T\Sigma_{PP}^{-1}\gamma(P, Z) \tag{2.30}$$

or $\gamma = \gamma(P, Z)$, there exists a linear factor model F for Z with N factors, such that F has the projection P onto the asset space, and $\gamma = \gamma(F, Z)$.

Proof: See Appendix.

As previously noted, the vector of risk premia for a given set of factors F and a given set of assets Z is unique and well-defined whenever F and Z satisfy the full-rank assumptions, and in particular when F is a linear factor model for Z. However, varying the unspanned components of F (while holding the spanned components fixed) can result in assignment of a nearly arbitrary vector of risk premia to F. As shown in Theorem 2, models with the same projection onto the asset space can have very different risk premia; any

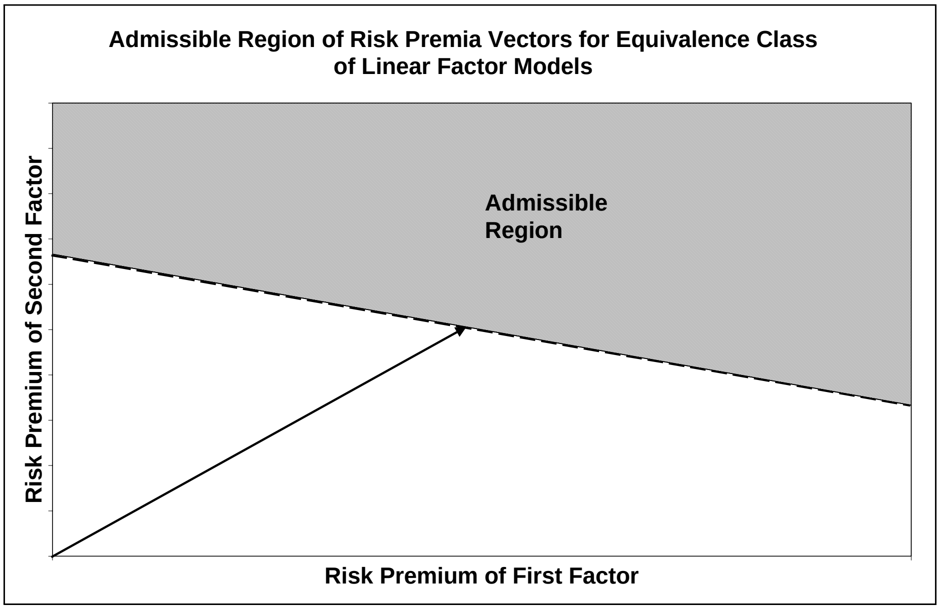

Figure 1: This figure shows a stylized representation of the admissible region for the vector of risk premia for a two factor model. The arrow shows the location of $\gamma(P, Z)$, which lies on the boundary of the half space. Any linear factor model for the assets Z with the projection P has a vector of risk premia within the labeled admissible region. The only point on the boundary that is included in the admissible region is $\gamma(P, Z)$ itself. Conversely, for any given point γ within the admissible region, there exists a linear factor model with projection P onto the assets with risk premia vector equal to γ.

point within an open half-space is achievable by varying the unspanned components associated with each factor. Figure 1 provides a graphical illustration of the subset of $\mathbb{R}^2$ that can be achieved in a two-factor model by adding unspanned components to the factors. This level of flexibility makes it impossible (without additional restrictions or assumptions) to define the risk premium of an unspanned factor independently of the other factors in the model, as is possible when factors are spanned. As shown in the next theorem, for any set of factors that is not a linear factor model, there exists a linear factor model containing the given set of factors, which assigns essentially any arbitrarily specified vector of risk premia to these factors, subject only to the constraints imposed on spanned factors (and spanned linear combinations of factors) by Theorem 1.

Theorem 3. *Let Z be a vector of M asset payoffs, and let F be a vector of N factor realizations, such that F and Z satisfy the full-rank assumptions, but such that F is not a linear factor model for Z. Let P be the projection of F onto the assets Z. Let γ be any N-vector such that $w^T\left[\gamma - \gamma\left(P, Z\right)\right] = 0$ for any N-vector w with $w^T \Sigma_{\eta\eta} w = 0$. Then there exists a linear factor model H for Z such that $\left(F - \mu_F\right) = \Psi^T\left(H - \mu_H\right)$ for some matrix Ψ, and such that $\gamma = \Psi^T \gamma\left(H, Z\right)$.*

Proof: See Appendix.

From Theorem 1, risk premia assigned to spanned factors are invariant to the addition of new factors. Consequently, if the rank of the $\Sigma_{\eta\eta}$ matrix is K, any linear factor model that includes F (in the sense that each element of F is a linear combination of the factors in the linear factor model) must assign to F a risk premia vector that lies within a K-dimensional hyperplane of $\mathbb{R}^N$. Theorem 3 shows that any point within this K-dimensional hyperplane can be achieved for some set of additional factors. Note that, if $K = 0$, then $w^T \Sigma_{\eta\eta} w = 0$ for any w, and the only value of γ that satisfies the conditions of Theorem 3 is $\gamma\left(P, Z\right)$, consistent with the previous result that risk premia of spanned factors are not dependent on the other factors in the model. If $K = N$, no factor (or linear combination of factors) is spanned, and the only vector w for which $w^T \Sigma_{\eta\eta} w = 0$ is $w = 0$. In this case, any vector γ satisfies the conditions of the theorem, and there exists a linear factor model containing F which assigns to F the specified risk premia vector. When no linear combination of factors is spanned, assignment of risk premia to F is completely arbitrary.

Theorem 3 requires that F not be a linear factor model for Z. By contrast, the next theorem shows that, if F is a linear factor model for Z, then the risk premia assigned to the elements of F are invariant to the introduction of additional factors, provided the new factors do not lead to a violation of the full-rank assumptions. Such factors, are, of course, unnecessary,[2] since F is already a linear factor model, but an econometrician most likely does not know a priori which factors are necessary and which are not. The risk premia assigned are also invariant to the removal of factors, provided the remaining factors still constitute a linear factor model.

---

[2] Kan and Zhang (1999a) and Kan and Zhang (1999b) discuss GMM and regression estimation, respectively, in models with "useless" factors, where "useless" is defined as being uncorrelated with any of the assets under study. Such factors are specifically precluded here, since they result in violations of the full-rank assumptions. Rather, the factors considered here are still correlated with the asset excess returns (from the full rank assumptions), but are considered unnecessary because, after their removal, the remaining factors are still a linear factor model.

Theorem 4. *Let F be linear factor model for Z, and let G be a set of factors such that G and Z satisfy the full-rank assumptions. If there exists a matrix Ψ such that $(F - \mu_F) = \Psi^T (G - \mu_G)$, then G is a linear factor model for Z and $\gamma(F, Z) = \Psi^T \gamma(G, Z)$.*

Proof: See appendix.

Theorem 4 shows that risk premia for factors in a linear factor model are invariant to the addition or removal of new factors, provided that the new set of factors is also a linear factor model. By contrast, Theorem 3 shows that the risk premia vector of an unspanned set of factors that does not constitute a linear factor model is highly dependent on the other factors included. Therefore, it is only meaningful to speak of the risk premium of a factor, in isolation of other factors, when that factor is spanned, or when it is a linear factor model by itself. When a set of unspanned factors is not a linear factor model for the assets, introduction of additional factors can result in the assignment of any arbitrary risk premia vector to the current factors.

3 Asset Space Selection

The results of the previous section tell us that an unspanned factor does not have an unambiguously defined risk premium (unless that factor is a linear factor model by itself); rather the risk premium depends on the other factors included in the model. If a factor is unspanned because the set of assets included in a study does not span the full universe of assets available, we might also wonder about the invariance of its risk premium to changes in the asset space. Omission of some assets may in fact be inevitable, if investors select portfolios based on information not available to the econometrician. In this situation, it is possible to construct asset pricing models based on the unconditional distribution of asset excess returns and factor realizations, provided the notion of "asset" includes conditional trading strategies as well; see Hansen and Richard (1987). The inclusion of additional information in a study then expands the space of "assets" available, possibly changing the risk premia vector assigned to a set of factors; even if a given set of factors is a linear factor model for a given set of assets, it may cease to be a linear factor model when the asset space is expanded. We therefore consider the effect of asset omission on the assignment of risk premia, and find that even seemingly innocuous changes in the asset space can result in significant changes in the risk premia assigned to unspanned factors.

First, we make precise the notion of "asset omission."

Definition 4. *Let Z be a set of $M > 1$ asset payoffs with a finite and full-rank covariance matrix Σ_{ZZ}, and let $Y = \Psi^T Z$ for some $M \times K$ full-rank matrix Ψ with $K < M$. We define the Y-reduction of Z, denoted by $Z \ominus Y$, as the equivalence class of all sets of asset payoffs $X = \Phi^T Z$ for all full-rank $M \times (M - K)$ matrices Φ such that $\Psi^T \Sigma_{ZZ} \Phi = 0_{K \times (M-K)}$.*

Each member of $Z \ominus Y$ is therefore a basis for the space containing all excess returns that are uncorrelated with Y. $Z \ominus Y$ is well-defined for any full-rank Ψ; however, the members of $Z \ominus Y$ may fail to satisfy the full-rank assumptions for a given set of factors F, even if F and Z do satisfy the full-rank assumptions.

Lemma 2. *Let F be a set of N factor realizations, and let Z be a set of $M > 1$ asset payoffs, such that F and Z satisfy the full-rank assumptions. Let $w \neq 0$ be an M-vector, and let Y be any member of $Z \ominus w^T Z$. F*

and Y satisfy the full-rank assumptions if and only if $w \neq \Gamma x$ for all N-vectors x.

Proof: See Appendix.

Since, for a given Z, F, and w, the full-rank assumptions are satisfied for all members of $Z \ominus w^T Z$ or for none, we say F and $Z \ominus w^T Z$ satisfy the full-rank assumptions, without reference to a particular member of $Z \ominus w^T Z$. Each member of this equivalence class is a different basis for the same $(M-1)$-dimensional subspace of the space spanned by Z; furthermore, all assets in the subspace have payoffs that are uncorrelated with the payoff of $w^T Z$. Reduction of the asset space by a single dimension will be the standard tool used throughout this section. We may construct a reduced space for any $w \in \mathbb{R}^M$ such that $w \neq 0$; however, w must lie outside the N-dimensional subspace of $\mathbb{R}^M$ spanned by the columns of Γ if F and $Z \ominus w^T Z$ are to satisfy the full-rank assumptions. As we have already noted in the previous section, two bases for the same asset space assign the same risk premia vector to a set of factors. Therefore, if F and $Z \ominus w^T Z$ satisfy the full-rank assumptions, we may refer to the risk premia vector $\gamma\left(F, Z \ominus w^T Z\right)$ given by the common value $\gamma(F, Y)$ where Y is any member of $Z \ominus w^T Z$. We also note that, for a given F, Z, and w, F is a linear factor model for all members of $Z \ominus w^T Z$ or for none. We therefore speak of F being a linear factor model for $Z \ominus w^T Z$, without reference to a particular element of $Z \ominus w^T Z$.

We now find that, if F is not a linear factor model for Z, then for any arbitrary vector γ_0, it is possible to remove a single asset from the asset space, such that F is a linear factor model for the remaining assets with risk premia vector given by γ_0. This result shows that omission of even a single asset can dramatically change the risk premia vector assigned to a set of factors.

Theorem 5. *Let F be a set of N factors, and let Z be a set of M assets, such that F and Z satisfy the full-rank assumptions, but such that F is not a linear factor model for Z. Let γ_0 be an N-element vector of constants. Then there exists an M-vector w such that $w \neq \Gamma x$ for any N-vector x, and such that F is a linear factor model for $Z \ominus w^T Z$ with $\gamma\left(F, Z \ominus w^T Z\right) = \gamma_0$. Furthermore, if w_0 is an M-vector such that $w_0 \neq \Gamma x$ for all N-vectors x and F is a linear factor model for $Z \ominus w_0^T Z$ with $\gamma\left(F, Z \ominus w_0^T Z\right) = \gamma_0$, then $w_0 = c \cdot w$ for some constant $c \neq 0$.*

Proof: See Appendix.

Theorem 5 illustrates the difficulty of attempting to assign risk premia to a set of factors when the set of assets observed is incomplete. If a set of factors explains the expected return of *every* asset *perfectly* (i.e., if the factor set is a linear factor model for the set of all available assets), then that set of factors has a unique and well-defined vector of risk premia; omission of some assets does not affect the risk premia vector (as long as the full-rank assumptions are still satisfied). However, should the set of factors predict a slightly incorrect expected excess return for even one asset, then that factor set can be assigned an arbitrary risk premia vector by omission of a single asset. Furthermore, the factor set will explain the expected excess returns for all assets in the reduced set perfectly. Consequently, the same set of factors can be a linear factor model for two sets of assets which span spaces with an $(M-2)$-dimensional intersection, and yet have very different vectors of risk premia for the two asset sets.

Through omission of an asset, it is possible to make an arbitrary set of factors F into a linear factor model for the reduced asset space with an arbitrary risk premium vector, provided F was not already a linear factor model for the original set of assets. One might suspect that, for some risk premia vectors, the necessary contraction of the asset space is severe, in the sense that the reduced space offers investors much less favorable investment opportunities than the original space. We use the maximum squared Sharpe ratio offered by reduced set of assets, relative to the maximum squared Sharpe ratio offered by the original set, as a measure of the severity of the contraction of the asset space.

Lemma 3. *Let Z be a set of M assets with finite and full-rank covariance matrix Σ_{ZZ}. The maximum squared Sharpe ratio offered by a set of assets Z, denoted by $S^2(Z)$, is equal to:*

$$S^2(Z) = \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z \tag{3.1}$$

For $M > 1$, let $Y = \Psi^T Z$ for some full-rank $M \times K$ matrix Ψ, with $K < M$. The maximum squared Sharpe ratio offered by $Z \ominus Y$, denoted by $S^2(Z \ominus Y)$, is equal to:

$$S^2(Z \ominus Y) = S^2(Z) - S^2(Y) = \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \mu_Z^T \Psi \left(\Psi^T \Sigma_{ZZ} \Psi\right)^{-1} \Psi^T \mu_Z \tag{3.2}$$

Proof: See Appendix.

The following theorem provides a measure of the severity of the contraction of the asset space described in Theorem 5.

Theorem 6. *Let F be a set of N factors, and Z a set of M assets, such that F and Z satisfy the full-rank assumptions, but such that F is not a linear factor model for Z. Let γ_0 be an N-element vector of constants. Let w be an M-vector, such that $w \neq \Gamma x$ for all N-vectors x, and such that F is a linear factor model for $Z \ominus w^T Z$ with $\gamma\left(F, Z \ominus w^T Z\right) = \gamma_0$. Then:*

$$S^2\left(Z \ominus w^T Z\right) = \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \frac{\left[\mu_Z^T \Sigma_{ZZ}^{-1} \left(\mu_Z - \beta^T \gamma_0\right)\right]^2}{\left(\mu_Z - \beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1} \left(\mu_Z - \beta^T \gamma_0\right)} \tag{3.3}$$

Proof: See Appendix.

Theorem 6 allows us to calculate directly the reduction in squared Sharpe ratio associated with the procedure described in Theorem 5. Note that this reduction depends on γ_0, but not on the particular choice of w used (recall, as per Theorem 5, all such choices differ only by a scaling factor). It is convenient to normalize $S^2\left(Z \ominus w^T Z\right)$ by $S^2(Z)$.

Definition 5. *Let F be a set of N factors and let Z be a set of M assets, such that F and Z satisfy the full-rank assumptions, but such that F is not a linear factor model for Z. Let γ_0 be an N-element vector of constants. We define the Squared Sharpe Ratio Reduction, denoted by $SSRR(F, Z, \gamma_0)$:*

$$SSRR(F, Z, \gamma_0) = \frac{S^2\left(Z \ominus w^T Z\right)}{S^2(Z)} \tag{3.4}$$

where w is an M-vector such that $w \neq \Gamma x$ for all N-vectors x, and such that F is a linear factor model for $Z \ominus w^T Z$ with $\gamma\left(F, Z \ominus w^T Z\right) = \gamma_0$.

The value of $SSRR\left(F,Z,\gamma_0\right)$ could conceivably range from 0 to 1. We might be inclined to think that there is a vector of risk premia that requires the smallest possible reduction, with increasingly larger reductions required for risk premia vectors that are increasingly far away from this minimum value. The reality of the situation is considerably more complex, as shown in the next theorem.

Theorem 7. *Let F be a set of N factors and Z a set of M assets, such that F and Z satisfy the full-rank assumptions, but such that F is not a linear factor model for Z. Let k_0 be the constant defined by:*

$$k_0 = 1 - \frac{\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z}{\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z} \tag{3.5}$$

Let k be some constant with $0 \leq k \leq 1$. For $N > 1$, the set of all $\gamma_0 \in \mathbb{R}^N$ such that $SSRR\left(F,Z,\gamma_0\right) = k$ is:

1. *The single point $\gamma_0 = 0_{N\times 1}$ if $k = 0$.*
2. *The $(N-1)$-dimensional hyperplane $\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z = \mu_Z^T \Sigma_{ZZ}^{-1} \beta^T \gamma_0$ if $k = 1$.*

For all other values of k, $SSRR\left(F,Z,\gamma_0\right) = k$ is equivalent to:

$$\left[\begin{array}{c} (1-k)\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)\left(\gamma_0^T \beta \Sigma_{ZZ}^{-1} \beta^T \gamma_0\right) + \left(\mu_Z^T \Sigma_{ZZ}^{-1} \beta^T \gamma_0\right)^2 \\ -k\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)^2 + 2k\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)\left(\mu_Z^T \Sigma_{ZZ}^{-1} \beta^T \gamma_0\right) \end{array}\right] = 0 \tag{3.6}$$

The set of all $\gamma_0 \in \mathbb{R}^N$ satisfying this equation for $k \neq 0$ and $k \neq 1$ is:

3. *An hyperboloid with branches on either side of the $k = 1$ hyperplane if $k_0 < k < 1$,*
4. *A paraboloid that separates the $k = 1$ hyperplane and the origin if $k = k_0$,*
5. *An ellipsoid containing the origin if $0 < k < k_0$.*

For $N = 1$, the hyperplane and paraboloid collapse to single points, and the hyperboloids and ellipsoids collapse to pairs of points.

Proof: See Appendix.

Theorem 7 provides a geometric interpretation for the analytic results of Theorem 6; the geometric objects are loci of vectors of risk premia for which it is equally difficult to construct a set of assets Y such that F is a linear factor model for Y (with the specified risk premia vector), where "difficulty" is measured by the reduction in squared Sharpe ratio needed. Figure 2 shows the vectors of risk premia that can be achieved for a variety of values of Squared Sharpe Ratio Reduction in a two-factor model.

The procedure described in Theorem 7 considers removal from the asset space of an asset $w^T Z$ where w lies within an N-dimensional hyperplane of $\mathbb{R}^M$; the set of factors is then a linear factor model for the reduced set of assets. If some asset $w^T Z$ is removed for a w that lies outside of this subspace, then F is not a linear factor model for the remaining assets, but, as per Definition 3, we can nonetheless assign risk premia to the factors. There are then typically many assets whose removal results in the assignment of the same risk premia vector to the factors. We might suspect that dropping the requirement that F be a linear factor model for the

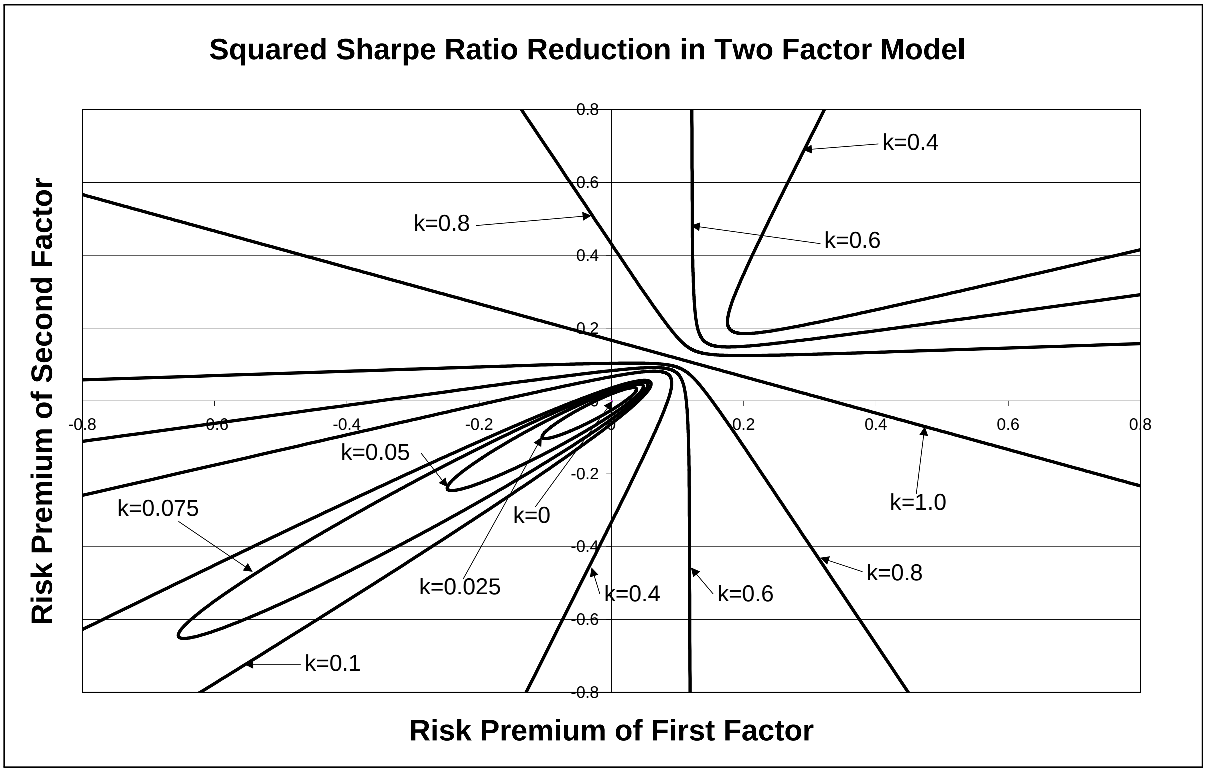

Figure 2: This figure shows the reduction in squared Sharpe ratio [i.e., the value of $SSRR\,(F, Z, \gamma_0)$] needed to make a given set of factors a linear factor model for a given set of assets. In this example, there are twenty excess returns, each with a mean of 0.1 and a variance of 0.8, and each uncorrelated with the others. There are two factors; one is the average value of ten of the asset returns, and the other is the average value of fifteen of the asset returns. Five asset returns are included in both factors. The mean-variance efficient portfolio is the average of all twenty asset returns, has a squared Sharpe ratio of 0.25, and is not spanned by the factors. The most difficult (in terms of squared Sharpe ratio reduction) vector of risk premia to achieve is the origin; the 19 assets remaining after the construction offer a maximum squared Share ratio of zero. The ellipses surrounding the origin require a slightly smaller reduction to achieve, while risk premia on the parabola can be achieved with a maximum squared Sharpe ratio of 0.025 (i.e., a reduction of 90% relative to the full set of 20 assets). Points on the hyperbola require progressively smaller reductions, and points on the line can be achieved with no reduction in squared Sharpe ratio at all.

remaining assets will make it easier (i.e., a smaller reduction in squared Sharpe ratio is needed) to construct a set of assets that assigns a desired vector of risk premia to the factors. This suspicion need not be correct, although it often will be for particular values of γ_0. We first characterize the set of assets whose removal results in a given vector of risk premia being assigned to the factors. There are two distinct cases to consider; the following theorem characterizes the situation when the given vector of risk premia is equal to the risk premia vector assigned without the removal of any assets.

Theorem 8. *Let Z be a set of M assets, and let F be a set of N factors, such that F and Z satisfy the full-rank assumptions, but such that F is not a linear factor model for Z. Let w be an M-vector such that $w \neq \Gamma x$ for any N-vector x. Then $\gamma\left(F, Z \ominus w^T Z\right) = \gamma(F, Z)$ if and only if either (i) $\mu_Z^T \left(\Sigma_{ZZ}^{-1} - \Gamma \Sigma_{PP}^{-1} \Gamma^T\right) \Sigma_{ZZ} w = 0$, or (ii) $\Sigma_{FZ} w = 0_{N \times 1}$.*

Proof: See Appendix.

Theorem 8 identifies those assets whose removal from Z does not change the risk premia vector assigned to F, i.e., those assets $w^T Z$ with the property that $\gamma\left(F, Z \ominus w^T Z\right) = \gamma(F, Z)$. These assets occupy two regions. Those satisfying criterion (i) lie inside an $(M-1)$-dimensional subspace of $\mathbb{R}^M$, but outside an N-dimensional subspace (wholly contained within the first subspace); this region is empty if $M = N+1$. Those assets satisfying criterion (ii) lie within an $(M-N)$-dimensional subspace of $\mathbb{R}^M$, excluding the origin. The two regions do not overlap. If the construction from Theorem 5 is applied to the case where $\gamma = \gamma(F, Z)$, then $w = \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma(F, Z)\right) = \Sigma_{ZZ}^{-1} \mu_Z - \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z$. This choice of w satisfies criterion (ii) in Theorem 8, and, as per the statement of Theorem 5, makes F a linear factor model for $Z \ominus w^T Z$.

The set of all assets whose removal results in a given risk premia vector different than $\gamma(F, Z)$ is not a union of subspaces, but rather a cone whose cross-sections are paraboloids,[3] as per the next theorem.

Theorem 9. *Let Z be a set of M assets, and let F be a set of N factors, such that F and Z satisfy the full-rank assumptions, but such that F is not a linear factor model for Z. Let $\gamma_0 \neq \gamma(F, Z)$ be an N-vector, and let w be an M-vector such that $w \neq \Gamma x$ for any N-vector x. Then F and $Z \ominus w^T Z$ satisfy the full-rank assumptions with $\gamma\left(F, Z \ominus w^T Z\right) = \gamma_0$ if and only if w can be expressed as:*

$$w = w(c, \zeta) = \left[\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right) + \zeta + \Gamma \Sigma_{PP}^{-1} \Gamma^T \left(\mu_Z - \beta^T \gamma_0\right) \left(\frac{\zeta^T \Sigma_{ZZ} \zeta}{\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z}\right)\right] c \tag{3.7}$$

for some constant $c \neq 0$ and some M-vector ζ such that $\mu_Z^T \zeta = 0$ and $\Sigma_{FZ} \zeta = 0_{N \times 1}$. If $M = N+1$, all such w are of the form:

$$w = c \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right) \tag{3.8}$$

and F is a linear factor model for $Z \ominus w^T Z$. If $M > N+1$, then for a fixed value of c, the set of all such w is a paraboloid, and F is a linear factor model for $Z \ominus w^T Z$ if and only if $\zeta = 0_{M \times 1}$.

Proof: See Appendix.

---

[3] This particular characterization is only one of several possible. By changing the orientation of the intersecting hyperplanes, it is possible to change the shape of the cross-sections.

Theorem 5 shows how removal of a particular asset from the asset space can assign a given vector of risk premia to the factors, which constitute a linear factor model for the remaining assets; the reduction in squared Sharpe ratio associated with this asset removal is described in Theorem 7. By contrast, Theorems 8 and 9 describe the set of all assets whose removal assigns a given vector of risk premia to the factors, whether or not the factors are then a linear factor model for the remaining assets. One might suspect that the weaker conditions of Theorem 9 can be met with a smaller reduction in squared Sharpe ratio than in Theorem 5. This will often be the case, but is not necessarily so, as shown by the following theorem.

Theorem 10. *Let F be a set of N factors, and let Z be a set of $M > N + 1$ asset payoffs, such that F and Z satisfy the full-rank assumptions, but such that F is not a linear factor model for Z. Let $\gamma_0 \neq \gamma(F, Z)$ be an N-vector, and let $w(c, \zeta)$ be as defined in Equation 3.7.*

1. *If the vector γ_0 satisfies the condition:*

$$\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z \leq \mu_Z^T \Sigma_{ZZ}^{-1} \beta^T \gamma_0 \tag{3.9}$$

then the value of ζ that maximizes $S^2\left(Z \ominus w(c,\zeta)^T Z\right)$ is $\zeta = 0_{M\times 1}$. In this case, F is a linear factor model for $Z \ominus w(c,0)^T Z$, and:

$$S^2\left(Z \ominus w(c,0)^T Z\right) = \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \frac{\left[\mu_Z^T \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)\right]^2}{\left(\mu_Z - \beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)} \tag{3.10}$$

This region consists of a closed half-space of $\mathbb{R}^N$.

2. *If the vector γ_0 satisfies the condition:*

$$\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z < \mu_Z^T \Sigma_{ZZ}^{-1} \beta^T \gamma_0 < \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z \tag{3.11}$$

then $S^2\left(Z \ominus w(c,\zeta)^T Z\right)$ is maximized by any ζ that satisfies:

$$\zeta^T \Sigma_{ZZ} \zeta = -\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right) \frac{\mu_Z^T \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)}{\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T\left(\mu_Z - \beta^T \gamma_0\right)} \tag{3.12}$$

In this case, $\gamma\left(F, Z \ominus w(c,\zeta)^T Z\right) = \gamma_0$, but F is not a linear factor model for $Z \ominus w(c,\zeta)^T Z$. The squared Sharpe ratio offered by the remaining assets is the same as the squared Sharpe ratio offered by the full set:

$$S^2\left(Z \ominus w(c,\zeta)^T Z\right) = \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z \tag{3.13}$$

This region lies between two $(N-1)$-dimensional hyperplanes.

3. *If the vector* γ_0 *satisfies both of the conditions:*

$$\begin{aligned}\mu_Z^T \Sigma_{ZZ}^{-1} \beta^T \gamma_0 &\leq \mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z \quad (3.14)\\ 9\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - 8\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z &\leq \left(3\mu_Z - 4\beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1} \left(3\mu_Z - 4\beta^T \gamma_0\right) \quad (3.15)\end{aligned}$$

then there is no value of ζ *that maximizes* $S^2\left(Z \ominus w\left(c,\zeta\right)^T Z\right)$*. However,* $S^2\left(Z \ominus w\left(c,\zeta\right)^T Z\right)$ *is bounded above, and the bound may be approached arbitrarily closely by choosing* ζ *such that* $\zeta^T \Sigma_{ZZ} \zeta$ *is arbitrarily large.* F *is not a linear factor model for* $Z \ominus w\left(c,\zeta\right)^T Z$ *(provided* $\zeta \neq 0$*), and the limiting value of* $S^2\left(Z \ominus w\left(c,\zeta\right)^T Z\right)$ *is given by:*

$$\lim_{\left(\zeta^T \Sigma_{ZZ} \zeta\right) \to +\infty} S^2\left(Z \ominus w\left(c,\zeta\right)^T Z\right) = \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \frac{\left[\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \left(\mu_Z - \beta^T \gamma_0\right)\right]^2}{\left(\mu_Z - \beta^T \gamma_0\right)^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \left(\mu_Z - \beta^T \gamma_0\right)} \quad (3.16)$$

This region is the set of all points lying within a closed half-space of $\mathbb{R}^N$*, but outside an open ellipsoid lying entirely within the closed half-space. The interior of the ellipsoid may not exist (i.e., the ellipsoid may consist of a single point or may be the empty set).*

4. *If the vector* γ_0 *satisfies the condition:*

$$\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z < \left(3\mu_Z - 4\beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1} \left(3\mu_Z - 4\beta^T \gamma_0\right) < 9\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - 8\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z \quad (3.17)$$

then $S^2\left(Z \ominus w\left(c,\zeta\right)^T Z\right)$ *is maximized by any* ζ *that satisfies:*

$$\zeta^T \Sigma_{ZZ} \zeta = -\frac{\left(\mu_Z - \beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1} \left(\mu_Z - 2\beta^T \gamma_0\right)}{\left(\mu_Z - \beta^T \gamma_0\right)^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \left(\mu_Z - 2\beta^T \gamma_0\right)} \quad (3.18)$$

In this case, $\gamma\left(F, Z \ominus w\left(c,\zeta\right)^T Z\right) = \gamma_0$*,* F *is not a linear factor model for* $Z \ominus w\left(c,\zeta\right)^T Z$*, and:*

$$S^2\left(Z \ominus w\left(c,\zeta\right)^T Z\right) = \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - 4\gamma_0^T \beta \Sigma_{ZZ}^{-1} \left(\mu_Z - \beta^T \gamma_0\right) \quad (3.19)$$

This region is the open interior of the previously referenced ellipse, but outside a smaller ellipse. Either the smaller ellipse or both ellipses may be degenerate.

5. *If the vector* γ_0 *satisfies the condition:*

$$\left(3\mu_Z - 4\beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1} \left(3\mu_Z - 4\beta^T \gamma_0\right) \leq \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z \quad (3.20)$$

then the value of ζ *that maximizes* $S^2\left(Z \ominus w\left(c,\zeta\right)^T Z\right)$ *is* $\zeta = 0_{M\times 1}$*. In this case,* F *is a linear factor* *model for* $Z \ominus w(c,0)^T Z$*, and:*

$$S^2\left(Z \ominus w(c,0)^T Z\right) = \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \frac{\left[\mu_Z^T \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)\right]^2}{\left(\mu_Z - \beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)} \tag{3.21}$$

This region is the closed interior of the smaller ellipsoid referenced in the previous case. Since the ellipsoid may be degenerate, this region may consist of a single point or the empty set.

Proof: See Appendix.

Theorem 10 identifies five distinct regions a given vector $\gamma_0 \in \mathbb{R}^N$ might occupy; Figure 3 shows these regions graphically. In the first region, occupying a closed half-space, the vector of risk premia can be achieved with the smallest possible reduction in squared Sharpe ratio by the construction of Theorem 5, which makes the factors a linear factor model for the remaining assets. Removal of any other asset within the space identified by Theorem 9 results in larger reduction in squared Sharpe ratio, and also fails to make the factors a linear factor model for the remaining factors; relaxing this requirement does not make it easier, in terms of squared Sharpe ratio reduction, to achieve a given risk premia vector in this region. The fifth region has the same properties as the first, except that the fifth region may consist of a single point, or may not exist at all. In the second and fourth regions, consisting of the region between two hyperplanes and the region between two ellipsoids, respectively, there exist assets (characterized by a common value of $\zeta^T \Sigma_{ZZ} \zeta$) whose removal results in the given risk premia vector with a smaller reduction in squared Sharpe ratio than is possible by the construction of Theorem 5. The second region is of particular note, since any risk premia vector within this region can be achieved with no reduction in squared Sharpe ratio at all. In the third region, the given risk premia vector can be achieved with reduction in squared Sharpe ratio arbitrarily close to the limiting value by choosing $\zeta^T \Sigma_{ZZ} \zeta$ to be arbitrarily large.

The results of this section demonstrate that omission of an asset, even if it does not decrease the squared Sharpe ratio offered to investors at all, can have dramatic effects on the risk premia vector assigned to a set of factors. Omission of a single asset can make the set of factors a linear factor model for the remaining assets with an arbitrarily specified vector of risk premia, and for many values of that vector, the reduction in Sharpe ratio caused by the asset omission can be small or even zero. Many values of the risk premia vector can be achieved with an even smaller reduction in Sharpe ratio, if the factors are not required to be a linear factor model for the remaining assets. By contrast, as per the results of the previous section, the risk premium assigned to a spanned factor is not sensitive at all to asset omission, provided the asset does not cease to be spanned as a result of the omission.

4 Shadow Returns

One might be interested in unspanned factors because markets are incomplete, or because it is impossible or impractical to identify enough assets to span a particular set of factors. Focusing on the latter case, we consider the "unspanned" components of a set of factors to be in fact spanned, but by assets not observed

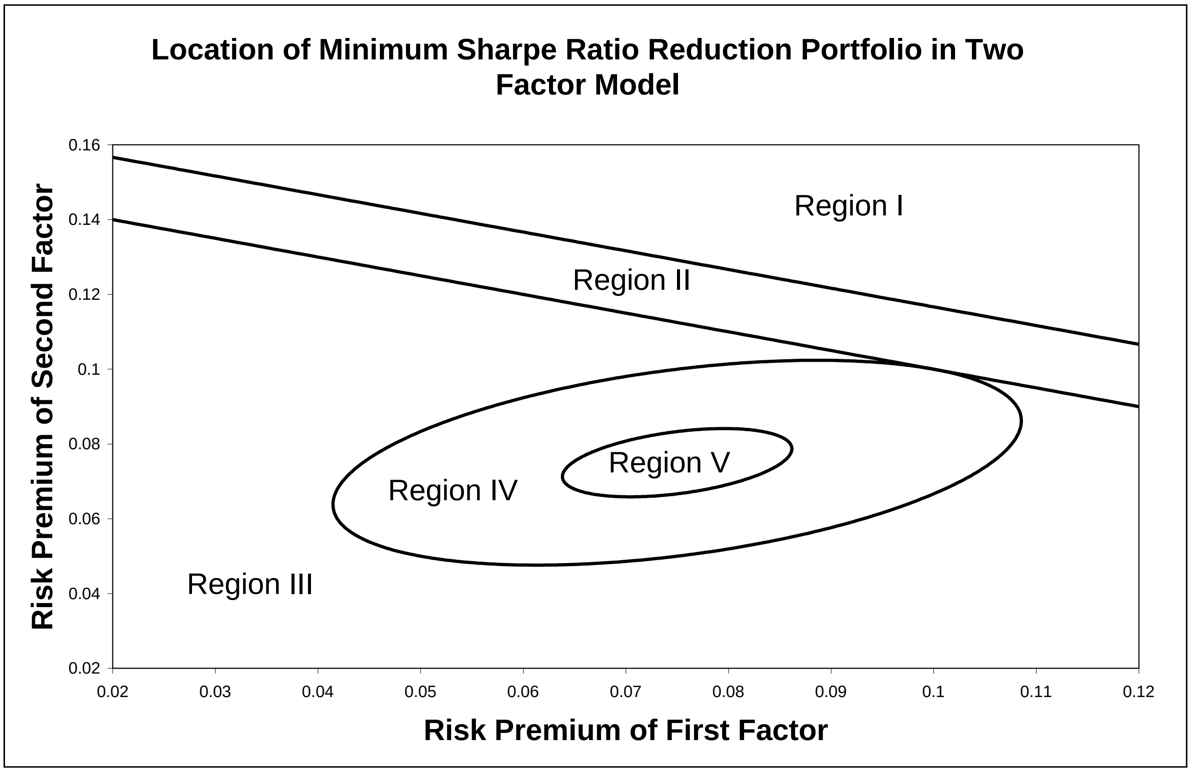

Figure 3: This figure identifies the five regions characterized by Theorem 10, for the same factors and assets used in Figure 2. Risk premia vectors lying within Regions I and V can be achieved with the smallest possible reduction of squared Sharpe ratio in the construction of Theorem 5, in which case the factors form a linear factor model for the remaining assets. Within Regions II and IV, a construction other than that of Theorem 5 results in the minimum reduction in squared Sharpe ratio; the factors are in this case not a linear factor model for the remaining asset. Within Region III, there is no minimum reduction in squared Sharpe ratio, but rather a limit which can be approached arbitrarily closely. As this limit is approached, the factors are not a linear factor model for the remaining assets. For some sets of factors and assets, Region V may consist of a single point, or vanish entirely. Similarly, Region IV may vanish entirely.

by the econometrician. For a set of factors F and a set of assets Z that satisfy the full-rank assumptions, let $K \leq N$ be the rank of $\Sigma_{\eta\eta}$. We assume the existence of a set of asset payoffs Y, with finite and full-rank covariance matrix Σ_{YY} and mean μ_Y, such that:

$$\left(\eta - \mu_\eta\right) = \Psi^T \left(Y - \mu_Y\right) \tag{4.1}$$

for some $K \times N$ matrix Ψ with rank K. It follows that:

$$\begin{aligned} \Sigma_{\eta\eta} &= \Psi^T \Sigma_{YY} \Psi & (4.2) \\ \Sigma_{YF} &= \Sigma_{Y\eta} = \Sigma_{YY}\Psi & (4.3) \end{aligned}$$

Y and Z are uncorrelated (since η and Z are uncorrelated), so F and the union of Z and Y satisfy the full-rank assumptions. Since the factors F are spanned by Z and Y, we have:

$$\gamma_0 = \gamma\left(F, \begin{bmatrix} Z \\ Y \end{bmatrix}\right) = \begin{bmatrix} \Gamma \\ \Psi \end{bmatrix}^T \begin{bmatrix} \mu_Z \\ \mu_Y \end{bmatrix} = \Gamma^T \mu_Z + \Psi^T \mu_Y \tag{4.4}$$

This value is the solution of the minimization problem:

$$\gamma_0 = \underset{\gamma \in \mathbb{R}^N}{\operatorname{argmin}} \left[\left(\mu_Z - \beta_Z^T \gamma\right)^T \Sigma_{ZZ}^{-1} \left(\mu_Z - \beta_Z^T \gamma\right) + \left(\mu_Y - \beta_Y^T \gamma\right)^T \Sigma_{YY}^{-1} \left(\mu_Y - \beta_Y^T \gamma\right) \right] \tag{4.5}$$

where $\beta_Z = \Sigma_{FF}^{-1}\Sigma_{FZ}$ and $\beta_Y = \Sigma_{FF}^{-1}\Sigma_{FY}$. However, since the expected excess returns of the assets Y are not observed, we can only calculate the vector of risk premia for F using the assets Z, as per Equation 2.15. If we set the two risk premium definitions (one requiring knowledge of μ_Y, and therefore infeasible) equal to each other, we find:

$$\begin{aligned} \gamma\left(F, \begin{bmatrix} Z \\ Y \end{bmatrix}\right) &= \gamma\left(F, Z\right) & (4.6) \\ \Gamma^T \mu_Z + \Psi^T \mu_Y &= \Gamma^T \mu_Z + \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Gamma^T \mu_Z & (4.7) \\ \mu_Y &= \left(\Psi\Psi^T\right)^{-1} \Psi \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Gamma^T \mu_Z = \Sigma_{YY}\Psi\Sigma_{PP}^{-1}\Gamma^T \mu_Z & (4.8) \end{aligned}$$

We can therefore interpret the assignment of risk premia to F when μ_Z is observed (but μ_Y is not) as assigning shadow expected excess returns to the unspanned components of F (i.e., to Y). From Equation 4.8, these shadow values are unique, whether or not F is a linear factor model for Z. If we attempt to predict the expected excess returns of the assets Y using the risk premia vector $\gamma\left(F, Z\right)$, we find:

$$\beta_Y^T \gamma\left(F, Z\right) = \Sigma_{YF}\Sigma_{FF}^{-1}\gamma\left(F, Z\right) = \Sigma_{YY}\Psi\Sigma_{PP}^{-1}\Gamma^T \mu_Z \tag{4.9}$$

Note that the expected excess returns predicted by this equation are the same as the shadow prices assigned by Equation 4.8, whether F is a linear factor model for Z or not. If F is a linear factor model for Z, and the actual expected excess returns of Y are equal to the values predicted by Equation 4.8, then F is a linear factor model for the union of Z and Y (F may fail to be a linear factor model for Y, since there is no guarantee that

F and Y satisfy the full rank assumptions).

Of course, if the actual value of μ_Y is something different than the shadow value specified in Equation 4.8, then the unobservability of Y results in a different (and incorrect) assignment of a vector of risk premia to F. Since the econometrician who does not observe the assets Y has no information concerning the true value of μ_Y, other than perhaps a theoretically motivated belief that a particular set of factors constitutes a linear factor model, it is not clear that the shadow values assigned are meaningful. One possible way to characterize the shadow expected excess returns is in terms of the mean-variance efficient portfolios formed from P and Y. By premultiplication of both sides of Equation 4.8 by $(Y - \mu_Y)^T \Sigma_{YY}^{-1}$, we find:

$$(Y - \mu_Y)^T \Sigma_{YY}^{-1} \mu_Y = (Y - \mu_Y)^T \Psi \Sigma_{PP}^{-1} \Gamma^T \mu_Z = \left(\eta - \mu_\eta\right)^T \Sigma_{PP}^{-1} \Gamma^T \mu_Z \tag{4.10}$$

If the shadow value of μ_Y is in fact the true value, then $Y^T \Sigma_{YY}^{-1} \mu_Y$ is the mean-variance efficient portfolio formed from the assets Y, and $P^T \Sigma_{PP}^{-1} \Gamma^T \mu_Z$ is the mean-variance efficient portfolio formed from the assets P. In other words, the linear combination of factors $F^T \Sigma_{PP}^{-1} \Gamma^T \mu_Z$ has a projection onto Z of $P^T \Sigma_{PP}^{-1} \Gamma^T \mu_Z$, and a projection onto Y of $\left(\eta - \mu_\eta\right)^T \Sigma_{PP}^{-1} \Gamma^T \mu_Z$. If the shadow value of μ_Y is equal to the true (but unobserved) expected excess returns of Y, then both of these projections are mean-variance efficient within the spaces spanned by Z and Y, respectively. The risk premia assigned to F therefore implies shadow values of μ_Y that ensure the same linear combination of factors with a mean-variance efficient projection onto P, also has a mean-variance efficient projection onto Y.

The shadow values of μ_Y, of course, do not necessarily bear any relation to the true values. Consider a set of assets Z and a set of spanned factors $P = \Gamma^T Z$, and a set of assets Y, uncorrelated with P. It is possible to construct a set of N factors F with projection P onto Z, such that the shadow value of μ_Y is equal to any arbitrarily chosen non-zero value; if the rank of Ψ is less than N, then any arbitrarily chosen value of μ_Y, including zero, is achievable.

Theorem 11. *Let Z be a set of M assets, and let $P = \Gamma^T Z$ be a set of N factors spanned by Z, such that Z and P satisfy the full rank assumptions. Let Y be a set of $K \leq N$ assets with a finite and full rank covariance matrix Σ_{YY}, uncorrelated with Z. Let μ_0 be any K-vector. If $K < N$ or $\mu_0 \neq 0$, then there exists a set of N factors F with projection P onto Z, such that the shadow value of μ_Y from Equation 4.8 is equal to μ_0.*

Proof: See Appendix.

If the allegedly unspanned components of a set of factors are in fact spanned, then, as per the results of the previous sections, the risk premia vector (under the assumption that all assets are observed) is unique, well-defined, and invariant to the addition or removal of assets or other factors, so long as the full-rank assumptions are satisfied (and as long as the removal of an asset does not cause a previously spanned factor to become unspanned). By contrast, the risk premia definition of Equation 2.15 implicitly assigns to the "unspanned" components a set of shadow expected excess returns that results in the closest fit of the assets Z to the factors F, without regard to the true value of μ_Y (which is unobserved). Theorem 11 shows how arbitrary this assignment can be; essentially any shadow value can be assigned simply by repackaging the spanned and unspanned components of the factors. Use of "unspanned" factors, which are in fact spanned by other assets not included in a study, therefore tends to result in overfitting; unless the shadow value of μ_Y is exactly equal to the true value, the risk premia assigned to a set of factors F by Equation 2.15 predicts expected excess returns for the assets Z that are a closer fit to the observed expected returns than those predicted by the true risk premia.

Empirical studies generally do not distinguish between the two types of risk premia, i.e., that which is due to μ_Z and that which is due to μ_Y. Since the assignment of risk premia to μ_Y is very sensitive to misspecification, recognition of such a distinction would be useful; reporting the risk premia of factor projections, in addition to or instead of the risk premia of the factors themselves, would be a useful guide to the reliability of the risk premia. The risk premia of the projected factors, as already noted, are given by:

$$\gamma\left(P, Z\right) = \Gamma^T \mu_Z \tag{4.11}$$

Considering the factors collectively, the maximum squared Sharpe ratio offered by the factors (treating them as if they were traded assets) is given by:

$$\begin{aligned}
\gamma^T\left(F, Z\right)\Sigma_{FF}^{-1}\gamma\left(F, Z\right) &= \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z & (4.12)\\
&= \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z + \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Gamma^T\mu_Z & (4.13)
\end{aligned}$$

By contrast, the maximum squared Sharpe ratio offered by the factor projections (treating them as if they were traded assets) is given by:

$$\gamma^T\left(P, Z\right)\Sigma_{PP}^{-1}\gamma\left(P, Z\right) = \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z \tag{4.14}$$

It is a fairly common practice to examine the difference between this latter quantity and the maximum squared Sharpe ratio offered by the assets (given by $\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z$); typical GMM formulations and the method of Gibbons, Ross, and Shanken (1989) test the hypothesis that this difference is zero (in which case the factors are a linear factor model). However, the difference between the quantities in Equations 4.13 and 4.14 is rarely, if ever, examined; nor is the difference between $\gamma\left(F, Z\right)$ and $\gamma\left(P, Z\right)$. Although there would appear to be little point in performing a statistical test to see if the two are different, the difference provides insight into how much of the risk premia assigned to a set of factors is directly observed in the asset returns, and how much is the result of extrapolation (which relies on model correctness). The risk premia $\gamma\left(P, Z\right)$ may be viewed as the product of a policy of conservatism, reflecting only risk premium for which the asset returns provide direct evidence. The risk premium of $\gamma\left(F, Z\right)$, by contrast, may be viewed as the result of a policy of optimism, in which possibly large risk premia, for which there are no direct evidence, are attributed to unobserved assets in an effort to fit the data better. Should the model turn out to be misspecified, much of the risk premia assigned to the factors by $\gamma\left(F, Z\right)$ may turn out to be fictional. By contrast, the risk premia assigned to the projection factors $\gamma\left(P, Z\right)$ are robust to misspecification.

The next section considers the problem of estimating risk premia of various types of factors, including the projections of unspanned factors.

5 Estimation

We now consider the problem of estimation of risk premia for a set of factors from a time series of observations of the factor realizations and asset payoffs. Such a problem is considered by Black, Jensen, and Scholes (1972) and by Fama and MacBeth (1973), who employ a two-pass regression method; Shanken (1992) derives the asymptotic distribution of the vector of risk premia estimated by such a method, but under an assumption that the factors constitute a linear factor model for the assets, and Kim (1995) modifies this analysis to account for heteroskedasticity. However, the combination of unspanned factors and possible misspecification adds some additional complexity to the analysis.

We take the estimation approach of replacing mean vectors and covariance matrices in the definition of the risk premia vector by their sample counterparts. This approach is straightforward, and also produces the same estimates as a two-pass regression technique, provided an appropriate GLS weighting matrix is used. The β coefficients for a set of assets can be estimated in a first pass regression of the time series of Z onto F:

$$\widehat{\beta} = \widehat{\Sigma}_{FF}^{-1} \widehat{\Sigma}_{FZ} \tag{5.1}$$

The risk premia estimates can then be estimated in a second pass regression of $\widehat{\mu}_Z$ on $\widehat{\beta}$:

$$\widehat{\gamma}(F, Z) = \left(\widehat{\beta} \widehat{\Sigma}_{ZZ}^{-1} \widehat{\beta}^T \right)^{-1} \left(\widehat{\beta} \widehat{\Sigma}_{ZZ}^{-1} \widehat{\mu}_Z \right) \tag{5.2}$$

where the GLS weighting matrix $\widehat{\Sigma}_{ZZ}^{-1}$ is used. If the joint distribution of F and Z is such that the sample estimates of Σ_{FF}, Σ_{FZ}, Σ_{ZZ}, and μ_Z are consistent, then:

$$\begin{aligned} \operatorname{plim} \widehat{\gamma}(F, Z) &= \left(\beta \Sigma_{ZZ}^{-1} \beta^T \right)^{-1} \left(\beta \Sigma_{ZZ}^{-1} \mu_Z \right) && (5.3) \\ &= \left(\Sigma_{FF}^{-1} \Sigma_{PP} \Sigma_{FF}^{-1} \right)^{-1} \left(\Sigma_{FF}^{-1} \Sigma_{FZ} \Sigma_{ZZ}^{-1} \mu_Z \right) && (5.4) \\ &= \Sigma_{FF} \Sigma_{PP}^{-1} \Gamma^T \mu_Z = \gamma(F, Z) && (5.5) \end{aligned}$$

This method of estimation is therefore consistent (provided, as stated above, the estimates of Σ_{FF}, Σ_{FZ}, Σ_{ZZ}, and μ_Z are consistent). If F is a linear factor model for Z, then it is consistent with other choices of weighting matrix as well, although in this case the results of the second regression are subject to the small sample problems discussed in Kandel and Stambaugh (1995). Furthermore, we note that no constant is included in the second regression, since we are dealing with excess returns rather than returns. If a constant is included in this regression, the results are subject to the Kandel and Stambaugh (1995) small sample problem even if the GLS weighting matrix Σ_{ZZ}^{-1} is used. In the case of spanned factors, although the two-pass regression estimator with GLS weighing matrix of $\widehat{\Sigma}_{ZZ}^{-1}$ is consistent, is is unnecessarily complicated; in this case, the risk premia are $\gamma = \Gamma^T \mu_Z$, which is more simply estimated by regressing the time series of factor realizations on the asset returns, and then multiplying the estimated coefficients by an estimate of μ_Z. We simply take the approach of replacing the matrices Σ_{FF}, Σ_{FZ}, Σ_{ZZ}, and μ_Z in the definition of the risk premia vector by their sample counterparts, since both of the two regression schemes described above yield the same result anyway.

First, we consider the case of spanned factors. The vector of risk premia is then:

$$\gamma(F, Z) = \Gamma^T \mu_Z = \Sigma_{FZ} \Sigma_{ZZ}^{-1} \mu_Z \tag{5.6}$$

Since the factors are spanned, the matrix Γ can be estimated without error; the coefficients of Γ are the coefficients of a regression of F onto Z, as per Equation 2.7, but in which the variance of each element of η is zero. The only variation in estimation is therefore due to variation in the estimation of μ_Z, which can be estimated by its sample mean:

$$\widehat{\gamma}(F, Z) = \widehat{\Sigma}_{FZ} \widehat{\Sigma}_{ZZ}^{-1} \widehat{\mu}_Z = \widehat{\Gamma}^T \widehat{\mu}_Z = \Gamma^T \widehat{\mu}_Z \tag{5.7}$$

Under an assumption that the observations of Z are i.i.d. distributed, the variance of this estimator has a particularly simple form.

Theorem 12. *Let Z be a set of assets and let $P = \Gamma^T Z$ be a set of factors such that P and Z satisfy the full-rank assumptions. Let $\widehat{\gamma}(P, Z)$ be the estimate of $\gamma(P, Z)$, as per Equation 5.7. If the time series of observations of F and Z is i.i.d., then $\widehat{\gamma}(P, Z)$ is a consistent estimate of $\gamma(P, Z)$, with asymptotic variance given by:*

$$AVar\left[\widehat{\gamma}(P, Z)\right] = \Sigma_{PP} \tag{5.8}$$

Proof: See Appendix.

The asymptotic variance of an individual factor does not depend on the other factors included in the factor set; furthermore, as long as the factors are spanned, the asymptotic variance of the risk premia vector is completely invariant to the choice of the spanning assets Z. This result can be viewed as a multivariate generalization of the result of Kandel and Stambaugh (1995), who consider a single spanned factor and expected returns (rather than expected excess returns) and find that the risk premium estimate differs from the zero-beta rate by the expected return of the factor's spanning portfolio when the GLS weighting matrix is used. Since we consider expected excess returns, the zero-beta rate is equal to zero, and the risk premia for factors are thus the expected excess return of the factors' spanning portfolios. No other characteristic of the assets or factors is relevant for the determination of the factor risk premia.

Before considering the fully general case of factors with unspanned components, we examine estimation of the risk premia of the projection P of a set of (not necessarily spanned) factors F onto the assets Z. If F and Z satisfy the full-rank assumptions, then so do P and Z, and if F is a linear factor model for Z, then so is P. However, in this case, the coefficients Γ are estimated with error. The risk premia of the projections of the factors are given by:

$$\gamma(P, Z) = \Gamma^T \mu_Z = \Sigma_{FZ} \Sigma_{ZZ}^{-1} \mu_Z \tag{5.9}$$

This value can be estimated by:

$$\widehat{\gamma}(P, Z) = \widehat{\Gamma}^T \widehat{\mu}_Z = \widehat{\Sigma}_{FZ} \widehat{\Sigma}_{ZZ}^{-1} \widehat{\mu}_Z \tag{5.10}$$

The uncertainty in the estimation of Γ leads to a larger asymptotic variance than when P is observed directly.

Theorem 13. *Let F be a set of factors and Z a set of assets, such that F and Z satisfy the full-rank assumptions. Let P be the projection of F onto Z. Let $\widehat{\gamma}(P,Z)$ be the estimate of $\gamma(P,Z)$, as per Equation 5.10. If the time series of observations of P and Z is i.i.d. with a multivariate Gaussian distribution, then $\widehat{\gamma}(P,Z)$ is a consistent estimate of $\gamma(P,Z)$, with asymptotic variance given by:*

$$AVar\left[\widehat{\gamma}(P,Z)\right] = \left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right) \Sigma_{\eta\eta} + \Sigma_{PP} \tag{5.11}$$

Proof: See Appendix.

Of course, Theorem 12 may be viewed as a special case of Theorem 13. The presence of unspanned components necessarily increases the asymptotic variance of the risk premia estimates, unless the expected excess return of every asset is equal to zero. Note that the asymptotic variance of the risk premia vector depends only on the covariance matrices of the spanned and unspanned components, and on the maximum squared Sharpe ratio offered by the assets Z. As in the case of Theorem 12, the asymptotic variance of the risk premium of an individual factor does not depend on the other factors included in the factor set. In particular, if an individual factor has no unspanned component, the estimate is not affected by the presence of unspanned components in other factors included in the factor set.

In the fully general case of unspanned factors, the risk premia vector is given by:

$$\gamma(F,Z) = \Sigma_{FF}\Sigma_{PP}^{-1}\gamma(P,Z) = \Sigma_{FF}\left(\Sigma_{FZ}\Sigma_{ZZ}^{-1}\Sigma_{ZF}\right)^{-1}\Sigma_{FZ}\Sigma_{ZZ}^{-1}\mu_Z \tag{5.12}$$

In addition to the variation due to the estimation of $\gamma(P,Z)$, the result is premultiplied by $\Sigma_{FF}\Sigma_{PP}^{-1}$, which must also be estimated. The estimate is therefore given by:

$$\widehat{\gamma}(F,Z) = \widehat{\Sigma}_{FF}\left(\widehat{\Sigma}_{FZ}\widehat{\Sigma}_{ZZ}^{-1}\widehat{\Sigma}_{ZF}\right)^{-1}\widehat{\Sigma}_{FZ}\widehat{\Sigma}_{ZZ}^{-1}\widehat{\mu}_Z \tag{5.13}$$

If F is a linear factor model for Z, then the asymptotic variance of $\widehat{\gamma}(F,Z)$ takes on a particularly simple form.

Theorem 14. *Let F be a linear factor model for Z, and let $\widehat{\gamma}(F,Z)$ be the estimate of $\gamma(F,Z)$, as per Equation 5.13. If F and Z are i.i.d. with a multivariate Gaussian distribution, then $\widehat{\gamma}(F,Z)$ is a consistent estimate of $\gamma(F,Z)$, with asymptotic variance given by:*

$$AVar\left[\widehat{\gamma}(F,Z)\right] = \Sigma_{FF} + \left[1 + \gamma^T(F,Z)\Sigma_{FF}^{-1}\gamma(F,Z)\right]\left(\Sigma_{\eta\eta} + \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right) \tag{5.14}$$

The asymptotic variance can also be expressed in terms of μ_Z instead of $\gamma(F,Z)$:

$$AVar\left[\widehat{\gamma}(F,Z)\right] = \Sigma_{FF} + \left[1 + \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z + \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right]\left(\Sigma_{\eta\eta} + \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right) \tag{5.15}$$

Proof: See Appendix.

A similar expression appears in Shanken (1992), who considers the estimation of risk premia by two-pass regression when dealing with expected returns (as opposed to expected excess returns). The incremental variance incurred by choosing F rather than P as the set of factors is given by:

$$\begin{aligned} AVar\left[\widehat{\gamma}\left(F,Z\right)\right]-AVar\left[\widehat{\gamma}\left(P,Z\right)\right] &= \left[1+\gamma^{T}\left(F,Z\right)\Sigma_{FF}^{-1}\gamma\left(F,Z\right)\right]\left(\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right) \\ &\quad +\left[2+\gamma^{T}\left(F,Z\right)\Sigma_{FF}^{-1}\Sigma_{\eta\eta}\Sigma_{FF}^{-1}\gamma\left(F,Z\right)\right]\Sigma_{\eta\eta} \end{aligned} \tag{5.16}$$

or, expressed in terms of μ_Z instead of $\gamma\left(F,Z\right)$:

$$\begin{aligned} AVar\left[\widehat{\gamma}\left(F,Z\right)\right]-AVar\left[\widehat{\gamma}\left(P,Z\right)\right] &= \begin{pmatrix} 1+\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z \\ +\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Gamma^T\mu_Z \end{pmatrix}\left(\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right) \\ &\quad +\left(2+\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right)\Sigma_{\eta\eta} \end{aligned} \tag{5.17}$$

The difference is clearly positive as long as $\Sigma_{\eta\eta}$ is not equal to zero. The presence of unspanned components causes $\widehat{\gamma}\left(F,Z\right)$ to have larger variance than $\widehat{\gamma}\left(P,Z\right)$, even if $\gamma\left(F,Z\right)=\gamma\left(P,Z\right)$. Although the variance of the risk premium estimate for an individual spanned factor is unaffected by the inclusion of unspanned components in other factors, it should also be noted that the asymptotic variance of the risk premium estimate of an individual factor with an unspanned component is, in contrast with the two previous cases, dependent on the other factors included in the factor set.

When F is not a linear factor model for Z, there are additional sources of variance.

Theorem 15. *Let F be a set of factors and Z be a set of assets such that F and Z satisfy the full-rank assumptions, and let $\widehat{\gamma}\left(F,Z\right)$ be the estimate of $\gamma\left(F,Z\right)$, as per Equation 5.13. If F and Z are i.i.d. with a multivariate Gaussian distribution, then $\widehat{\gamma}\left(F,Z\right)$ is a consistent estimate of $\gamma\left(F,Z\right)$, with asymptotic variance given by:*

$$AVar\left[\widehat{\gamma}\left(F,Z\right)\right] = \Sigma_{FF}+\left[\Sigma_{\eta\eta}+\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right]\left(1+\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z+\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right) \tag{5.18}$$

$$+\left[\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}+\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right]\left(\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z-\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right) \tag{5.19}$$

Proof: See Appendix.

To distinguish between the two versions of the asymptotic variance of $\widehat{\gamma}\left(F,Z\right)$ (derived under differing assumptions about the asset expected excess returns), we denote the asymptotic variance from Theorem 14 by $AVar_{LF}$, since it is derived under the assumption that F is a linear factor model for Z. The difference between the two asymptotic variances (from Theorems 15 and 14) is then given by:

$$AVar\left[\widehat{\gamma}\left(F,Z\right)\right]-AVar_{LF}\left[\widehat{\gamma}\left(F,Z\right)\right]=\begin{bmatrix}\Sigma_{\eta\eta}+ \\ 2\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}+ \\ \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\end{bmatrix}\left[\varepsilon^T\left(F,Z\right)\Sigma_{ZZ}^{-1}\varepsilon\left(F,Z\right)\right] \tag{5.20}$$

The right-hand side is always positive whenever $\Sigma_{\eta\eta}\neq 0$ and $\varepsilon\left(F,Z\right)\neq 0$. Holding the covariance matrices and $\gamma\left(F,Z\right)$ fixed, Equation 5.20 indicates the degree to which misspecification (i.e., that F is not a linear factor model for Z) affects the variance of the estimate of $\gamma\left(F,Z\right)$. The greater the misspecification (i.e., the larger the magnitude of $\varepsilon^T\left(F,Z\right)\Sigma_{ZZ}^{-1}\varepsilon\left(F,Z\right)$), the larger is the asymptotic variance of the risk premia vector.

It is worth noting a few points on hypothesis testing. A common practice is to test whether a given factor is "priced," in the sense that it has a risk premium different from zero. When risk premia are estimated using a two-pass regression technique, the t-statistic of a risk premium estimate is often used to determine whether the corresponding factor is priced; Chen, Kan, and Zhang (1999) offer a criticism of this approach based on the statistical properties of the risk premia estimates. However, quite apart from any econometric issues, this approach is conceptually flawed. If a factor is spanned, its risk premium is equal to the expected excess return of the spanning portfolio. This result does not depend on what other factors are included in a factor set, whether those factors constitute a linear factor model for the assets, etc. For example, given a set of $M > 1$ assets Z, consider the factors F defined as:

$$F^T = \mu_F^T + (Z - \mu_Z)^T \Sigma_{ZZ}^{-1} \begin{bmatrix} \mu_Z & w \end{bmatrix} \tag{5.21}$$

where w is an M-vector with $w \neq \mu_Z k$ for any constant k. It is a brief exercise to verify that F is indeed a linear factor model for Z, and that the risk premia of the two factors are $\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z$ and $\mu_Z^T \Sigma_{ZZ}^{-1} w$, respectively. However, one quickly verifies that the first factor is by itself a linear factor model for Z, with the same risk premium. If $\mu_Z^T \Sigma_{ZZ}^{-1} w \neq 0$, then the second factor has a non-zero risk premium despite its non-importance, and a consistent estimation technique allows an econometrician, given enough data, to reject the hypothesis that the risk premium of this second factor is zero. Despite this finding that the factor is "priced," it has absolutely no relevance, given the other factors in the model, for asset pricing.

Similarly, let w be an M-vector such that $\mu_Z^T \Sigma_{ZZ}^{-1} w = 0$ and $w \neq \mu_Z k$ for any value of k. Then define the factors:

$$F^T = (Z - \mu_Z)^T \Sigma_{ZZ}^{-1} \begin{bmatrix} \mu_Z - w & w \end{bmatrix} \tag{5.22}$$

One also readily verifies that this set of factors is a linear factor model for Z. However, if the second factor, which by definition has a risk premium of zero, is removed, the remaining factor is not a linear factor model. One can in fact make the fit of the first factor arbitrarily bad (i.e., the value of $\varepsilon^T (F, Z) \Sigma_{ZZ}^{-1} \varepsilon (F, Z)$ is arbitrarily large) simply by choosing w such that $w^T \Sigma_{ZZ}^{-1} w$ is large enough. For sufficiently large w, the first factor does a very poor job predicting the expected excess returns of the assets, but when the second factor (with its risk premium of zero) is added, the model predicts the expected excess returns of all assets perfectly. However, conventional statistical tests will fail to reject the hypothesis that the risk premium of the second factor is zero (since the hypothesis is true) 95% of the time. Despite its importance in asset pricing, the second factor is not "priced."

Whether the risk premium of a spanned factor is equal to zero or not has essentially nothing to do with its importance in a model for expected excess returns. As per the results of earlier sections, holding a set of assets fixed, the risk premium of an unspanned factor is not even uniquely defined, unless the factor is a linear factor model by itself. However, even if the risk premium of a particular factor is "correct," in the sense that it is spanned by additional assets (not included in the study) with actual expected excess returns equal to the shadow values, the use of t-statistics of risk premia, or similar techniques, in model selection is fundamentally flawed, because it tests the wrong hypothesis. An appropriate hypothesis to test is not whether the risk premium of a factor is equal to zero, but whether the removal of the factor from the model results in a statistically significant change in the fitted values of the expected excess returns of the assets.

6 Conclusion

We have examined the problem of definition and estimation of risk premia for linear factor models in the presence of misspecification. The results show that factors that are spanned by the assets under study can be assigned risk premia that are invariant to the choice of other factors included in the model, and also to the choice of assets under study; furthermore, model misspecification does not increase the asymptotic variance of estimates of such risk premia. By contrast, it is impossible even to define unambiguously the risk premium of an unspanned factor in the presence of model misspecification; such a risk premium is dependent on the other factors included in a model, and also on the assets under study. Furthermore, even holding the set of factors fixed, misspecification leads to larger asymptotic variance for estimates of the risk premia of unspanned factors. The risk premium of a spanned factor can be considered an intrinsic property of the factor itself, as it is invariant to the model in which the factor is placed and the assets used to span it. By contrast, the risk premium of an unspanned factor is only meaningful in the context of the other factors in the model and the assets included in a study.

If it is not possible or practical to extend the set of assets included in a study so that all factors are spanned, then a researcher may nonetheless avoid the hard consequences of potential misspecification on factor risk premia by focusing instead on the risk premia of the projections of the factors onto the asset space. These projections, being spanned factors themselves, are completely immune to the troubles associated with unspanned factors, although, since the projections onto the asset space are estimated with error, the asymptotic variance of the risk premia are larger than they would be if the projections were estimated without error. However, the asymptotic variances of the risk premia estimates of the projections are still smaller than the asymptotic variances of the risk premia estimates for the factors themselves. If a model is possibly misspecified, distinguishing between the risk premia of the factors and their projections provides an indication of how much of a factor risk premium is due to extrapolation, which is sensitive to misspecification, rather than direct observation, which is not. The risk premia of the projections are completely robust to misspecification, and may be viewed as the result of a policy of conservatism, reflecting only risk premium for which there is direct evidence from asset returns. By contrast, the risk premia of unspanned factors are highly sensitive to misspecification, and may be viewed as the result of a policy of optimism, reflecting extrapolated components of risk premium for which there is no direct evidence in asset return data. The robustness of an empirical study can thus be improved by reporting the risk premia of factor projections, in addition to or instead of the risk premia of unspanned factors.

References

AMIHUD, Y., B. J. CHRISTENSEN, AND H. MENDELSON (1992): "Further Evidence on the Risk-Return Relationship," Working paper, New York University and Stanford University.

BLACK, F., M. JENSEN, AND M. SCHOLES (1972): "The Capital Asset Pricing Model: Some Empirical Results," in *Studies in the Theory of Capital Markets*, ed. by M. Jensen, pp. 79–121. Praeger.

BREEDEN, D. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–296.

CHEN, N.-F., R. KAN, AND C. ZHANG (1999): "A Critique of the Use of T-Ratios in Model Selection," Working paper, Hong Kong University of Science and Technology and University of Toronto.

DANIEL, K., AND S. TITMAN (1997): "Evidence on the Characteristics of Cross-Sectional Variation in Stock Returns," *Journal of Finance*, 52, 1–33.

DANIEL, K., S. TITMAN, AND K. J. WEI (2001): "Explaining the Cross-Section of Stock Returns in Japan: Factors or Characteristics?," *Journal of Finance*, 55, 743–766.

FAMA, E. F., AND K. R. FRENCH (1992): "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 427–465.

——— (1996): "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 51, 55–84.

FAMA, E. F., AND J. D. MACBETH (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

GIBBONS, M. R., S. A. ROSS, AND J. SHANKEN (1989): "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121–1152.

HANSEN, L. P., AND S. F. RICHARD (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–613.

JAGANNATHAN, R., AND Z. WANG (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3–53.

KAN, R., AND C. ZHANG (1999a): "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, 54, 103–127.

KAN, R., AND C. ZHANG (1999b): "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 203–235.

KAN, R., AND G. ZHOU (1999): "A Critique of the Stochastic Discount Factor Methodolgy," *Journal of Finance*, 54, 1221–1248.

KANDEL, S., AND R. F. STAMBAUGH (1995): "Portfolio Inefficiency and the Cross-Section of Expected Returns," *Journal of Finance*, 50, 157–184.

KIM, D. (1995): "The Errors in the Variables Problem in the Cross-Section of Expected Stock Returns," *Journal of Finance*, 50, 1605–1634.

LINTNER, J. (1965): "The Valuation of Risky Assets and the Selection of Risky Investments in the Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13–37.

MERTON, R. C. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–887.

ROLL, R., AND S. A. ROSS (1994): "On the Cross-Sectional Relation Between Expected Returns and Betas," *Journal of Finance*, 49, 101–121.

Ross, S. A. (1976): “The Arbitrage Theory of Capital Asset Pricing,” *Journal of Economic Theory*, 13, 341–360.

Shanken, J. (1992): “On the Estimation of Beta-Pricing Models,” *Review of Financial Studies*, 5, 1–33.

Sharpe, W. F. (1964): “Capital Asset Prices: A Theory of Market Equilibrium,” *Journal of Finance*, 19, 425–442.

7 Appendix

7.1 Proof of Lemma 1

Since F is a linear factor model for Z, we have:

$$\mu_Z = \beta^T \gamma = \Sigma_{ZF} \Sigma_{FF}^{-1} \gamma \tag{7.1}$$

From the full-rank assumptions, the quantity $\Sigma_{FF} \left(\Sigma_{FZ} \Sigma_{ZZ}^{-1} \Sigma_{ZF} \right)^{-1} \Sigma_{FZ} \Sigma_{ZZ}^{-1}$ is well-defined and unique. Premultiplying both sides of Equation 7.1 by this quantity yields:

$$\Sigma_{FF} \left(\Sigma_{FZ} \Sigma_{ZZ}^{-1} \Sigma_{ZF} \right)^{-1} \Sigma_{FZ} \Sigma_{ZZ}^{-1} \mu_Z = \Sigma_{FF} \left(\Sigma_{FZ} \Sigma_{ZZ}^{-1} \Sigma_{ZF} \right)^{-1} \Sigma_{FZ} \Sigma_{ZZ}^{-1} \Sigma_{ZF} \Sigma_{FF}^{-1} \gamma = \gamma \tag{7.2}$$

The last expression is the desired result.QED

7.2 Proof of Theorem 1

We show that the other four quantities in the chain of equalities are all equal to $z^T \mu_Z$. Beginning with F, the risk premia vector is:

$$\begin{aligned} \gamma(F, Z) &= \Sigma_{FF} \left(\Sigma_{FZ} \Sigma_{ZZ}^{-1} \Sigma_{ZF} \right)^{-1} \Sigma_{FZ} \Sigma_{ZZ}^{-1} \mu_Z = \Sigma_{FF} \Sigma_{PP}^{-1} \Gamma_F^T \mu_Z & (7.3) \\ &= \left(\Sigma_{PP} + \Sigma_{\eta\eta} \right) \Sigma_{PP}^{-1} \Gamma_F^T \mu_Z = \Gamma_F^T \mu_Z + \Sigma_{\eta\eta} \Sigma_{PP}^{-1} \Gamma_F^T \mu_Z & (7.4) \end{aligned}$$

We premultiply both sides of the equation describing the projection of F onto Z by x^T:

$$x^T (F - \mu_F) = x^T \Gamma_F^T (Z - \mu_Z) + x^T \eta \tag{7.5}$$

Since η and Z are uncorrelated, and by assumption $x^T (F - \mu_F) = z^T (Z - \mu_Z)$, it must be the case that:

$$\begin{aligned} x^T \eta &= 0 & (7.6) \\ z^T &= x^T \Gamma_F^T & (7.7) \end{aligned}$$

From the last expression, we have $z^T \mu_Z = x^T \Gamma_F^T \mu_Z$. Premultiplying both sides of Equation 7.3 by x^T, we find:

$$x^T \gamma(F, Z) = x^T \Gamma_F^T \mu_Z + x^T \Sigma_{\eta\eta} \Sigma_{PP}^{-1} \Gamma_F^T \mu_Z = z^T \mu_Z \tag{7.8}$$

Proceeding analogously with G in place of F establishes the remaining identities.QED

7.3 Proof of Theorem 2

Consider the deviation of the vector $\gamma(F, Z)$ from $\gamma(P, Z)$:

$$\gamma(F, Z) - \gamma(P, Z) = \Sigma_{\eta\eta} \Sigma_{PP}^{-1} \gamma(P, Z) \tag{7.9}$$

If we premultiply both sides of Equation 7.9 by $\gamma(P, Z)^T \Sigma_{PP}^{-1}$, we find:

$$\gamma(P, Z)^T \Sigma_{PP}^{-1} \left[\gamma(F, Z) - \gamma(P, Z) \right] = \gamma(P, Z)^T \Sigma_{PP}^{-1} \Sigma_{\eta\eta} \Sigma_{PP}^{-1} \gamma(P, Z) \tag{7.10}$$

The right-hand side is clearly non-negative, since $\Sigma_{\eta\eta}$ is positive semidefinite. If the right-hand side is zero, then the right-hand side of Equation 7.9 is also zero, and $\gamma(F,Z) = \gamma(P,Z)$. If the right-hand side is positive, we have:

$$\begin{aligned} \gamma(P,Z)^T \Sigma_{PP}^{-1} [\gamma(F,Z) - \gamma(P,Z)] &> 0 && (7.11) \\ \gamma(P,Z)^T \Sigma_{PP}^{-1} \gamma(F,Z) &> \gamma(P,Z)^T \Sigma_{PP}^{-1} \gamma(P,Z) && (7.12) \end{aligned}$$

To show the other direction of the implication, we take as given a vector γ falling within the specified region (i.e., either $\gamma = \gamma(P,Z)$ or $\gamma(P,Z)^T \Sigma_{PP}^{-1} \gamma > \gamma(P,Z)^T \Sigma_{PP}^{-1} \gamma(P,Z)$), and construct a linear factor model F such that $\gamma(F,Z) = \gamma$. If $\gamma = \gamma(P,Z)$, then we can trivially take $F = P$. If $\gamma \neq \gamma(P,Z)$, then we must choose $\Sigma_{\eta\eta}$ such that:[4]

$$\gamma(F,Z) = \gamma(P,Z) + \Sigma_{\eta\eta} \Sigma_{PP}^{-1} \gamma(P,Z) = \gamma \tag{7.13}$$

If $N = 1$, we can trivially choose any random variable η such that:

$$Var[\eta] = \Sigma_{\eta\eta} = \frac{\Sigma_{PP}^{-1} \gamma(P,Z)}{\gamma - \gamma(P,Z)} \tag{7.14}$$

Since:

$$[\gamma - \gamma(P,Z)]^T \Sigma_{PP}^{-1} \gamma(P,Z) > 0 \tag{7.15}$$

it must be the case that either:

$$\gamma - \gamma(P,Z) > 0 \text{ and } \Sigma_{PP}^{-1} \gamma(P,Z) > 0 \tag{7.16}$$

or:

$$\gamma - \gamma(P,Z) < 0 \text{ and } \Sigma_{PP}^{-1} \gamma(P,Z) < 0 \tag{7.17}$$

In either case, the required value of $\Sigma_{\eta\eta}$ is positive.

For $N > 1$, rather than choosing $\Sigma_{\eta\eta}$ directly, we consider a non-singular transformation of this matrix. The required identity can be expressed as:

$$\gamma - \gamma(P,Z) = \Sigma_{\eta\eta} \Sigma_{PP}^{-1} \gamma(P,Z) = K \left[K^{-1} \Sigma_{\eta\eta} \left(K^T\right)^{-1} \right] K^T \Sigma_{PP}^{-1} \gamma(P,Z) \tag{7.18}$$

where K is an arbitrarily specified non-singular $N \times N$ matrix. We choose for K:

$$K = \left[\begin{array}{ccc} \frac{\Sigma_{PP}^{-1} \gamma(P,Z)}{\gamma(P,Z)^T \Sigma_{PP}^{-1} \Sigma_{PP}^{-1} \gamma(P,Z)} & \varepsilon & H \end{array} \right] \tag{7.19}$$

where ε is defined as:

$$\varepsilon = [\gamma - \gamma(P,Z)] - \Sigma_{PP}^{-1} \gamma(P,Z) \frac{\gamma(P,Z)^T \Sigma_{PP}^{-1} [\gamma - \gamma(P,Z)]}{\gamma(P,Z)^T \Sigma_{PP}^{-1} \Sigma_{PP}^{-1} \gamma(P,Z)} \tag{7.20}$$

and H is an $N \times (N-2)$ matrix orthogonal to the first two columns of K:

$$\begin{aligned} \gamma(P,Z)^T \Sigma_{PP}^{-1} H &= 0 && (7.21) \\ \varepsilon^T H &= 0 && (7.22) \end{aligned}$$

---

[4] We assume the probability space contains at least N random variables with finite variance that are uncorrelated with Z and each other.

(If $N = 2$, then H is empty and K consists only of the first two columns.) Note that ε is also orthogonal to the first column of K:

$$\frac{\gamma(P,Z)^T \Sigma_{PP}^{-1}\varepsilon}{\gamma(P,Z)^T \Sigma_{PP}^{-1}\Sigma_{PP}^{-1}\gamma(P,Z)} = \frac{\gamma(P,Z)^T \Sigma_{PP}^{-1}\begin{pmatrix} [\gamma-\gamma(P,Z)] - \\ \Sigma_{PP}^{-1}\gamma(P,Z)\frac{\gamma(P,Z)^T\Sigma_{PP}^{-1}[\gamma-\gamma(P,Z)]}{\gamma(P,Z)^T\Sigma_{PP}^{-1}\Sigma_{PP}^{-1}\gamma(P,Z)}\end{pmatrix}}{\gamma(P,Z)^T \Sigma_{PP}^{-1}\Sigma_{PP}^{-1}\gamma(P,Z)} = 0 \tag{7.23}$$

We define the matrix Ω:

$$\Omega = K^{-1}\Sigma_{\eta\eta}\left(K^T\right)^{-1} \tag{7.24}$$

We can now rewrite Equation 7.18 as:

$$\gamma - \gamma(P,Z) = K\Omega K^T \Sigma_{PP}^{-1}\gamma(P,Z) \tag{7.25}$$

where Ω is a positive semidefinite matrix. Substituting in the definition of K, we find:

$$\gamma - \gamma(P,Z) = \begin{bmatrix} \frac{\Sigma_{PP}^{-1}\gamma(P,Z)}{\gamma(P,Z)^T\Sigma_{PP}^{-1}\Sigma_{PP}^{-1}\gamma(P,Z)} & \varepsilon & H \end{bmatrix}\Omega\begin{bmatrix} 1 \\ 0 \\ 0_{(N-2)\times 1}\end{bmatrix} \tag{7.26}$$

We choose Ω such that:

$$\Omega\begin{bmatrix} 1 \\ 0 \\ 0_{(N-2)\times 1}\end{bmatrix} = \begin{bmatrix} \gamma(P,Z)^T\Sigma_{PP}^{-1}[\gamma-\gamma(P,Z)] \\ 1 \\ 0_{(N-2)\times 1}\end{bmatrix} \tag{7.27}$$

We then have:

$$\begin{aligned} \gamma - \gamma(P,Z) &= \begin{bmatrix} \frac{\Sigma_{PP}^{-1}\gamma(P,Z)}{\gamma(P,Z)^T\Sigma_{PP}^{-1}\Sigma_{PP}^{-1}\gamma(P,Z)} & \varepsilon & H \end{bmatrix}\begin{bmatrix} \gamma(P,Z)^T\Sigma_{PP}^{-1}[\gamma-\gamma(P,Z)] \\ 1 \\ 0_{(N-2)\times 1}\end{bmatrix} && (7.28) \\ &= \frac{\Sigma_{PP}^{-1}\gamma(P,Z)}{\gamma(P,Z)^T\Sigma_{PP}^{-1}\Sigma_{PP}^{-1}\gamma(P,Z)}\gamma(P,Z)^T\Sigma_{PP}^{-1}[\gamma-\gamma(P,Z)] + \varepsilon && (7.29) \\ &= \gamma - \gamma(P,Z) && (7.30) \end{aligned}$$

By assumption, $\gamma(P,Z)^T\Sigma_{PP}^{-1}[\gamma-\gamma(P,Z)]$ is positive, so the matrix Ω can be chosen to be positive semidefinite. Since $\Sigma_{\eta\eta} = K\Omega K^T$ and K is non-singular, $\Sigma_{\eta\eta}$ is also positive semidefinite; from Equation 7.30, this choice of $\Sigma_{\eta\eta}$ generates the specified vector of risk premia γ.QED

7.4 Proof of Theorem 3

Let G be a set of additional factors, such that the union of F and G, called H, is a linear factor model for Z. Consider the partitioned model:

$$H = \begin{bmatrix} F \\ G \end{bmatrix} = \begin{bmatrix} \Gamma_F^T \\ \Gamma_G^T \end{bmatrix}(Z - \mu_Z) + \begin{bmatrix} \eta \\ \delta \end{bmatrix} = \begin{bmatrix} P \\ Q \end{bmatrix} + \begin{bmatrix} \eta \\ \delta \end{bmatrix} \tag{7.31}$$

where η and δ are uncorrelated with Z. The vector of risk premia for the projection of H onto the asset space is given by:

$$\gamma\left(\begin{bmatrix} P \\ Q \end{bmatrix}, Z\right) = \begin{bmatrix} \Gamma_F^T \\ \Gamma_G^T \end{bmatrix} \mu_Z \tag{7.32}$$

The vector of risk premia for H is given by:

$$\begin{aligned} \gamma\left(H, Z\right) &= \gamma\left(\begin{bmatrix} P \\ Q \end{bmatrix}, Z\right) + \begin{bmatrix} \Sigma_{\eta\eta} & \Sigma_{\eta\delta} \\ \Sigma_{\delta\eta} & \Sigma_{\delta\delta} \end{bmatrix} \begin{bmatrix} \Gamma_F^T \Sigma_{ZZ} \Gamma_F & \Gamma_F^T \Sigma_{ZZ} \Gamma_G \\ \Gamma_G^T \Sigma_{ZZ} \Gamma_F & \Gamma_G^T \Sigma_{ZZ} \Gamma_G \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_F^T \\ \Gamma_G^T \end{bmatrix} \mu_Z \qquad (7.33) \\ &= \gamma\left(\begin{bmatrix} P \\ Q \end{bmatrix}, Z\right) + \begin{bmatrix} \Sigma_{\eta\eta} & \Sigma_{\eta\delta} \\ \Sigma_{\delta\eta} & \Sigma_{\delta\delta} \end{bmatrix} \Phi \mu_Z \qquad (7.34) \end{aligned}$$

where:

$$\Phi = \begin{bmatrix} \begin{bmatrix} \Gamma_F^T \Sigma_{ZZ} \Gamma_F - \\ \Gamma_F^T \Sigma_{ZZ} \Gamma_G \left(\Gamma_G^T \Sigma_{ZZ} \Gamma_G\right)^{-1} \Gamma_G^T \Sigma_{ZZ} \Gamma_F \end{bmatrix}^{-1} \cdot \begin{bmatrix} \Gamma_F^T - \\ \Gamma_F^T \Sigma_{ZZ} \Gamma_G \left(\Gamma_G^T \Sigma_{ZZ} \Gamma_G\right)^{-1} \Gamma_G^T \end{bmatrix} \\ \begin{bmatrix} \Gamma_G^T \Sigma_{ZZ} \Gamma_G - \\ \Gamma_G^T \Sigma_{ZZ} \Gamma_F \left(\Gamma_F^T \Sigma_{ZZ} \Gamma_F\right)^{-1} \Gamma_F^T \Sigma_{ZZ} \Gamma_G \end{bmatrix}^{-1} \cdot \begin{bmatrix} \Gamma_G^T - \\ \Gamma_G^T \Sigma_{ZZ} \Gamma_F \left(\Gamma_F^T \Sigma_{ZZ} \Gamma_F\right)^{-1} \Gamma_F^T \end{bmatrix} \end{bmatrix} \tag{7.35}$$

Since H is a linear factor model, it must be the case that:

$$\begin{aligned} \mu_Z &= \beta_H^T \gamma\left(H, Z\right) = \Sigma_{ZH} \Sigma_{HH}^{-1} \gamma\left(H, Z\right) = \begin{bmatrix} \Sigma_{ZF} \\ \Sigma_{ZG} \end{bmatrix} \Sigma_{HH}^{-1} \gamma\left(H, Z\right) \qquad (7.36) \\ &= \Sigma_{ZF} \lambda_F + \Sigma_{ZG} \lambda_G = \Sigma_{ZZ} \Gamma_F \lambda_F + \Sigma_{ZZ} \Gamma_G \lambda_G \qquad (7.37) \end{aligned}$$

for some λ_F and λ_G. Plugging this value for μ_Z into Equation 7.34, we find:

$$\gamma\left(H, Z\right) - \gamma\left(\begin{bmatrix} P \\ Q \end{bmatrix}, Z\right) = \begin{bmatrix} \Sigma_{\eta\eta} \lambda_F + \Sigma_{\eta\delta} \lambda_G \\ \Sigma_{\delta\eta} \lambda_F + \Sigma_{\delta\delta} \lambda_G \end{bmatrix} \tag{7.38}$$

Consider the case in which the factors G have no unspanned components (i.e., $\delta = 0$). In this case, we have:

$$\gamma\left(H, Z\right) - \gamma\left(\begin{bmatrix} P \\ Q \end{bmatrix}, Z\right) = \begin{bmatrix} \Sigma_{\eta\eta} \lambda_F \\ 0 \end{bmatrix} \tag{7.39}$$

The matrix Ψ in this case is given by:

$$\Psi = \begin{bmatrix} I_N \\ 0_{K \times N} \end{bmatrix} \tag{7.40}$$

where K is the number of additional factors contained in G. Then the desired relation is:

$$\begin{aligned} \gamma &= \Psi^T \gamma\left(H, Z\right) = \Psi^T \gamma\left(\begin{bmatrix} P \\ Q \end{bmatrix}, Z\right) + \Psi^T \begin{bmatrix} \Sigma_{\eta\eta} \lambda_F \\ 0 \end{bmatrix} = \gamma\left(P, Z\right) + \Sigma_{\eta\eta} \lambda_F \qquad (7.41) \\ \gamma - \gamma\left(P, Z\right) &= \Sigma_{\eta\eta} \lambda_F \qquad (7.42) \end{aligned}$$

Let L be the rank of $\Sigma_{\eta\eta}$. Let $w_1, ..., w_{N-L}$ be $N - L$ linearly independent vectors such that $w_i^T \Sigma_{\eta\eta} w_i = 0$ for each $1 \leq i \leq N - L$, and let $v_1, ..., v_L$ be a set of L vectors linearly independent of each other and of

$w_1, ..., w_{L-N}$. By assumption, $w_i^T(\gamma - \gamma(P, Z)) = 0$ for each $1 \leq i \leq N - L$. $\gamma - \gamma(P, Z)$ can therefore be expressed as a linear combination of the vectors $v_1, ..., v_L$; furthermore, each v_i, $1 \leq i \leq L$ can be expressed as $v_i = \Sigma_{\eta\eta} x$ for some vector x. There therefore exists a vector λ such that:

$$\gamma - \gamma(P, Z) = \Sigma_{\eta\eta}\lambda \tag{7.43}$$

We therefore need only choose a set of factors G, such that each factor in G is spanned, the union of F and G is a linear factor model for Z, and the value of λ_F corresponding to this choice of G is equal to λ. We choose:

$$G = \left(\mu_Z^T \Sigma_{ZZ}^{-1} - \lambda^T \Gamma_F^T\right)(Z - \mu_Z) \tag{7.44}$$

G is clearly spanned by the assets Z. Σ_{HH} can only be singular if the factor G can be expressed as a linear combination of P. Suppose this is the case for some N-vector x:

$$\begin{aligned}
\left(\mu_Z^T \Sigma_{ZZ}^{-1} - \lambda^T \Gamma_F^T\right)(Z - \mu_Z) &= x^T P = x^T \Gamma^T (Z - \mu_Z) && (7.45)\\
\left(\mu_Z^T \Sigma_{ZZ}^{-1} - \lambda^T \Gamma_F^T - x^T \Gamma^T\right) &= 0 && (7.46)\\
\mu_Z &= \Sigma_{ZZ}\Gamma_F(\lambda + x) = \Sigma_{ZF}(\lambda + x) && (7.47)\\
&= \Sigma_{ZF}\Sigma_{FF}^{-1}\Sigma_{FF}(\lambda + x) = \beta_F^T \Sigma_{FF}(\lambda + x) && (7.48)
\end{aligned}$$

The last result shows that, contrary to the assumptions of the theorem, F is a linear factor model for Z. Σ_{HH} is therefore non-singular. Similarly, Σ_{HZ} is given by:

$$\Sigma_{HZ} = \begin{bmatrix} \Sigma_{FZ} \\ \Sigma_{GZ} \end{bmatrix} = \begin{bmatrix} \Sigma_{FZ} \\ \left(\mu_Z^T \Sigma_{ZZ}^{-1} - \lambda^T \Gamma_F^T\right)\Sigma_{ZZ} \end{bmatrix} = \begin{bmatrix} \Sigma_{FZ} \\ \left(\mu_Z^T - \lambda^T \Sigma_{FZ}\right) \end{bmatrix} \tag{7.49}$$

For this matrix to have rank N (rather than $N + 1$), there must exist an N-vector x and constant y such that either $x \neq 0$ or $y \neq 0$ and:

$$\begin{bmatrix} x^T & y \end{bmatrix} \Sigma_{HZ} = \begin{bmatrix} x^T & y \end{bmatrix} \begin{bmatrix} \Sigma_{FZ} \\ \left(\mu_Z^T - \lambda^T \Sigma_{FZ}\right) \end{bmatrix} = x^T \Sigma_{FZ} + y\left(\mu_Z^T - \lambda^T \Sigma_{FZ}\right) = 0 \tag{7.50}$$

If $y = 0$, and there exists such an $x \neq 0$, then F and Z, contrary to the assumptions of the theorem, do not satisfy the full-rank assumptions. But if $y \neq 0$, then the last equation becomes:

$$\begin{aligned}
x^T \Sigma_{FZ} + y\left(\mu_Z^T - \lambda^T \Sigma_{FZ}\right) &= 0 && (7.51)\\
\mu_Z y &= \Sigma_{ZF}\lambda y - \Sigma_{ZF} x && (7.52)\\
\mu_Z &= \Sigma_{ZF}\left(\lambda - \frac{x}{y}\right) = \beta_F^T \Sigma_{FF}\left(\lambda - \frac{x}{y}\right) && (7.53)
\end{aligned}$$

F is then, contrary to the theorem assumptions, a linear factor model for Z. It must therefore be the case that Σ_{HZ} has rank $N + 1$. H and Z therefore satisfy the full-rank assumptions. Finally, we note that:

$$\begin{aligned}
\Sigma_{ZH} &= \begin{bmatrix} \Sigma_{ZF} & (\mu_Z - \Sigma_{ZF}\lambda) \end{bmatrix} && (7.54)\\
\Sigma_{ZH}\begin{bmatrix} \lambda \\ 1 \end{bmatrix} &= \begin{bmatrix} \Sigma_{ZF} & (\mu_Z - \Sigma_{ZF}\lambda) \end{bmatrix}\begin{bmatrix} \lambda \\ 1 \end{bmatrix} = \Sigma_{ZF}\lambda + \mu_Z - \Sigma_{ZF}\lambda = \mu_Z && (7.55)\\
\mu_Z &= \Sigma_{ZH}\begin{bmatrix} \lambda \\ 1 \end{bmatrix} = \Sigma_{ZH}\Sigma_{HH}^{-1}\Sigma_{HH}\begin{bmatrix} \lambda \\ 1 \end{bmatrix} = \beta_H^T \Sigma_{HH}\begin{bmatrix} \lambda \\ 1 \end{bmatrix} && (7.56)
\end{aligned}$$

So H is a linear factor model for Z. Construction of H therefore requires selection of a vector λ that satisfies Equation 7.43, and then selecting G according to Equation 7.44.QED

7.5 Proof of Theorem 4

Since F is a linear factor model for Z, we have:

$$\mu_Z = \beta_F^T \gamma(F, Z) = \Sigma_{ZF} \Sigma_{FF}^{-1} \gamma(F, Z) \tag{7.57}$$

This expression can be manipulated to yield:

$$\mu_Z = \Sigma_{ZF} \Sigma_{FF}^{-1} \gamma(F, Z) = \Sigma_{ZG} \Psi \Sigma_{FF}^{-1} \gamma(F, Z) = \Sigma_{ZG} \Sigma_{GG}^{-1} \Sigma_{GG} \Psi \Sigma_{FF}^{-1} \gamma(F, Z) = \beta_G^T \Sigma_{GG} \Psi \Sigma_{FF}^{-1} \gamma(F, Z) \tag{7.58}$$

From the last expression, G is a linear factor model for Z with:

$$\gamma(G, Z) = \Sigma_{GG} \Psi \Sigma_{FF}^{-1} \gamma(F, Z) = \Sigma_{GG} \Psi \left(\Psi^T \Sigma_{GG} \Psi\right)^{-1} \gamma(F, Z) \tag{7.59}$$

Premultiplication of both sides by Ψ^T yields the desired result:

$$\Psi^T \gamma(G, Z) = \Psi^T \Sigma_{GG} \Psi \left(\Psi^T \Sigma_{GG} \Psi\right)^{-1} \gamma(F, Z) = \gamma(F, Z) \tag{7.60}$$

QED

7.6 Proof of Lemma 2

From the definition, $Y = \Phi^T Z$ for some $M \times (M-1)$ full-rank matrix Φ such that $w^T \Sigma_{ZZ} \Phi = 0$. It is clear that Σ_{YY} is finite and full-rank. Σ_{FF} is finite and full-rank by assumption. It remains only to show that Σ_{ZY} has rank N.

Suppose there exists an N-vector x such that $\Sigma_{ZF} x = \Sigma_{ZZ} w$. Then:

$$\Sigma_{YF} x = \Phi^T \Sigma_{ZF} x = \Phi^T \Sigma_{ZZ} w = 0 \tag{7.61}$$

The existence of such an x implies that Σ_{YF} has rank of less than N. To show the other direction of the implication, suppose there exists an $x \neq 0$ such that $\Sigma_{YF} x = 0$. Then:

$$\Sigma_{YF} x = \Phi^T \Sigma_{ZF} x = \Phi^T \Sigma_{ZZ} \Sigma_{ZZ}^{-1} \Sigma_{ZF} x = 0 \tag{7.62}$$

However, Φ is an $M \times (M-1)$ matrix with rank $M-1$, and:

$$\Phi^T \Sigma_{ZZ} w = 0 \tag{7.63}$$

For any vector y such that:

$$\Phi^T \Sigma_{ZZ} y = 0 \tag{7.64}$$

it must be the case that $y = w \cdot k$ for some constant k; furthermore, if $y \neq 0$, then $k \neq 0$. Equation 7.62 therefore implies:

$$\begin{aligned} \Sigma_{ZZ}^{-1} \Sigma_{ZF} x &= w \cdot k & (7.65) \\ \Sigma_{ZF} \frac{x}{k} &= \Sigma_{ZZ} w & (7.66) \end{aligned}$$

QED

7.7 Proof of Theorem 5

First, we note that, if Z consists of a single asset, then F is trivially a linear factor model for Z. Since we have assumed F is not a linear factor model for Z, it must be the case that $M \geq 2$. Choose $w = \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T\gamma_0\right)$, and let $Y \in Z \ominus w^T Z$. The $Y = \Phi^T Z$ for some $M \times (M-1)$ full-rank matrix Φ, with $\Phi^T\Sigma_{ZZ}w = 0$. Substituting in the value for w, we find:

$$\begin{aligned}\Phi^T\Sigma_{ZZ}\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T\gamma_0\right) &= 0 && (7.67)\\ \Phi^T\mu_Z &= \Phi^T\beta^T\gamma_0 && (7.68)\\ \Phi^T\mu_Z &= \Phi^T\Sigma_{ZF}\Sigma_{FF}^{-1}\gamma_0 && (7.69)\end{aligned}$$

Since $\mu_Y = \Phi^T\mu_Z$ and $\Sigma_{YF} = \Phi^T\Sigma_{ZF}$, it follows that:

$$\mu_Y = \Sigma_{YF}\Sigma_{FF}^{-1}\gamma_0 = \beta_Y^T\gamma_0 \tag{7.70}$$

It is clear then that the expected excess returns of Y obey the required relationship to their β coefficients on F; it remains only to show that F and Y satisfy the full-rank assumptions. By Lemma 2, this will fail to be the case only if there exists some N-vector x such that $\Sigma_{ZF}x = \Sigma_{ZZ}w$. Let x be such a vector:

$$\begin{aligned}\Sigma_{ZF}x &= \Sigma_{ZZ}w = \Sigma_{ZZ}\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T\gamma_0\right) = \left(\mu_Z - \beta^T\gamma_0\right) && (7.71)\\ \mu_Z &= \beta^T\gamma_0 + \Sigma_{ZF}x = \beta^T\gamma_0 + \Sigma_{ZF}\Sigma_{FF}^{-1}\Sigma_{FF}x = \beta^T\left(\gamma_0 + \Sigma_{FF}x\right) && (7.72)\end{aligned}$$

The last result implies that F is a linear factor model for Z, with risk premia vector equal to $(\gamma_0 + \Sigma_{FF}x)$. Since, by assumption, F is not a linear factor model for Z, no such x exists.

Let w_0 be another vector such that $\gamma\left(F, Z \ominus w_0^T Z\right) = \gamma_0$. Let X be any element of $Z \ominus w_0^T Z$, and let Φ be an $M \times (M-1)$ full-rank matrix such that $X = \Phi^T Z$. Then:

$$\begin{aligned}\mu_X &= \beta_X^T\gamma_0 = \Sigma_{XF}\Sigma_{FF}^{-1}\gamma_0 && (7.73)\\ \Phi^T\mu_Z &= \Phi^T\Sigma_{ZF}\Sigma_{FF}^{-1}\gamma_0 && (7.74)\\ \Phi^T\left(\mu_Z - \beta\gamma_0\right) &= 0 && (7.75)\\ \Phi^T\Sigma_{ZZ}\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta\gamma_0\right) &= 0 && (7.76)\end{aligned}$$

$\Phi^T\Sigma_{ZZ}w_0 = 0$, by definition, and from the last result, $\Phi^T\Sigma_{ZZ}w = 0$. Since Φ has rank $M-1$ and $w \neq 0$, it must be the case that $w_0 = w \cdot c$ for some constant c.QED

7.8 Proof of Lemma 3

That $\mu_Z^T\Sigma_{ZZ}^{-1}Z$ is a mean-variance efficient portfolio is well-known. Direct evaluation shows that its expected excess return is $\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z$, and that its variance is also $\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z$. The squared Sharpe ratio is therefore:

$$S^2(Z) = \frac{\left(\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z\right)^2}{\left(\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z\right)} = \mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z \tag{7.77}$$

It is also well-known that the maximum squared Sharpe ratio offered by two uncorrelated sets of assets is equal to the sum of the maximum squared Sharpe ratios offered be each set. Since Y and $Z \ominus Y$ are uncorrelated, and their union is a basis for Z, we have:

$$\begin{aligned} S^2(Z) &= S^2(Z \ominus Y) + S^2(Y) && (7.78) \\ S^2(Y) &= \mu_Y^T \Sigma_{YY}^{-1} \mu_Y = \mu_Z^T \Psi \left(\Psi^T \Sigma_{ZZ} \Psi\right)^{-1} \Psi^T \mu_Z && (7.79) \\ S^2(Z \ominus Y) &= S^2(Z) - S^2(Y) = \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \mu_Z^T \Psi \left(\Psi^T \Sigma_{ZZ} \Psi\right)^{-1} \Psi^T \mu_Z && (7.80) \end{aligned}$$

QED

7.9 Proof of Theorem 6

From Theorem 5, we can take $w = \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)$. The result follows by direct substitution of this value of w for Ψ in Lemma 3:

$$\begin{aligned} S^2(Z \ominus Y) &= \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \mu_Z^T w \left(w^T \Sigma_{ZZ} w\right)^{-1} w^T \mu_Z && (7.81) \\ &= \frac{\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \mu_Z^T \left[\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)\right] \left[\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)\right]^T \mu_Z}{\left[\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)\right]^T \Sigma_{ZZ} \left[\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)\right]} && (7.82) \\ &= \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \frac{\left[\mu_Z^T \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)\right]^2}{\left(\mu_Z^T - \gamma_0^T \beta\right) \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)} && (7.83) \end{aligned}$$

QED

7.10 Proof of Theorem 7

From Definition 5 and Theorem 6:

$$SSRR(F, Z, \gamma_0) = \frac{\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \frac{\left[\mu_Z^T \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)\right]^2}{\left(\mu_Z - \beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T \gamma_0\right)}}{\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z} \qquad (7.84)$$

Setting this value equal to k and rearranging, we find:

$$\left[\begin{array}{r} (1-k)\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)\left(\gamma_0^T \beta \Sigma_{ZZ}^{-1} \beta^T \gamma_0\right) - \left(\mu_Z^T \Sigma_{ZZ}^{-1} \beta^T \gamma_0\right)^2 \\ +2k\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)\left(\mu_Z^T \Sigma_{ZZ}^{-1} \beta^T \gamma_0\right) - k\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)^2 \end{array}\right] = 0 \qquad (7.85)$$

Note that this equation is quadratic in γ_0; the nature of the equation (i.e., parabolic, elliptical, hyperbolic, or degenerate) depends on the value of k. We can express any vector of risk premia γ_0 as:

$$\gamma_0 = \gamma(F, Z)\, b + \xi \qquad (7.86)$$

where b is a scalar and ξ is an N-element vector, satisfying the condition:

$$\gamma^T(F, Z)\left(\Sigma_{FF}^{-1} \Sigma_{PP} \Sigma_{FF}^{-1}\right) \xi = 0 \qquad (7.87)$$

Substituting in the definition of $\gamma(F, Z)$, this condition simplifies to:

$$\mu_Z^T \Gamma \Sigma_{FF}^{-1} \xi = \mu_Z^T \Sigma_{ZZ}^{-1} \beta^T \xi = 0 \tag{7.88}$$

Substituting the value for γ_0 from Equation 7.86 into Equation 7.85, we find:

$$\left[\begin{array}{c} b^2 \left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right) \left[(1-k)\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right) - \left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right)\right] \\ +(1-k)\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)\left(\xi^T \Sigma_{FF}^{-1} \Sigma_{PP} \Sigma_{FF}^{-1} \xi\right) \\ +2bk\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)\left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right) - k\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)^2 \end{array}\right] = 0 \tag{7.89}$$

We first consider the case $k = 0$:

$$b^2 \left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right)\left[\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right) - \mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right] + \left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)\left(\xi^T \Sigma_{FF}^{-1} \Sigma_{PP} \Sigma_{FF}^{-1} \xi\right) = 0 \tag{7.90}$$

At this point, we note:

$$S^2\left(\mu_Z^T \Sigma_{ZZ}^{-1} Z\right) = \left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right) > \mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z = S^2\left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T Z\right) \tag{7.91}$$

This inequality follows from the fact that $\mu_Z^T \Sigma_{ZZ}^{-1} Z$ is mean-variance efficient, whereas $\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma Z$ is not. All terms on the left-hand side of Equation 7.90 are non-negative, so the only solution is $b = 0$ and $\xi = 0_{N\times 1}$; the corresponding value of γ_0 is $0_{N\times 1}$. This proves the first assertion. For the second, $k = 1$, Equation 7.89 simplifies to:

$$-b^2\left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right)^2 + 2b\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)\left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right) - \left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)^2 = 0 \tag{7.92}$$

which is the negative of a perfect square:

$$-\left[b\left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right) - \left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)\right]^2 = 0 \tag{7.93}$$

The unique value of b which solves this equation is:

$$b = \frac{\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z}{\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z} \tag{7.94}$$

The solution does not depend on ξ; γ_0 therefore lies within an $(N-1)$-dimensional hyperplane, proving the second assertion. We now consider the boundary case:

$$k = k_0 = 1 - \frac{\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z}{\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z} \tag{7.95}$$

In this case, Equation 7.89 becomes:

$$\left[\begin{array}{c} \left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right)\left(\xi^T \Sigma_{FF}^{-1} \Sigma_{PP} \Sigma_{FF}^{-1} \xi\right) \\ +2b\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right)\left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right) \\ -\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right)\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right) \end{array}\right] = 0 \tag{7.96}$$

This equation is parabolic; note that the first term is positive for any non-zero value of ξ. The value of b that corresponds to $\xi^T \Sigma_{FF}^{-1} \Sigma_{PP} \Sigma_{FF}^{-1} \xi = 0$ (which, since $\Sigma_{FF}^{-1} \Sigma_{PP} \Sigma_{FF}^{-1}$ is full-rank, requires $\xi = 0$) is equal to half the value that specifies the $k = 1$ hyperplane:

$$b = \frac{\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)}{2\left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right)} \tag{7.97}$$

If we set $b = 0$, we find that this hyperplane intersects the paraboloid:

$$\xi^T \Sigma_{FF}^{-1} \Sigma_{PP} \Sigma_{FF}^{-1} \xi = \left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - \mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right) \frac{\left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)}{\left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right)} \tag{7.98}$$

We can therefore conclude that this paraboloid, whose vertex is halfway between the origin and the $k = 1$ hyperplane, opens towards the origin. The third assertion now proven, we consider the hyperbolic case:

$$1 - \frac{\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z}{\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z} < k < 1 \tag{7.99}$$

We can then rewrite Equation 7.89 as:

$$\left[\begin{array}{c} -b^2 \left(k - k_0\right) \left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right) \left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right) \\ + \left(1 - k\right) \left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right) \left(\xi^T \Sigma_{FF}^{-1} \Sigma_{PP} \Sigma_{FF}^{-1} \xi\right) \\ + 2bk \left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right) \left(\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z\right) - k \left(\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right)^2 \end{array}\right] = 0 \tag{7.100}$$

where $(k - k_0) > 0$. Note that this equation is hyperbolic; furthermore, if we set $\xi = 0$, the two values of b that solve the equation lie on opposite sides of the $k = 1$ hyperplane. The two branches of the hyperboloid open outward, facing away from the $k = 1$ hyperplane. The fourth assertion now demonstrated, we turn to the elliptical case:

$$0 < k < 1 - \frac{\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z}{\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z} \tag{7.101}$$

In this case, we still arrive at Equation 7.100, but the leading term is non-negative rather than non-positive, making the equation elliptical. Note that the entire ellipsoid lies on one side of the $k = 1$ hyperplane.

The five assertions have now been demonstrated for $N > 1$; the results for $N = 1$ are trivial restrictions of the general results obtained by setting $\xi = 0$.QED

7.11 Proof of Theorem 8

Let Y be any element of $Z \ominus w^T Z$, and let Φ be an $M \times (M - 1)$ full-rank matrix such that $Y = \Phi^T Z$. The desired relation is:

$$\begin{align} \gamma\left(F, Z \ominus w^T Z\right) &= \gamma_0 \tag{7.102} \\ \Sigma_{FF} \left(\Sigma_{FY} \Sigma_{YY}^{-1} \Sigma_{YF}\right)^{-1} \Sigma_{FY} \Sigma_{YY}^{-1} \mu_Y &= \gamma_0 \tag{7.103} \\ \Sigma_{FF} \left[\Sigma_{FZ} \Phi \left(\Phi^T \Sigma_{ZZ} \Phi\right)^{-1} \Phi^T \Sigma_{ZF}\right]^{-1} \Sigma_{FZ} \Phi \left(\Phi^T \Sigma_{ZZ} \Phi\right)^{-1} \Phi^T \mu_Z &= \gamma_0 \tag{7.104} \\ \left[\Sigma_{FZ} \Phi \left(\Phi^T \Sigma_{ZZ} \Phi\right)^{-1} \Phi^T \Sigma_{ZF}\right]^{-1} \Sigma_{FZ} \Phi \left(\Phi^T \Sigma_{ZZ} \Phi\right)^{-1} \Phi^T \mu_Z - \Sigma_{FF}^{-1} \gamma_0 &= 0 \tag{7.105} \\ \Sigma_{FZ} \Phi \left(\Phi^T \Sigma_{ZZ} \Phi\right)^{-1} \Phi^T \left[\mu_Z - \beta^T \gamma_0\right] &= 0 \tag{7.106} \end{align}$$

where $\gamma_0 = \gamma(F, Z)$. One readily verifies that:

$$\left[\Phi \left(\Phi^T \Sigma_{ZZ} \Phi\right)^{-1} \Phi^T \Sigma_{ZZ}\right] h = h - \left[w \left(w^T \Sigma_{ZZ} w\right)^{-1} w^T \Sigma_{ZZ}\right] h \tag{7.107}$$

for any M-vector h. Equation 7.106 can therefore be rewritten as:

$$\begin{aligned} 0 &= \Sigma_{FZ}\left[I - w\left(w^T\Sigma_{ZZ}w\right)^{-1}w^T\Sigma_{ZZ}\right]\Sigma_{ZZ}^{-1}\left[\mu_Z - \beta^T\gamma_0\right] && (7.108)\\ \left[\Gamma^T\mu_Z - \Gamma^T\beta^T\gamma_0\right] &= \Sigma_{FZ}w\left(w^T\Sigma_{ZZ}w\right)^{-1}w^T\left[\mu_Z - \beta^T\gamma_0\right] && (7.109)\\ \left[\Gamma^T\mu_Z - \Sigma_{PP}\Sigma_{FF}^{-1}\gamma_0\right] &= \Sigma_{FZ}w\left(w^T\Sigma_{ZZ}w\right)^{-1}w^T\left[\mu_Z - \beta^T\gamma_0\right] && (7.110)\end{aligned}$$

Substituting in the value of γ_0, we find:

$$\begin{aligned} \left[\Gamma^T\mu_Z - \Gamma^T\mu_Z\right] &= \Sigma_{FZ}w\left(w^T\Sigma_{ZZ}w\right)^{-1}w^T\left[\mu_Z - \beta^T\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right] && (7.111)\\ 0 &= \Sigma_{FZ}w\left(w^T\Sigma_{ZZ}w\right)^{-1}w^T\Sigma_{ZZ}\left[\Sigma_{ZZ}^{-1}\mu_Z - \Sigma_{ZZ}^{-1}\Sigma_{ZF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right] && (7.112)\\ 0 &= \Sigma_{FZ}w\left(w^T\Sigma_{ZZ}w\right)^{-1}w^T\Sigma_{ZZ}\left[\Sigma_{ZZ}^{-1} - \Gamma\Sigma_{PP}^{-1}\Gamma^T\right]\mu_Z && (7.113)\end{aligned}$$

We now express w as follows:

$$w = \Gamma\varphi + \varepsilon c + \varsigma \qquad (7.114)$$

where φ is an N-vector, c is a constant, ε is defined as:

$$\varepsilon = \left[\Sigma_{ZZ}^{-1} - \Gamma\Sigma_{PP}^{-1}\Gamma^T\right]\mu_Z \qquad (7.115)$$

and ς is any M-vector such that:

$$\begin{aligned} \Gamma^T\Sigma_{ZZ}\varsigma &= 0 && (7.116)\\ \varepsilon^T\Sigma_{ZZ}\varsigma &= 0 && (7.117)\end{aligned}$$

Note that:

$$\begin{aligned} \Gamma^T\Sigma_{ZZ}\varepsilon &= \Gamma^T\Sigma_{ZZ}\left[\Sigma_{ZZ}^{-1} - \Gamma\Sigma_{PP}^{-1}\Gamma^T\right]\mu_Z = \left[\Gamma^T - \Gamma^T\Sigma_{ZF}\Sigma_{PP}^{-1}\Gamma^T\right]\mu_Z && (7.118)\\ &= \left[\Gamma^T - \Sigma_{PP}\Sigma_{PP}^{-1}\Gamma^T\right]\mu_Z = \left[\Gamma^T - \Gamma^T\right]\mu_Z = 0 && (7.119)\end{aligned}$$

Also note that the assumptions of the theorem require either $c \neq 0$ or $\varsigma \neq 0$. Substituting this decomposition of w into Equation 7.113, we find:

$$\begin{aligned} 0 &= \Sigma_{FZ}\left(\Gamma\varphi + \varepsilon c + \varsigma\right)\frac{\left(\Gamma\varphi + \varepsilon c + \varsigma\right)^T\Sigma_{ZZ}\left[\Sigma_{ZZ}^{-1} - \Gamma\Sigma_{PP}^{-1}\Gamma^T\right]\mu_Z}{\left(\Gamma\varphi + \varepsilon c + \varsigma\right)^T\Sigma_{ZZ}\left(\Gamma\varphi + \varepsilon c + \varsigma\right)} && (7.120)\\ 0 &= \frac{\Sigma_{FZ}\left(\Gamma\varphi + \varepsilon c + \varsigma\right)\left(\Gamma\varphi + \varepsilon c + \varsigma\right)^T\Sigma_{ZZ}\varepsilon}{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\varphi + c^2\varepsilon^T\Sigma_{ZZ}\varepsilon + \varsigma^T\Sigma_{ZZ}\varsigma} = \frac{\Sigma_{FZ}\Sigma_{ZZ}^{-1}\Sigma_{ZZ}\left(\Gamma\varphi + \varepsilon c + \varsigma\right)\left[c\varepsilon^T\Sigma_{ZZ}\varepsilon\right]}{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\varphi + c^2\varepsilon^T\Sigma_{ZZ}\varepsilon + \varsigma^T\Sigma_{ZZ}\varsigma} && (7.121)\\ 0 &= \frac{\Gamma^T\Sigma_{ZZ}\left(\Gamma\varphi + \varepsilon c + \varsigma\right)\left[c\varepsilon^T\Sigma_{ZZ}\varepsilon\right]}{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\varphi + c^2\varepsilon^T\Sigma_{ZZ}\varepsilon + \varsigma^T\Sigma_{ZZ}\varsigma} = \frac{\Gamma^T\Sigma_{ZZ}\Gamma\varphi\left[c\varepsilon^T\Sigma_{ZZ}\varepsilon\right]}{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\varphi + c^2\varepsilon^T\Sigma_{ZZ}\varepsilon + \varsigma^T\Sigma_{ZZ}\varsigma} && (7.122)\\ 0 &= \frac{\Sigma_{FZ}\Sigma_{ZZ}^{-1}\Sigma_{ZF}\varphi\left[c\varepsilon^T\Sigma_{ZZ}\varepsilon\right]}{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\varphi + c^2\varepsilon^T\Sigma_{ZZ}\varepsilon + \varsigma^T\Sigma_{ZZ}\varsigma} = \frac{\Sigma_{PP}\varphi\left[c\varepsilon^T\Sigma_{ZZ}\varepsilon\right]}{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\varphi + c^2\varepsilon^T\Sigma_{ZZ}\varepsilon + \varsigma^T\Sigma_{ZZ}\varsigma} && (7.123)\end{aligned}$$

First, note that $\varepsilon \neq 0.$, since (by assumption) F is not a linear factor model for Z. Therefore, for the last expression to be zero, we must have either $c = 0$ or $\varphi = 0$: These correspond to conditions (i) and (ii) of the theorem, respectively.QED

7.12 Proof of Theorem 9

The proof is the same as in Theorem 8 up until Equation 7.110. We then express the vector as the sum of two components:

$$\Sigma_{ZZ}^{-1}\left(\mu_Z-\beta^T\gamma_0\right)=\Gamma\phi+\varepsilon \tag{7.124}$$

where:

$$\phi=\Sigma_{PP}^{-1}\Gamma^T\mu_Z-\Sigma_{FF}^{-1}\gamma_0 \tag{7.125}$$

and ε is as before. Using the same decomposition for w as in Theorem 8, we express Equation 7.110 as:

$$\begin{aligned}
\left[\Gamma^T\mu_Z-\Sigma_{PP}\Sigma_{FF}^{-1}\gamma_0\right] &= \Sigma_{FZ}\left(\Gamma\varphi+\varepsilon c+\varsigma\right)\frac{\left(\Gamma\varphi+\varepsilon c+\varsigma\right)^T\left[\mu_Z-\beta^T\gamma_0\right]}{\left(\Gamma\varphi+\varepsilon c+\varsigma\right)^T\Sigma_{ZZ}\left(\Gamma\varphi+\varepsilon c+\varsigma\right)} && (7.126)\\
\Gamma^T\left[\mu_Z-\beta^T\gamma_0\right] &= \frac{\Sigma_{FZ}\left(\Gamma\varphi+\varepsilon c+\varsigma\right)\left(\Gamma\varphi+\varepsilon c+\varsigma\right)^T\left[\mu_Z-\beta^T\gamma_0\right]}{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\varphi+c^2\varepsilon^T\Sigma_{ZZ}\varepsilon+\varsigma^T\Sigma_{ZZ}\varsigma} && (7.127)\\
\Sigma_{FZ}\left[\Gamma\phi+\varepsilon\right] &= \frac{\Gamma^T\Sigma_{ZZ}\left(\Gamma\varphi+\varepsilon c+\varsigma\right)\left(\Gamma\varphi+\varepsilon c+\varsigma\right)^T\Sigma_{ZZ}\left[\Gamma\phi+\varepsilon\right]}{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\varphi+c^2\varepsilon^T\Sigma_{ZZ}\varepsilon+\varsigma^T\Sigma_{ZZ}\varsigma} && (7.128)\\
\Gamma^T\Sigma_{ZZ}\Gamma\phi &= \frac{\left(\Gamma^T\Sigma_{ZZ}\Gamma\varphi\right)\left(\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\phi+c\varepsilon^T\Sigma_{ZZ}\varepsilon\right)}{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\varphi+c^2\varepsilon^T\Sigma_{ZZ}\varepsilon+\varsigma^T\Sigma_{ZZ}\varsigma} && (7.129)\\
0 &= \Gamma^T\Sigma_{ZZ}\Gamma\left[\phi-\varphi\frac{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\phi+c\varepsilon^T\Sigma_{ZZ}\varepsilon}{\varphi^T\Gamma^T\Sigma_{ZZ}\Gamma\varphi+c^2\varepsilon^T\Sigma_{ZZ}\varepsilon+\varsigma^T\Sigma_{ZZ}\varsigma}\right] && (7.130)
\end{aligned}$$

Since $\Gamma^T\Sigma_{ZZ}\Gamma$ is full rank, the quantity within brackets must be equal to zero. The cases where $\phi=0$ are covered in Theorem 8; it must therefore be the case that $\phi\neq 0$. It must therefore be the case that $\varphi=\phi\cdot k$ for some $k\neq 0$. If then the above simplifies to:

$$\begin{aligned}
\phi\cdot k\frac{k\phi^T\Gamma^T\Sigma_{ZZ}\Gamma\phi+c\varepsilon^T\Sigma_{ZZ}\varepsilon}{k^2\phi^T\Gamma^T\Sigma_{ZZ}\Gamma\phi+c^2\varepsilon^T\Sigma_{ZZ}\varepsilon+\varsigma^T\Sigma_{ZZ}\varsigma} &= \phi && (7.131)\\
k^2\phi^T\Gamma^T\Sigma_{ZZ}\Gamma\phi+kc\varepsilon^T\Sigma_{ZZ}\varepsilon &= k^2\phi^T\Gamma^T\Sigma_{ZZ}\Gamma\phi+c^2\varepsilon^T\Sigma_{ZZ}\varepsilon+\varsigma^T\Sigma_{ZZ}\varsigma && (7.132)\\
\left(kc-c^2\right)\varepsilon^T\Sigma_{ZZ}\varepsilon &= \varsigma^T\Sigma_{ZZ}\varsigma && (7.133)\\
k &= c+\frac{\varsigma^T\Sigma_{ZZ}\varsigma}{c\left(\varepsilon^T\Sigma_{ZZ}\varepsilon\right)} && (7.134)
\end{aligned}$$

At this point, we note:

$$\varepsilon^T\Sigma_{ZZ}\varepsilon=\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z-\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z \tag{7.135}$$

We can express w as:

$$\begin{aligned}
w &= \Gamma(\Sigma_{PP}^{-1}\Gamma^T\mu_Z-\Sigma_{FF}^{-1}\gamma_0)\left(c+\frac{\varsigma^T\Sigma_{ZZ}\varsigma}{c\left(\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z-\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right)}\right) && (7.136)\\
&\quad+\left[\Sigma_{ZZ}^{-1}-\Gamma\Sigma_{PP}^{-1}\Gamma^T\right]\mu_Z c+\varsigma && (7.137)\\
w &= \begin{bmatrix}\Gamma\Sigma_{PP}^{-1}\Gamma^T(\mu_Z-\beta^T\gamma_0)\left(1+\frac{\varsigma^T\Sigma_{ZZ}\varsigma}{c^2\left(\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z-\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right)}\right)\\ +\left[\Sigma_{ZZ}^{-1}-\Gamma\Sigma_{PP}^{-1}\Gamma^T\right]\mu_Z+\frac{\varsigma}{c}\end{bmatrix}c && (7.138)
\end{aligned}$$

Normalizing ς, the result follows immediately. Note that, for a fixed value of c, this equation is linear in w but quadratic in ξ, describing a paraboloid.QED

7.13 Proof of Theorem 10

$S^2\left(Z \ominus w\left(c,\zeta\right)^T Z\right)$ is maximized when $S^2\left(w\left(c,\zeta\right)^T Z\right)$ is minimized. It will be useful to employ the familiar decomposition:

$$\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T\gamma_0\right) = \Gamma\phi + \varepsilon \tag{7.139}$$

where:

$$\phi^T\Gamma^T\Sigma_{ZZ}\varepsilon = 0 \tag{7.140}$$

Recall that ϕ and ε are given by:

$$\begin{aligned} \phi &= \left[\Sigma_{PP}^{-1}\Gamma^T\mu_Z - \Sigma_{FF}^{-1}\gamma_0\right] && (7.141) \\ \varepsilon &= \left[\Sigma_{ZZ}^{-1} - \Gamma\Sigma_{PP}^{-1}\Gamma^T\right]\mu_Z && (7.142) \end{aligned}$$

The expected excess return and variance of return of $w\left(c,\zeta\right)^T Z$ can then be expressed as:

$$\begin{aligned} E\left[w\left(c,\zeta\right)^T Z\right] &= c\left[\left(\phi^T\Sigma_{PP}\phi + \gamma_0^T\beta\Gamma\phi\right)\left(1 + \frac{\varsigma^T\Sigma_{ZZ}\varsigma}{\varepsilon^T\Sigma_{ZZ}\varepsilon}\right) + \varepsilon^T\Sigma_{ZZ}\varepsilon\right] && (7.143) \\ Var\left[w\left(c,\zeta\right)^T Z\right] &= c^2\left[\phi^T\Sigma_{PP}\phi\left(1 + \frac{\varsigma^T\Sigma_{ZZ}\varsigma}{\varepsilon^T\Sigma_{ZZ}\varepsilon}\right)^2 + \varepsilon^T\Sigma_{ZZ}\varepsilon + \varsigma^T\Sigma_{ZZ}\varsigma\right] && (7.144) \end{aligned}$$

$S^2\left(w\left(c,\zeta\right)^T Z\right)$ therefore does not depend on c, and can be expressed as:

$$S^2\left(w\left(c,\zeta\right)^T Z\right) = \frac{\left[\left(\phi^T\Sigma_{PP}\phi + \gamma_0^T\beta\Gamma\phi\right)\left(1 + \frac{\varsigma^T\Sigma_{ZZ}\varsigma}{\varepsilon^T\Sigma_{ZZ}\varepsilon}\right) + \varepsilon^T\Sigma_{ZZ}\varepsilon\right]^2}{\left[\phi^T\Sigma_{PP}\phi\left(1 + \frac{\varsigma^T\Sigma_{ZZ}\varsigma}{\varepsilon^T\Sigma_{ZZ}\varepsilon}\right)^2 + \varepsilon^T\Sigma_{ZZ}\varepsilon + \varsigma^T\Sigma_{ZZ}\varsigma\right]} \tag{7.145}$$

The minimum value could conceivably occur when $\varsigma^T\Sigma_{ZZ}\varsigma$ is equal to zero, as $\varsigma^T\Sigma_{ZZ}\varsigma$ approaches infinity, or at an interior minimum point. $S^2\left(w\left(c,\zeta\right)^T Z\right)$ is clearly continuously differentiable in $S^2\left(w\left(c,\zeta\right)^T Z\right)$, so calculating the derivative and setting equal to zero, we find there are two potential interior minimum values of $\varsigma^T\Sigma_{ZZ}\varsigma$, which we denote by k_A and k_B:

$$\begin{aligned} k_A &= -\left(\varepsilon^T\Sigma_{ZZ}\varepsilon\right)\left(1 + \frac{\varepsilon^T\Sigma_{ZZ}\varepsilon}{\left(\phi + \Sigma_{FF}^{-1}\gamma_0\right)^T\Sigma_{PP}\phi}\right) && (7.146) \\ k_B &= -\left(\varepsilon^T\Sigma_{ZZ}\varepsilon\right)\left(1 + \frac{\varepsilon^T\Sigma_{ZZ}\varepsilon}{\left(\phi - \Sigma_{FF}^{-1}\gamma_0\right)^T\Sigma_{PP}\phi}\right) && (7.147) \end{aligned}$$

Consider k_A first. This value is positive if and only if:

$$0 > \left(\phi + \Sigma_{FF}^{-1}\gamma_0\right)^T\Sigma_{PP}\phi > -\varepsilon^T\Sigma_{ZZ}\varepsilon \tag{7.148}$$

Plugging in the definitions of ϕ and ε, this condition becomes:

$$\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z < \mu_Z^T\Sigma_{ZZ}^{-1}\beta^T\gamma_0 < \mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z \tag{7.149}$$

which is the region between two parallel hyperplanes. Plugging in the values ϕ and ε into the expression for k_A, we find:

$$k_A = -\left(\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z - \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right)\frac{\mu_Z^T\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T\gamma_0\right)}{\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\left(\mu_Z - \beta^T\gamma_0\right)} \tag{7.150}$$

Finally, the value of $S^2\left(w\left(c,\zeta\right)^T Z\right)$ at k_A is given by:

$$S^2\left(w\left(c,\zeta\right)^T Z\right) = 0 \tag{7.151}$$

Clearly, whenever $k_A > 0$, it will be the minimum point. This proves the statements regarding the second region. To show the other statements, we consider when $k_B > 0$:

$$0 > \left(\phi - \Sigma_{FF}^{-1}\gamma_0\right)^T \Sigma_{PP}\phi > -\varepsilon^T\Sigma_{ZZ}\varepsilon \tag{7.152}$$

Substituting in the values of ϕ and ε, this condition becomes:

$$\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z < \left(3\mu_Z - 4\beta^T\gamma_0\right)^T \Sigma_{ZZ}^{-1}\left(3\mu_Z - 4\beta^T\gamma_0\right) < 9\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z - 8\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z \tag{7.153}$$

This region is evidently that between two ellipsoids. Plugging in the values ϕ and ε into the expression for k_B, we find:

$$k_B = -\left(\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z - \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right)\left[\frac{\left(\mu_Z - \beta^T\gamma_0\right)^T \Sigma_{ZZ}^{-1}\left(\mu_Z - 2\beta^T\gamma_0\right)}{\left(\mu_Z - \beta^T\gamma_0\right)^T \Gamma\Sigma_{PP}^{-1}\Gamma^T\left(\mu_Z - 2\beta^T\gamma_0\right)}\right] \tag{7.154}$$

The value of $S^2\left(w\left(c,\zeta\right)^T Z\right)$ at k_B is given by:

$$S^2\left(w\left(c,\zeta\right)^T Z\right) = 4\left(\gamma_0^T\Sigma_{FF}^{-1}\Sigma_{PP}\phi\right) \tag{7.155}$$

Plugging in the definition of ϕ, we find:

$$S^2\left(w\left(c,\zeta\right)^T Z\right) = 4\gamma_0^T\beta\Sigma_{ZZ}^{-1}\left(\mu_Z - \beta^T\gamma_0\right) \tag{7.156}$$

Since this value is greater than zero, it is not immediately obvious that the minimum occurs at k_B. To check, we consider the value of $S^2\left(w\left(c,\zeta\right)^T Z\right)$ when $\varsigma^T\Sigma_{ZZ}\varsigma$ is equal to zero, and when $\varsigma^T\Sigma_{ZZ}\varsigma$ tends to positive infinity. These values can be found directly from the expression for $S^2\left(w\left(c,\zeta\right)^T Z\right)$:

$$\begin{aligned} S^2\left(w\left(c,0\right)^T Z\right) &= \left[\phi^T\Sigma_{PP}\phi + \gamma_0^T\beta\Gamma\phi + \varepsilon^T\Sigma_{ZZ}\varepsilon\right]\left[1 + \frac{\gamma_0^T\beta\Gamma\phi}{\phi^T\Sigma_{PP}\phi + \varepsilon^T\Sigma_{ZZ}\varepsilon}\right] && (7.157) \\ \lim_{\varsigma^T\Sigma_{ZZ}\varsigma\to+\infty} S^2\left(w\left(c,\varsigma\right)^T Z\right) &= \frac{\left[\phi^T\Sigma_{PP}\phi + \gamma_0^T\beta\Gamma\phi\right]^2}{\left[\phi^T\Sigma_{PP}\phi\right]} && (7.158) \end{aligned}$$

The value at k_B is smaller than both values when both these conditions are met:

$$\begin{aligned} 4\left(\gamma_0^T\beta\Gamma\phi\right) &< \left[\phi^T\Sigma_{PP}\phi + \gamma_0^T\beta\Gamma\phi + \varepsilon^T\Sigma_{ZZ}\varepsilon\right]\left[1 + \frac{\gamma_0^T\beta\Gamma\phi}{\phi^T\Sigma_{PP}\phi + \varepsilon^T\Sigma_{ZZ}\varepsilon}\right] && (7.159) \\ 4\left(\gamma_0^T\beta\Gamma\phi\right) &< \frac{\left[\phi^T\Sigma_{PP}\phi + \gamma_0^T\beta\Gamma\phi\right]^2}{\left[\phi^T\Sigma_{PP}\phi\right]} && (7.160) \end{aligned}$$

The second of these conditions simplifies to:

$$0 < \left[\phi^T\Sigma_{PP}\phi - \gamma_0^T\beta\Gamma\phi\right]^2 \tag{7.161}$$

which is satisfied trivially whenever $k_B > 0$. The first condition simplifies to:

$$0 < \left[\phi^T \Sigma_{PP} \phi - \gamma_0^T \beta \Gamma \phi + \varepsilon^T \Sigma_{ZZ} \varepsilon\right]^2 \tag{7.162}$$

which is also satisfied trivially whenever $k_B > 0$. Therefore, whenever $k_B > 0$, the minimum value of $S^2\left(w\left(c,\zeta\right)^T Z\right)$ is achieved at k_B. This proves the statements concerning the fourth region.

The remaining values of γ_0, can be placed into three distinct categories by a close examination of the $k_A > 0$ and $k_B > 0$ conditions:

$$\gamma_0^T \Sigma_{FF}^{-1} \Sigma_{PP} \phi \geq \phi^T \Sigma_{PP} \phi + \varepsilon^T \Sigma_{ZZ} \varepsilon \tag{7.163}$$

$$\phi^T \Sigma_{PP} \phi \geq \gamma_0^T \Sigma_{FF}^{-1} \Sigma_{PP} \phi \geq -\phi^T \Sigma_{PP} \phi \tag{7.164}$$

$$-\phi^T \Sigma_{PP} \phi - \varepsilon^T \Sigma_{ZZ} \varepsilon \geq \gamma_0^T \Sigma_{FF}^{-1} \Sigma_{PP} \phi \tag{7.165}$$

In all three of these regions, the minimum must be achieved either when $\varsigma^T \Sigma_{ZZ} \varsigma$ is zero or when $\varsigma^T \Sigma_{ZZ} \varsigma$ approaches positive infinity. The limit at infinity has the smaller squared Sharpe ratio whenever:

$$\left[\gamma_0^T \Sigma_{FF}^{-1} \Sigma_{PP} \phi\right]^2 < \left[\phi^T \Sigma_{PP} \phi\right]^2 + \left[\phi^T \Sigma_{PP} \phi\right]\left[\varepsilon^T \Sigma_{ZZ} \varepsilon\right] \tag{7.166}$$

One verifies immediately that this condition is satisfied whenever Equation 7.164 is satisfied, but never when Equations 7.163 or 7.165 is satisfied. Substituting the values for ϕ and ε into Equation 7.163, we find:

$$\left(3\mu_Z - 4\beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1} \left(3\mu_Z - 4\beta^T \gamma_0\right) \leq \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z \tag{7.167}$$

The values of γ_0 that satisfy this inequality correspond to the fifth region in the theorem statement. In this area, the minimum value of $S^2\left(w\left(c,\zeta\right)^T Z\right)$ occurs at $\varsigma^T \Sigma_{ZZ} \varsigma = 0$, and has the value of $S^2\left(w\left(c,0\right)^T Z\right)$ shown above. This demonstrates the statements concerning the fifth region. Similarly, we can substitute in the values of ϕ and ε into 7.164:

$$\gamma_0^T \beta \Sigma_{ZZ}^{-1} \mu_Z \geq \mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z \tag{7.168}$$

The set of all such γ_0^T corresponds to the first region; the minimum value of $S^2\left(w\left(c,\zeta\right)^T Z\right)$ occurs at $\varsigma^T \Sigma_{ZZ} \varsigma = 0$. This demonstrates the statements about the first region, leaving only the third region. Substituting in the values of ϕ and ε into Equation 7.164, we find:

$$\left(3\mu_Z - 4\beta^T \gamma_0\right)^T \Sigma_{ZZ}^{-1} \left(3\mu_Z - 4\beta^T \gamma_0\right) \geq 9\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z - 8\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z \tag{7.169}$$

$$\mu_Z^T \Gamma \Sigma_{PP}^{-1} \Gamma^T \mu_Z \geq \mu_Z^T \Sigma_{ZZ}^{-1} \beta^T \gamma_0 \tag{7.170}$$

Equation 7.164 therefore describes the third region. We have already noted that the minimum value of $S^2\left(w\left(c,\zeta\right)^T Z\right)$ in this region occurs when $\varsigma^T \Sigma_{ZZ} \varsigma$ approaches positive infinity, and the value is given above. All statements have now been demonstrated.QED

7.14 Proof of Theorem 11

The task at hand is to choose Ψ so that:

$$\mu_0 = \left(\Psi \Psi^T\right)^{-1} \Psi \Sigma_{\eta\eta} \Sigma_{PP}^{-1} \Gamma^T \mu_Z \tag{7.171}$$

But recall:

$$\Sigma_{\eta\eta} = \Psi^T \Sigma_{YY} \Psi \tag{7.172}$$

Equation 7.171 therefore simplifies to:

$$\begin{aligned} \mu_0 &= \left(\Psi\Psi^T\right)^{-1} \Psi \left(\Psi^T \Sigma_{YY} \Psi\right) \Sigma_{PP}^{-1} \Gamma^T \mu_Z = \Sigma_{YY} \Psi \Sigma_{PP}^{-1} \Gamma^T \mu_Z && (7.173) \\ \Sigma_{YY}^{-1} \mu_0 &= \Psi \Sigma_{PP}^{-1} \Gamma^T \mu_Z && (7.174) \end{aligned}$$

The left-hand side of the last equation is a K-vector, and the right-hand side is a $K \times N$ matrix multiplied by an N-vector. Since $K \leq N$, we can choose a full-rank matrix Ψ that satisfies the equation.QED

7.15 Proof of Theorem 12

Under the assumptions of the theorem, the sample mean $\widehat{\mu}_Z$ of the excess return vector Z is a consistent estimator with asymptotic variance Σ_{ZZ}. The sample estimate is:

$$\widehat{\gamma}\left(P, Z\right) = \widehat{\Gamma}^T \widehat{\mu}_Z \tag{7.175}$$

Note, however, that $\widehat{\Gamma} = \Gamma$; the factors P are spanned, and their projection onto the assets can therefore be estimated without error. The asymptotic variance of the estimator is therefore:

$$AVar\left[\widehat{\gamma}\left(P, Z\right)\right] = AVar\left[\Gamma^T \widehat{\mu}_Z\right] = \Gamma^T AVar\left[\widehat{\mu}_Z\right] \Gamma = \Gamma^T \Sigma_{ZZ} \Gamma = \Sigma_{PP} \tag{7.176}$$

QED

7.16 Proof of Theorem 13

The estimate is given by:

$$\widehat{\gamma}\left(P, Z\right) = \widehat{\Gamma}^T \widehat{\mu}_Z = \widehat{\Sigma}_{FZ} \widehat{\Sigma}_{ZZ}^{-1} \widehat{\mu}_Z \tag{7.177}$$

A first-order Taylor expansion is given by:

$$\widehat{\gamma}\left(P, Z\right) - \gamma\left(P, Z\right) \approx \left(\widehat{\Sigma}_{FZ} - \Sigma_{FZ}\right) \Sigma_{ZZ}^{-1} \mu_Z - \Gamma^T \left(\widehat{\Sigma}_{ZZ} - \Sigma_{ZZ}\right) \Sigma_{ZZ}^{-1} \mu_Z + \Gamma^T \left(\widehat{\mu}_Z - \mu_Z\right) \tag{7.178}$$

We denote the three terms on the right-hand side as $\widehat{\gamma}_1$, $\widehat{\gamma}_2$ and $\widehat{\gamma}_3$. The asymptotic covariance between each pair of terms can be found element by element as follows:

$$\begin{aligned} ACovar\left(\left[\widehat{\gamma}_1\right]_{i1}, \left[\widehat{\gamma}_1\right]_{j1}\right) &= ACovar\left(\begin{array}{c} \sum_{k=1}^{M} \left[\widehat{\Sigma}_{FZ} - \Sigma_{FZ}\right]_{ik} \left[\Sigma_{ZZ}^{-1} \mu_Z\right]_{k1}, \\ \sum_{l=1}^{M} \left[\widehat{\Sigma}_{FZ} - \Sigma_{FZ}\right]_{jl} \left[\Sigma_{ZZ}^{-1} \mu_Z\right]_{l1} \end{array}\right) && (7.179) \\ &= \sum_{k=1}^{M} \sum_{l=1}^{M} \left[\Sigma_{ZZ}^{-1} \mu_Z\right]_{k1} \left[\Sigma_{ZZ}^{-1} \mu_Z\right]_{l1} ACovar\left(\begin{array}{c} \left[\widehat{\Sigma}_{FZ} - \Sigma_{FZ}\right]_{ik}, \\ \left[\widehat{\Sigma}_{FZ} - \Sigma_{FZ}\right]_{jl} \end{array}\right) && (7.180) \\ &= \sum_{k=1}^{M} \sum_{l=1}^{M} \left[\Sigma_{ZZ}^{-1} \mu_Z\right]_{k1} \left[\Sigma_{ZZ}^{-1} \mu_Z\right]_{l1} \left(\begin{array}{c} \left[\Sigma_{FF}\right]_{ij} \left[\Sigma_{ZZ}\right]_{kl} + \\ \left[\Sigma_{FZ}\right]_{il} \left[\Sigma_{ZF}\right]_{kj} \end{array}\right) && (7.181) \\ &= \left[\Sigma_{FF}\right]_{ij} \left[\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right] + \left[\Gamma^T \mu_Z\right]_{i1} \left[\Gamma^T \mu_Z\right]_{j1} && (7.182) \end{aligned}$$

Proceeding similarly with the other pairs, we find:

$$\begin{aligned}
ACovar\left([\widehat{\gamma}_1]_{i1}, [\widehat{\gamma}_2]_{j1}\right) &= [\Sigma_{PP}]_{ij}\left[\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right] + \left[\Gamma^T \mu_Z\right]_{i1}\left[\Gamma^T \mu_Z\right]_{j1} && (7.183)\\
ACovar\left([\widehat{\gamma}_1]_{i1}, [\widehat{\gamma}_3]_{j1}\right) &= 0 && (7.184)\\
ACovar\left([\widehat{\gamma}_2]_{i1}, [\widehat{\gamma}_2]_{j1}\right) &= [\Sigma_{PP}]_{ij}\left[\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right] + \left[\Gamma^T \mu_Z\right]_{i1}\left[\Gamma^T \mu_Z\right]_{j1} && (7.185)\\
ACovar\left([\widehat{\gamma}_2]_{i1}, [\widehat{\gamma}_3]_{j1}\right) &= 0 && (7.186)\\
ACovar\left([\widehat{\gamma}_3]_{i1}, [\widehat{\gamma}_3]_{j1}\right) &= [\Sigma_{PP}]_{ij} && (7.187)
\end{aligned}$$

Putting together the above results, we find:

$$ACovar\left([\widehat{\gamma}(P,Z)]_i, [\widehat{\gamma}(P,Z)]_j\right) = [\Sigma_{PP}]_{ij} + [\Sigma_{\eta\eta}]_{ij}\left[\mu_Z^T \Sigma_{ZZ}^{-1} \mu_Z\right] \tag{7.188}$$

The desired result follows immediately.QED

7.17 Proof of Theorem 14

The estimator is:

$$\widehat{\gamma}(F,Z) = \widehat{\Sigma}_{FF}\widehat{\Sigma}_{PP}^{-1}\widehat{\Gamma}^T\widehat{\mu}_Z = \widehat{\Sigma}_{FF}\left(\widehat{\Sigma}_{FZ}\widehat{\Sigma}_{ZZ}^{-1}\widehat{\Sigma}_{ZF}\right)^{-1}\widehat{\Sigma}_{FZ}\widehat{\Sigma}_{ZZ}^{-1}\widehat{\mu}_Z \tag{7.189}$$

A first order Taylor expansion is:

$$\begin{aligned}
\widehat{\gamma}(F,Z) - \gamma(F,Z) \approx\; & \left(\widehat{\Sigma}_{FF} - \Sigma_{FF}\right)\Sigma_{PP}^{-1}\Gamma^T\mu_Z && (7.190)\\
& -\Sigma_{FF}\Sigma_{PP}^{-1}\left(\widehat{\Sigma}_{FZ} - \Sigma_{FZ}\right)\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z && (7.191)\\
& +\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\left(\widehat{\Sigma}_{ZZ} - \Sigma_{ZZ}\right)\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z && (7.192)\\
& -\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\left(\widehat{\Sigma}_{ZF} - \Sigma_{ZF}\right)\Sigma_{PP}^{-1}\Gamma^T\mu_Z && (7.193)\\
& +\Sigma_{FF}\Sigma_{PP}^{-1}\left(\widehat{\Sigma}_{FZ} - \Sigma_{FZ}\right)\Sigma_{ZZ}^{-1}\mu_Z && (7.194)\\
& -\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\left(\widehat{\Sigma}_{ZZ} - \Sigma_{ZZ}\right)\Sigma_{ZZ}^{-1}\mu_Z && (7.195)\\
& +\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\left(\widehat{\mu}_Z - \mu_Z\right) && (7.196)
\end{aligned}$$

However, since F is a linear factor model for Z, we have $\mu_Z = \beta^T\gamma(F,Z)$. Making this substitution, four of the seven terms on the right-hand side cancel:

$$\begin{aligned}
\widehat{\gamma}(F,Z) - \gamma(F,Z) \approx\; & \left(\widehat{\Sigma}_{FF} - \Sigma_{FF}\right)\Sigma_{FF}^{-1}\gamma(F,Z) && (7.197)\\
& -\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\left(\widehat{\Sigma}_{ZF} - \Sigma_{ZF}\right)\Sigma_{FF}^{-1}\gamma(F,Z) && (7.198)\\
& +\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\left(\widehat{\mu}_Z - \mu_Z\right) && (7.199)
\end{aligned}$$

Denoting the terms on the right-hand side by $\widehat{\gamma}_1$ through $\widehat{\gamma}_3$, and applying the technique used for the previous theorem, the asymptotic covariances between each pair of terms is given by:

$$ACovar\left([\widehat{\gamma}_1]_i, [\widehat{\gamma}_1]_j\right) = [\Sigma_{FF}]_{ij}\left[\gamma^T(F,Z)\,\Sigma_{FF}^{-1}\gamma(F,Z)\right] + [\gamma(F,Z)]_{i1}\,[\gamma(F,Z)]_{j1} \tag{7.200}$$

$$ACovar\left([\widehat{\gamma}_1]_i, [\widehat{\gamma}_2]_j\right) = [\Sigma_{FF}]_{ij}\left[\gamma^T(F,Z)\,\Sigma_{FF}^{-1}\gamma(F,Z)\right] + [\gamma(F,Z)]_{i1}\,[\gamma(F,Z)]_{j1} \tag{7.201}$$

$$ACovar\left([\widehat{\gamma}_1]_i, [\widehat{\gamma}_3]_j\right) = 0 \tag{7.202}$$

$$ACovar\left([\widehat{\gamma}_2]_i, [\widehat{\gamma}_2]_j\right) = \left[\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\right]_{ij}\left[\gamma^T(F,Z)\,\Sigma_{FF}^{-1}\gamma(F,Z)\right] + [\gamma(F,Z)]_{i1}\,[\gamma(F,Z)]_{j1} \tag{7.203}$$

$$ACovar\left([\widehat{\gamma}_2]_i, [\widehat{\gamma}_3]_j\right) = 0 \tag{7.204}$$

$$ACovar\left([\widehat{\gamma}_3]_i, [\widehat{\gamma}_3]_j\right) = \left[\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\right]_{ij} \tag{7.205}$$

Adding up all the terms, we find:

$$ACovar\left([\widehat{\gamma}(F,Z)]_i, [\widehat{\gamma}(F,Z)]_j\right) = [\Sigma_{FF}]_{ij} + \tag{7.206}$$

$$\left[\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF} - \Sigma_{FF}\right]_{ij}\left[1 + \gamma^T(F,Z)\,\Sigma_{FF}^{-1}\gamma(F,Z)\right] \tag{7.207}$$

$$= [\Sigma_{FF}]_{ij} + \tag{7.208}$$

$$\left[\Sigma_{\eta\eta} + \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right]_{ij}\left[1 + \gamma^T(F,Z)\,\Sigma_{FF}^{-1}\gamma(F,Z)\right] \tag{7.209}$$

This last expression yields the first desired result. To express this asymptotic covariance in terms of μ_Z instead of $\gamma(F,Z)$, we substitute in $\gamma(F,Z) = \Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z$:

$$ACovar\left([\widehat{\gamma}(F,Z)]_i, [\widehat{\gamma}(F,Z)]_j\right) = [\Sigma_{FF}]_{ij} + \tag{7.210}$$

$$\left[\Sigma_{\eta\eta} + \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right]_{ij}\left[1 + \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right] \tag{7.211}$$

$$= [\Sigma_{FF}]_{ij} + \tag{7.212}$$

$$\left[\Sigma_{\eta\eta} + \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right]_{ij}\left[\begin{array}{c} 1 + \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z + \\ \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Gamma^T\mu_Z \end{array}\right] \tag{7.213}$$

QED

7.18 Proof of Theorem 15

The first order Taylor expansion can be expressed as:

$$\widehat{\gamma}(F,Z) - \gamma(F,Z) \approx \left(\widehat{\Sigma}_{FF} - \Sigma_{FF}\right)\Sigma_{PP}^{-1}\Gamma^T\mu_Z \tag{7.214}$$

$$+\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\left(\widehat{\Sigma}_{ZZ} - \Sigma_{ZZ}\right)\left(\Gamma\Sigma_{PP}^{-1}\Gamma^T - \Sigma_{ZZ}^{-1}\right)\mu_Z \tag{7.215}$$

$$-\Sigma_{FF}\Sigma_{PP}^{-1}\left(\widehat{\Sigma}_{FZ} - \Sigma_{FZ}\right)\left(\Gamma\Sigma_{PP}^{-1}\Gamma^T - \Sigma_{ZZ}^{-1}\right)\mu_Z \tag{7.216}$$

$$-\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\left(\widehat{\Sigma}_{ZF} - \Sigma_{ZF}\right)\Sigma_{PP}^{-1}\Gamma^T\mu_Z \tag{7.217}$$

$$+\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\left(\widehat{\mu}_Z - \mu_Z\right) \tag{7.218}$$

Denoting the terms on the right-hand side by $\widehat{\gamma}_1$ through $\widehat{\gamma}_5$, and applying the technique used for the previous theorems, the asymptotic covariances between each pair of terms is given by:

$$ACovar\left([\widehat{\gamma}_1]_i\,,[\widehat{\gamma}_1]_j\right) = [\Sigma_{FF}]_{ij}\left[\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right] \tag{7.219}$$

$$+\left[\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right]_{i1}\left[\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right]_{j1} \tag{7.220}$$

$$ACovar\left([\widehat{\gamma}_1]_i\,,[\widehat{\gamma}_2]_j\right) = 0 \tag{7.221}$$

$$ACovar\left([\widehat{\gamma}_1]_i\,,[\widehat{\gamma}_3]_j\right) = 0 \tag{7.222}$$

$$ACovar\left([\widehat{\gamma}_1]_i\,,[\widehat{\gamma}_4]_j\right) = [\Sigma_{FF}]_{ij}\left[\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right] \tag{7.223}$$

$$+\left[\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right]_{i1}\left[\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right]_{j1} \tag{7.224}$$

$$ACovar\left([\widehat{\gamma}_1]_i\,,[\widehat{\gamma}_5]_j\right) = 0 \tag{7.225}$$

$$ACovar\left([\widehat{\gamma}_2]_i\,,[\widehat{\gamma}_2]_j\right) = \left[\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\right]_{ij}\left[\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z - \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right] \tag{7.226}$$

$$ACovar\left([\widehat{\gamma}_2]_i\,,[\widehat{\gamma}_3]_j\right) = \left[\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\right]_{ij}\left[\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z - \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right] \tag{7.227}$$

$$ACovar\left([\widehat{\gamma}_2]_i\,,[\widehat{\gamma}_4]_j\right) = 0 \tag{7.228}$$

$$ACovar\left([\widehat{\gamma}_2]_i\,,[\widehat{\gamma}_5]_j\right) = 0 \tag{7.229}$$

$$ACovar\left([\widehat{\gamma}_3]_i\,,[\widehat{\gamma}_3]_j\right) = \left[\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\right]_{ij}\left[\mu_Z^T\left(\Sigma_{ZZ}^{-1} - \Gamma\Sigma_{PP}^{-1}\Gamma^T\right)\mu_Z\right] \tag{7.230}$$

$$ACovar\left([\widehat{\gamma}_3]_i\,,[\widehat{\gamma}_4]_j\right) = 0 \tag{7.231}$$

$$ACovar\left([\widehat{\gamma}_3]_i\,,[\widehat{\gamma}_5]_j\right) = 0 \tag{7.232}$$

$$ACovar\left([\widehat{\gamma}_4]_i\,,[\widehat{\gamma}_4]_j\right) = \left[\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\right]_{ij}\left(\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right) \tag{7.233}$$

$$+\left[\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right]_{i1}\left[\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right]_{j1} \tag{7.234}$$

$$ACovar\left([\widehat{\gamma}_4]_i\,,[\widehat{\gamma}_5]_j\right) = 0 \tag{7.235}$$

$$ACovar\left([\widehat{\gamma}_5]_i\,,[\widehat{\gamma}_5]_j\right) = \left[\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\right]_{ij} \tag{7.236}$$

Summing up the terms, we find:

$$ACovar\left([\widehat{\gamma}\,(F,Z)]_i\,,[\widehat{\gamma}\,(F,Z)]_j\right) = -\,[\Sigma_{FF}]_{ij}\left[\mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{FF}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right] \tag{7.237}$$

$$+\left[\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\right]_{ij}\left[1 - \mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z + \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z\right] \tag{7.238}$$

$$+\left[\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\Sigma_{PP}^{-1}\Sigma_{FF}\right]_{ij}\left[\begin{array}{c}\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z - \\ \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z\end{array}\right] \tag{7.239}$$

$$= [\Sigma_{FF}]_{ij} \tag{7.240}$$

$$+\left[\Sigma_{\eta\eta} + \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\right]_{ij}\left[\begin{array}{c}1 + \mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z + \\ \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Gamma^T\mu_Z\end{array}\right] \tag{7.241}$$

$$+\left[\begin{array}{c}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta} + \\ \Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\Sigma_{PP}^{-1}\Sigma_{\eta\eta}\end{array}\right]_{ij}\left[\begin{array}{c}\mu_Z^T\Sigma_{ZZ}^{-1}\mu_Z - \\ \mu_Z^T\Gamma\Sigma_{PP}^{-1}\Gamma^T\mu_Z\end{array}\right] \tag{7.242}$$

The last expression yields the desired result.QED