Identification of Factor Risk Premia*

Peter G. Hansen[†]
New York University, 19 West 4th St, New York, NY 10012, USA
peter.hansen@nyu.edu

Maziar M. Kazemi
100 Main St, Cambridge, MA 02141
kazemi@mit.edu

October 1, 2021

Abstract

This paper a develops novel statistical test of whether individual factor risk premia are identified from return data in multi-factor models. We give a necessary and sufficient condition for population identification of individual risk premia, which we call the *kernel-orthogonality condition*. This condition is weaker than the standard rank condition commonly assumed for linear factor models. Under misspecification, our condition ensures point identification of the risk premium with minimal pricing error. We show how to test this restriction directly in reduced-rank models. Finally, we apply our test methodology to assess identification of risk premia associated with consumption growth and intermediary leverage.

Keywords: Linear factor models, Underidentification test, Risk premia

JEL Codes: G12, C12, C58

*The authors gratefully acknowledge helpful comments and suggestions from Hui Chen, Sharada Dharmasankar, Lars Hansen, Leonid Kogan, Andrew Lo, Andrey Malenko, Maarten Meeuwis, Per Mykland, Jonathan Parker, Stephen Penman (discussant), Pari Sastry, Grace Tsiang, and Xiangyu Zhang, and participants at the MIT Sloan Doctoral Research Forum, the SoFiE 2021 Annual Conference, and the 2021 World Finance Conference.

†Corresponding author.

1 Introduction

Since the seminal work of Sharpe (1964) and Lintner (1965) on the CAPM, linear factor models have become ubiquitous in empirical finance. Notable examples include Chen et al. (1986), Fama and French (1993), Jagannathan and Wang (1996), Carhart (1997), Lettau and Ludvigson (2001), Lustig and Verdelhan (2007), Adrian et al. (2014), and He et al. (2017). These models posit a stochastic discount factor (SDF) that is linear in risk factors. These risk factors can be either returns on traded securities, macroeconomic variables, or a combination of the two.

Rather than converge on a single "correct" model, the literature has instead become saturated with numerous proposed factors which have been claimed to be statistically important in explaining cross-sectional variation in stock returns. Harvey et al. (2016) document 316 such factors. One possible reason for this excessive assignment of statistically significant risk factors is a lack of identification. Conventional methods, such as cross-sectional regression, Fama-MacBeth regressions, or GMM, for estimating factor risk premia rely on the true factor risk premia being point identified. This identification condition is equivalent to a rank condition on the matrix of betas. If some factors are not correlated with returns, or are redundant for explaining cross-sectional variation, then the standard estimation and inference techniques become unreliable. Estimated risk premia can have substantially different distributions than what would be predicted by standard asymptotics assuming point identification, leading researchers to conclude that risk premia are statistically significant even in cases where the population risk premium is not well-defined. This is illustrated in figure 1 where the the simulated distribution of t-statistics for the risk premium of a noisy and redundant factor is clearly far from a standard normal distribution.

This paper proposes diagnostic procedures to detect whether individual factor risk premia are in fact identified from return data. We illustrate how a regularity condition, which we call the *kernel-orthogonality condition*, is necessary and sufficient for a given factor risk premia to be identified. This condition is weaker than the standard rank condition, precisely because individual factor risk premia can be identified even when the full vector of factor risk premia is not. We show that this condition is mathematically equivalent to the existence of a "true" (population) factor mimicking portfolio. Even if the linear factor model is misspecified, this condition remains informative about the identification of the factor risk premium consistent with minimal pricing error. We describe statistical tests

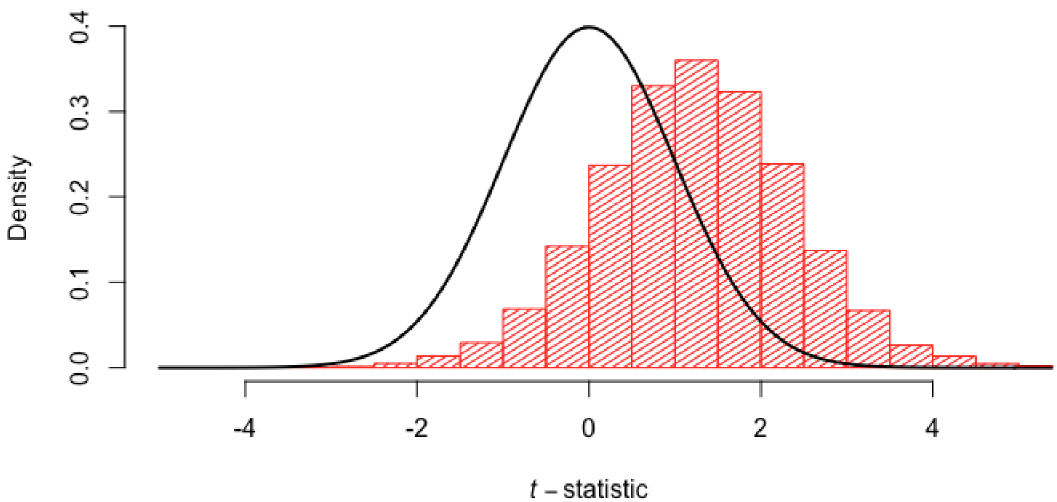

Figure 1: Simulated distribution of t-statistics for estimated risk premium of a noisy, redundant factor, using sample size of $T = 50$ observations and $n = 25$ test assets based on cross-sectional regression. Simulated distribution shown in **red**. Standard normal density (predicted by standard asymptotics under point identification) shown in **black**. t-statistics computed using Shanken (1992) correction. Rejection probabilities given in table D1. Further simulation details given in appendix D.

which can together determine the dimension of rank deficiency in the matrix of betas, as well as whether individual factor risk premia are identified. Finally, we apply our testing to two linear factor specifications in the literature. The first specification, due to Lustig and Verdelhan (2007), estimates the risk premium of consumption growth based on portfolios of currency returns. The second specification, due to Adrian et al. (2014), estimates the risk premium associated with the leverage of financial intermediaries based on equity excess returns. In both cases we are unable to reject the null hypothesis of reduced rank, giving a key role for our testing methodology to determine whether individual factor risk premia are identified.

1.1 Related Literature

Our paper contributes to a significant econometrics literature which analyzes the cross-sectional regression procedure introduced by Black et al. (1972) and Fama and MacBeth

(1973) for the analysis of linear factor models. The closest papers to ours focus on identification in linear factor models. Burnside (2016) applies the rank tests of Cragg and Donald (1997) and Kleibergen and Paap (2006) to test the standard rank condition for identification in linear factor models. Our testing methodology is closely related to Arellano et al. (2012), Manresa et al. (2017), and Windmeijer et al. (2018), who develop and apply tests of underidentification and discuss estimation of the direction of underidentification in linear factor models. Unlike this existing literature, our methods can be useful to infer whether an individual factor risk premium is identified even when the null of rank deficiency is not rejected.

Kleibergen (2009) and Kleibergen and Zhan (2018) propose a tests of factor risk premia that are valid under "weak identification" asymptotics, i.e. model specifications in which the matrix of factor betas drifts towards rank deficiency as the sample size becomes large. This approach differs from ours in two ways. First, under "weak identification" asymptotics point identification only fails in the limit, while our asymptotics allow for general identification failure in finite samples. Second, while these papers construct tests of whether a risk premium has a particular value (for instance zero), our methodology instead tests whether a given factor risk premium is identified in a population sense. We view this approach to be distinct but complementary to our own.

Several papers propose estimation methods to deal with identification issues caused by spurious factors which are uncorrelated with asset returns. These papers typically resolve the corresponding identification failure in a subjective or ad-hoc manner, for instance by automatically setting the corresponding risk premia to zero. Examples include Gospodinov et al. (2014), Bryzgalova (2015), Giglio and Xiu (2021), Kazemi (2019), and Kozak et al. (2020). We view these approaches as having the disadvantage that they do not distinguish between cases in which a factor risk premium is not identified and cases in which a factor risk premium is point identified and (close to) zero, whereas our test procedure is able to distinguish between these two cases.

1.2 Outline

The outline of the paper is as follows. Section 2 studies population identification of factor risk premia in linear factor models. We formally state the kernel-orthogonality condition as well as it's equivalence to the point identification of an individual factor risk premium. We show that this condition is equivalent to the existence of a factor mimicking portfolio. Additionally, we demonstrate that the condition is meaningful under model misspecification as necessary and sufficient for point identification of the risk premium consistent with minimal pricing error. Section 3 develops a novel test procedure which can determine whether individual factor risk premia are identified. We give asymptotic distribution theory for the test statistic, and describe an asymptotically conservative test which has asymptotically exact size in the case of one-dimension rank deficiency. Section 4 applies our test procedures to several well-known linear factor specifications in the literature. Section 5 concludes.

2 Identification in Linear Factor Models

We begin by reviewing the standard theory of linear factor models. Consider the standard asset pricing equation for expected returns

$$\mathbb{E}[m\mathbf{r}^e] = \mathbf{0} \in \mathbb{R}^n \tag{1}$$

where the random variable m is the (scalar) stochastic discount factor and $\mathbf{r}^e$ is an n-dimensional vector of excess returns. We consider factor models where m lies in the affine span of a random vector $\mathbf{f} \in \mathbb{R}^k$ of mean-zero factors.

Assumption 2.1. *The stochastic discount factor can be represented as $m = R_f^{-1}(1 - \mathbf{f}'\boldsymbol{\theta})$ where $\mathbb{E}[\mathbf{f}] = \mathbf{0} \in \mathbb{R}^k$.*

We think of the gross risk-free rate R_f as known or exogenously imposed by the econometrician. We show in appendix B how the assumption of a constant risk-free rate can be relaxed without changing the implications. By construction,

$$\mathbb{E}[mR_f] = 1. \tag{2}$$

The no arbitrage condition (1) can be expressed as

$$\mathbb{E}\left[-\frac{1}{R_f}\mathbf{r}^e[\mathbf{f}' \;\; 1]\right]\left[\begin{array}{c}\boldsymbol{\theta} \\ -1\end{array}\right] = \mathbf{0}. \tag{3}$$

Observe that since R_f is a non-zero scalar, $-R_f^{-1}$ can be factored out of equation (3). For notational simplicity, we introduce the following definition:

Definition 2.2. *Define a random matrix* $\mathbf{\Psi} \in \mathbb{R}^{n\times(k+1)}$ *as*

$$\mathbf{\Psi} \doteq \mathbf{r}^e[\mathbf{f}' \ 1]. \tag{4}$$

and $\overline{\mathbf{\Psi}} = \mathbb{E}[\mathbf{\Psi}]$.

Equipped with this definition, equation (3) can be expressed concisely as

$$\overline{\mathbf{\Psi}} \begin{bmatrix} \boldsymbol{\theta} \\ -1 \end{bmatrix} = \mathbf{0}. \tag{5}$$

For some of our results, we will maintain the following assumption:

Assumption 2.3. *Equation (5) admits at least one solution* $\boldsymbol{\theta} \in \mathbb{R}^k$.

Assumption 2.3 ensures that the model is correctly specified, which guarantees that the identified set of risk-price vectors is non-empty. It does not, however, rule out cases where risk-prices may not be unique.

It is useful to partition the random matrix $\mathbf{\Psi}$ as $\mathbf{\Psi} = [\mathbf{\Psi}^{(1)} \ \mathbf{\Psi}^{(2)}]$ where

$$\mathbf{\Psi}^{(1)} \doteq \mathbf{r}^e\mathbf{f}' \in \mathbb{R}^{n\times k}$$
$$\mathbf{\Psi}^{(2)} \doteq \mathbf{r}^e \in \mathbb{R}^{n}.$$

Further, we define $\overline{\mathbf{\Psi}}^{(1)} \doteq \mathbb{E}\left[\mathbf{\Psi}^{(1)}\right]$ and $\boldsymbol{\mu} \doteq \mathbb{E}\left[\mathbf{\Psi}^{(2)}\right] = \mathbb{E}\left[\mathbf{r}^e\right]$. Under this notation, equation (5) can be equivalently expressed as

$$\overline{\mathbf{\Psi}}^{(1)}\boldsymbol{\theta} = \boldsymbol{\mu}. \tag{6}$$

Remark 2.4. *Under correct specification (i.e. if assumption 2.3 holds),* $\operatorname{rank}(\overline{\mathbf{\Psi}}) = \operatorname{rank}(\overline{\mathbf{\Psi}}^{(1)})$. *Furthermore,* $\operatorname{rank}(\overline{\mathbf{\Psi}}) = \operatorname{rank}(\overline{\mathbf{\Psi}}^{(1)}) = k$ *is a necessary and sufficient condition for the vector* $\boldsymbol{\theta}$ *to be point identified.*

Equation (6) is equivalent to the standard β-representation of expected returns. To see this, it is helpful to impose the following definition:

Definition 2.5. *Given* $\boldsymbol{\theta} \in \mathbb{R}^k$ *which satisfies equation (5), we define the associated vector of factor risk premia* $\boldsymbol{\lambda} \doteq \mathbf{\Sigma}_f\boldsymbol{\theta}$, *where* $\mathbf{\Sigma}_f \doteq \mathbb{E}[\mathbf{f}\ \mathbf{f}']$.

Throughout, we maintain the following assumption.

Assumption 2.6. $\mathbf{\Sigma}_f$ *is nonsingular.*

Note that assumption 2.6 does not rule out the possibility that factors are either uncorrelated with returns, or redundant for pricing. This can be seen in examples 2.9 and 2.10.

We can write the stochastic discount factor as $m = R_f^{-1}(1 - \boldsymbol{f}'\mathbf{\Sigma}^{-1}\boldsymbol{\lambda})$. Equation (6) can be re-written equivalently as

$$\boldsymbol{\mu} = \overline{\mathbf{\Psi}}^{(1)}\mathbf{\Sigma}_f^{-1}\boldsymbol{\lambda} \tag{7}$$

which is the standard beta representation of expected returns. To see this, observe that $\mathbf{B} \doteq \overline{\mathbf{\Psi}}^{(1)}\mathbf{\Sigma}_f^{-1}$ is a matrix of regression coefficients, so we have

$$\boldsymbol{\mu} = \mathbf{B}\boldsymbol{\lambda} \tag{8}$$

or in component form

$$\mathbb{E}[\mathbf{r}_i^e] = \sum_{j=1}^{k} \mathbf{B}_{i,j}\boldsymbol{\lambda}_j. \tag{9}$$

If the vector $\boldsymbol{\lambda}$ is point identified, then Fama and MacBeth (1973) style methods will consistently estimate the vector $\boldsymbol{\lambda}$. Without point identification, such methods can exhibit strange behavior (Bryzgalova (2015)) and GMM specification tests of the linear factor model can give highly misleading results (Gospodinov et al. (2017)).

Remark 2.7. *In our analysis, we follow the standard convention in the empirical asset pricing literature which defines factor risk premia for multi-factor models in terms of multivariate regression betas. Several authors instead define factor risk premia relative to univariate regression betas, and advocate for this alternate convention. See for instance Chen et al. (1986), Jagannathan and Wang (1996), and Jagannathan and Wang (1998). It is straightforward to extend our analysis to obtain analogous results under this alternative convention. Moreover, it would lead to significant econometric simplifications.*

2.1 Identification of Factor Risk Premia

We are interested in characterizing the identification properties of the vector of factor risk premia $\boldsymbol{\lambda}$ as well as the identification of its components $\boldsymbol{\lambda}_i$. We begin with a definition.

Definition 2.8. *Define the sets* I_Λ *and* I_Λ^i

$$
\begin{aligned}
I_\Lambda &\doteq \left\{\boldsymbol{\lambda} \in \mathbb{R}^k : \mathbf{B}\boldsymbol{\lambda} = \boldsymbol{\mu}\right\} \\
I_\Lambda^i &\doteq \left\{l \in \mathbb{R} : \exists \boldsymbol{\lambda} \in I_\Lambda \ s.t. \ \boldsymbol{\lambda}_i = l\right\}.
\end{aligned}
$$

We will say that $\boldsymbol{\lambda}$ *is point identified if* I_Λ *is a singleton, and that* $\boldsymbol{\lambda}_i$ *is point identified if* I_Λ^i *is a singleton.*

Observe that I_Λ^i is the projection of I_Λ onto the i-th coordinate. Since I_Λ is a hyperplane, it must be the case that I_Λ^i is either a single number or the whole real line. This is proved formally as proposition A.1 in the appendix. If assumption 2.3 holds, then both sets will be non-empty. If, however, the linear factor model is misspecified, then both sets will be empty.

Next we give a simple example in which the vector of risk premia $\boldsymbol{\lambda}$ is not point identified, but an individual component is point identified.

Example 2.9. *Let* $k = 2$ *and assume*

$$
\mathbf{B} = [\boldsymbol{\beta} \ \mathbf{0}]
$$

where the vector $\boldsymbol{\beta} \in \mathbb{R}^n$ *contains at least one non-zero element. Additionally, assume* $\mathbb{E}[\mathbf{r}^e] = \lambda^* \boldsymbol{\beta}$ *for a fixed* $\lambda^* \in \mathbb{R}$. *Then equation (7) will have infinitely many solutions so the vector* $\boldsymbol{\lambda} \in \mathbb{R}^2$ *is not point-identified. However all solutions are of the form* $\boldsymbol{\lambda} = (\lambda^*, \lambda_2)$, *so the price-of-risk* $\boldsymbol{\lambda}_1 = \lambda^*$ *of the first factor is point identified.*

Of course it can also be the case that the vector of risk premia $\boldsymbol{\lambda}$ is not point identified, and none of its components are point identified. We illustrate this in the next example.

Example 2.10. *Suppose there is a true latent one-factor model in* f^* *where* $\mathbb{E}[f^*] = 0$, $\operatorname{var}(f^*) = 1$ *so that*

$$
\mathbb{E}[\mathbf{r}^e] = \lambda^* \boldsymbol{\beta}
$$

where $\lambda^* \neq 0$, $\boldsymbol{\beta} = \operatorname{cov}(\mathbf{r}^e, f^*)$. *Suppose further that* $n \geqslant 2$ *and that* $\boldsymbol{\beta}$ *has at least two non-zero elements which are not equal. The econometrician does not observe* f^* *directly, but instead observes two noisy proxies* $\mathbf{f} = (\mathbf{f}_1, \mathbf{f}_2)$ *where*

$$
\mathbf{f}_i = \rho_i f^* + \sqrt{1 - \rho_i^2}\ \epsilon_i
$$

with $0 < |\rho_i| < 1$, $\mathbb{E}[\epsilon_i] = 0$, $\operatorname{var}(\epsilon_i) = 1$ *and* ϵ_i *is uncorrelated with either returns,* f^**, or* ϵ_{-i}*. In this case*

$$\overline{\boldsymbol{\Psi}}^{(1)} = \left[\rho_1 \boldsymbol{\beta} \ \rho_2 \boldsymbol{\beta}\right], \ \overline{\boldsymbol{\Psi}}^{(2)} = \lambda^* \boldsymbol{\beta}, \ \boldsymbol{\Sigma}_f = \begin{bmatrix} 1 & \rho_1 \rho_2 \\ \rho_1 \rho_2 & 1 \end{bmatrix}$$

We see that $I_\Lambda = \{\boldsymbol{\lambda} \in \mathbb{R}^2 : \boldsymbol{\lambda}_1 \rho_1 + \boldsymbol{\lambda}_2 \rho_2 = \lambda^*(1 + \rho_1 \rho_2)\}$*, and that neither* $\boldsymbol{\lambda}_1$ *nor* $\boldsymbol{\lambda}_2$ *are point identified.*

Remark 2.11. *Under correct specification (i.e. if assumption 2.3 holds),* $\operatorname{rank}(\overline{\boldsymbol{\Psi}}) = \operatorname{rank}(\overline{\boldsymbol{\Psi}}^{(1)}) = k$ *is a necessary and sufficient condition for the vector* $\boldsymbol{\lambda}$ *to be point identified.*

One might be tempted to conclude that $\operatorname{rank}(\overline{\boldsymbol{\Psi}}^{(1)}) = k$ is a necessary and sufficient condition for the individual price of risk $\boldsymbol{\lambda}_i$ to be point identified. However this is incorrect. While $\overline{\boldsymbol{\Psi}}^{(1)}$ having full rank is a sufficient condition for $\boldsymbol{\lambda}_i$ to be point identified, it is not a necessary condition. We state precise necessary and sufficient conditions below.

Condition 2.12. $\ker(\mathbf{B}) \perp \mathbf{e}_i$*, where* $\mathbf{e}_i$ *is the* i*-th Euclidean basis vector.*

We refer to condition 2.12 as a *kernel-orthogonality* condition for coordinate i. This condition will play a key role both in our characterization of identification of $\boldsymbol{\lambda}_i$ (see below) as well as our subsequent identification tests.

Proposition 2.13. *Assume the model is correctly specified (i.e. assumption 2.3 holds). Then condition 2.12 is a necessary and sufficient condition for the risk premium* $\boldsymbol{\lambda}_i$ *to be point identified.*

Observe that if $\overline{\boldsymbol{\Psi}}^{(1)}$ has full rank, then $\ker(\mathbf{B}) = \{\mathbf{0}\}$, which is clearly orthogonal to $\mathbf{e}_i$. Thus point identification of individual risk prices will hold under weaker conditions than those required for point identification of the entire price of risk vector.

2.2 Factor Mimicking Portfolios

Next we discuss the connection between mimicking portfolios and the kernel-orthogonality condition. We represent a portfolio of excess returns by a vector $\boldsymbol{\phi} \in \mathbb{R}^n$ of portfolio weights. The associated portfolio return is then given by $\boldsymbol{\phi}' \mathbf{r}^e$. We call $\boldsymbol{\phi}$ a mimicking portfolio for factor $\mathbf{f}_i$ if the factor beta of the portfolio return is 1 for factor $\mathbf{f}_i$ and 0 for all factors $\mathbf{f}_{-i}$. Formally, this can be expressed as

$$\mathbf{B}' \boldsymbol{\phi} = \mathbf{e}_i. \tag{10}$$

If the factor model is correctly specified and a mimicking portfolio exists, it necessarily has expected return equal to $\boldsymbol{\lambda}_i$. This is shown in appendix C. Mimicking portfolios are generally not unique whenever one has more returns than factors. In empirical applications, researchers often select the mimicking portfolio that has minimal variance.[1] Such considerations are not relevant to our analysis.

It is in no way obvious that a mimicking portfolio for a given factor $\mathbf{f}_i$ exists. For instance, if $\mathbf{f}_i$ is uncorrelated with excess returns, then a mimicking portfolio will never exist. Alternatively, if $\mathbf{f}_i$ and $\mathbf{f}_j$, $j \neq i$ differ only by an error term which is uncorrelated with returns, then no mimicking portfolio will exist for either factor $\mathbf{f}_i$ or $\mathbf{f}_j$. It is straightforward to see that under correct specification, existence of a mimicking portfolio $\boldsymbol{\phi}$ for factor $\mathbf{f}_i$ is sufficient for point identification of $\boldsymbol{\lambda}_i$, since it is identified as $\boldsymbol{\lambda}_i = \mathbb{E}[\boldsymbol{\phi}'\mathbf{r}^e]$. The following result shows that it is in fact necessary and sufficient.

Proposition 2.14. *Condition 2.12 is a necessary and sufficient condition for the existence of a mimicking portfolio for factor* $\mathbf{f}_i$.

Proposition 2.14 follows immediately from proposition A.2 in the appendix. It shows that condition 2.12 is equivalent to the existence of a mimicking portfolio for factor $\mathbf{f}_i$. Perhaps surprisingly, we did not need to assume that the factor model was correctly specified for the equivalence in proposition 2.14 to hold.

2.3 Misspecified Models and "Pseudo-True" Identification

Previously, we characterized identification of the risk price vector $\boldsymbol{\lambda}$ in the case where the linear factor model correctly captures the expected return relationship in the economy. One may still be interested in parameter identification even if we dispense with assumption 2.3 and instead treat the linear factor model as an approximation. In this case we are interested in the identification properties of the best approximate model, sometimes referred to as the "least-misspecified" or "pseudo-true" model.

To even begin to discuss identification under model misspecification, one needs a formal measure of approximation error or misspecification. Under misspecification, the vector $\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda}$ is naturally thought of as a vector of pricing errors. One way to form a measure of

[1] This is equivalent to finding a portfolio with maximal correlation with the given factor $\mathbf{f}_i$. See for instance Huberman et al. (1987) or Breeden et al. (1989).

model misspecification is to consider a quadratic form of the vector of pricing errors, i.e.

$$c_W(\boldsymbol{\lambda}) = (\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda})' W (\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda})$$

for some positive semidefinite matrix $W \in \mathbb{R}^{n\times n}$. Subsequently, we will discuss two choices of W commonly used in the literature.

We can discuss the identified set risk premia consistent with minimizing $c_W(\boldsymbol{\lambda})$. Analogous to the correct specification setting, we define two sets:

Definition 2.15. *Define the sets*

$$\begin{aligned} I_{\Lambda,W} &\doteq \underset{\boldsymbol{\lambda}}{\operatorname{argmin}}\, (\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda})' W (\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda}) \\ I^i_{\Lambda,W} &\doteq \{l \in \mathbb{R} : \exists \boldsymbol{\lambda} \in I_{\Lambda,W} \ s.t.\ \boldsymbol{\lambda}_i = l\} \end{aligned}$$

We will say that $\boldsymbol{\lambda}$ is pseudo-point identified if $I_{\Lambda,W}$ is a singleton and $\boldsymbol{\lambda}_i$ is pseudo-point identified if $I^i_{\Lambda,W}$ is a singleton.

The next proposition shows that even under model misspecification, condition 2.12 plays a key role in the (pseudo) identification of individual factor risk premia.

Proposition 2.16. *For any positive semidefinite matrix $W \in \mathbb{R}^{n\times n}$, condition 2.12 is a necessary condition for $I^i_{\Lambda,W}$ to be a singleton. Furthermore, if W is positive definite, then condition 2.12 is also a sufficient for $I^i_{\Lambda,W}$ to be a singleton.*

The proof of proposition 2.16 is given in the appendix.

We consider measures of misspecification corresponding to two special cases of W. First, Fama-MacBeth methods implicitly estimate the vector of risk premia which minimize the average squared pricing error, which corresponds to $W = I_{n\times n}$. To see this, observe that the population counterpart to the second-stage Fama-MacBeth regression is given by

$$\min_{\boldsymbol{\lambda}} (\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda})^2 = \min_{\boldsymbol{\lambda}} (\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda})' (\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda}).$$

Since the identity matrix is strictly positive definite, we see that condition 2.12 is both necessary and sufficient for the pseudo-identified set $I_{\Lambda,I_{n\times n}}$ associated with the Fama-MacBeth procedure to be a singleton.

Next, we consider the first Hansen-Jagannathan (HJ) distance, a measure of misspecification proposed by Hansen and Jagannathan (1997) and widely used in the empirical asset pricing literature (See for instance Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Kan and Robotti (2009), Lewellen et al. (2010)).

In our context, the HJ distance δ associated with an approximate stochastic discount factor $\widehat{m}$ is defined as the solution to the optimization problem

$$\begin{aligned}\delta(\widehat{m}) \doteq \min_{m \in L^2} \ & \left(\mathbb{E}\left[(m-\widehat{m})^2\right]\right)^{\frac{1}{2}} \\ \text{s.t. } & \mathbb{E}[m\mathbf{r}^e] = \mathbf{0} \\ & \mathbb{E}[mR_f] = 1\end{aligned}$$

for a given gross risk-free rate R_f. The minimization problem will have a unique minimum as long as $\widehat{m}$ has a finite second moment[2]. For linear factor models, our stochastic discount factor proxy is given by $\widehat{m}(\boldsymbol{\lambda}) \doteq R_f^{-1}(1-\mathbf{f}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda})$. We write $\delta(\boldsymbol{\lambda})$ for the HJ distance associated with the stochastic discount factor proxy $\widehat{m}(\boldsymbol{\lambda})$.

Proposition 2.17. *Assume that $\mathbb{E}[(\mathbf{f})(\mathbf{f})'] < \infty$ and $\mathbb{E}[(\mathbf{r}^e)(\mathbf{r}^e)'] < \infty$. Then*

$$\delta^2(\boldsymbol{\lambda}) = R_f^{-2}\left(\boldsymbol{\mu}-\mathbf{B}\boldsymbol{\lambda}\right)'\boldsymbol{\Sigma}^+\left(\boldsymbol{\mu}-\mathbf{B}\boldsymbol{\lambda}\right).$$

where $\boldsymbol{\Sigma} \doteq \mathrm{cov}(\mathbf{r}^e)$ and the $+$ superscript denotes the Moore-Penrose inverse.

Proposition 2.17 is proved in the appendix. Thus we see that the HJ distance can be thought of as taking $W = \boldsymbol{\Sigma}^+$.

A few observations are in order. First the factor model is misspecified, so $\mathrm{rank}(\overline{\boldsymbol{\Psi}}) = \mathrm{rank}(\overline{\boldsymbol{\Psi}}^{(1)})+1$. Second, the vector $\boldsymbol{\lambda}$ is pseudo-point identified if and only if $\mathrm{rank}(\overline{\boldsymbol{\Psi}}^{(1)}) = k$. Thus we consider $\mathrm{rank}(\overline{\boldsymbol{\Psi}}^{(1)})$ to be more informative about identification than $\mathrm{rank}(\overline{\boldsymbol{\Psi}})$.

Remark 2.18. *Arellano et al. (2012) and Manresa et al. (2017) advocate tests of hypotheses on rank($\overline{\boldsymbol{\Psi}}$). If one views the linear factor model as an approximation which may potentially be misspecified, we consider it more revealing to test hypotheses on rank($\overline{\boldsymbol{\Psi}}^{(1)}$) than on rank($\overline{\boldsymbol{\Psi}}$).*

3 Testing the Kernel-Orthogonality Condition

In this section we describe a procedure to test whether the kernel-orthogonality condition 2.12 holds for a given factor $\mathbf{f}_i$. Throught we will take the null hypothesis to be that the

[2]See, for example, Luenberger (1997).

kernel-orthogonality condition holds for factor i, i.e. that the risk premium associated with factor i is point identified. We first describe the test procedure treating $r \doteq \text{rank}(\mathbf{B})$ as known to the econometrician. Then we show how to extend this test to a "plug-in" test which accomodates cases where r is unknown but a consistent estimator $\widehat{r}_T$ is available. Finally we discuss approaches from the literature for constructing such rank estimators.

Throughout we make the following assumption.

Assumption 3.1. *Define a matrix*

$$\mathbf{\Psi}^{(a)} \doteq \begin{bmatrix} \mathbf{\Psi}^{(1)} \\ \mathbf{f}\ \mathbf{f}' \end{bmatrix}' = \begin{bmatrix} \mathbf{f}\,(\mathbf{r}^e)' & \mathbf{f}\ \mathbf{f}' \end{bmatrix}.$$

As $T \to \infty$

$$\sqrt{T}\left[\text{vec}\left(\overline{\mathbf{\Psi}}_T^{(a)}\right) - \text{vec}\left(\overline{\mathbf{\Psi}}^{(a)}\right)\right] \xrightarrow{d} \text{Normal}\left(\mathbf{0}, \mathbf{\Omega}^{(a)}\right)$$

The standard estimator of $\mathbf{B}$ is given by $\widehat{\mathbf{B}}_T \doteq \overline{\mathbf{\Psi}}_T^{(1)} \widehat{\mathbf{\Sigma}}_{f,T}^{-1}$ where $\widehat{\mathbf{\Sigma}}_{f,T}$ is the sample covariance matrix of $\mathbf{f}$. We show in the appendix that under assumption 3.1,

$$\sqrt{T}\left(\text{vec}\left(\widehat{\mathbf{B}}_T' - \mathbf{B}'\right)\right) \xrightarrow{d} \text{Normal}\left(\mathbf{0}, \mathbf{\Omega}^{\mathbf{B}}\right)$$

as $T \to \infty$. An expression for $\mathbf{\Omega}^{\mathbf{B}}$ in terms of $\mathbf{\Omega}^{(a)}$ is given in the statement of lemma A.6. In practice, we find it easiest to estimate the matrix $\mathbf{\Omega}^{\mathbf{B}}$ by bootstrapping.

3.1 Kernel-Orthogonality Test with Known Rank Deficiency

We start with the assumption that $r = \text{rank}(\mathbf{B})$ is known. Then $d = k - r$ is the dimension of rank-deficiency, i.e $\dim(\ker(\mathbf{B})) = d$. Our approach tests for the existence of a d-dimensional orthonormal basis of the kernel of $\mathbf{B}$ which is orthogonal to $\mathbf{e}_i$.

We first consider the case in which the dimension d of rank deficiency is equal to one. He we are able to construct an asymptotically exact chi-square test of the kernel-orthogonality condition. In both our applications, this will be the empirically relevant case. We will then consider the case higher-dimensional rank deficiency $1 < d < k$ and construct an asymptotically conservative test.

When $d = 1$, we have that $\text{rank}(B) = k - 1$ so that the kernel of $\mathbf{B}$ is a one-dimensional subspace of $\mathbb{R}^k$. Our null hypothesis is that there exists a vector $\mathbf{u} \in \mathbb{R}^k$ such that

$$\mathbf{B}\mathbf{u} = \mathbf{0} \in \mathbb{R}^n \tag{11}$$

and furthermore that $\mathbf{u}_i = \mathbf{u}'\mathbf{e}_i = 0$. Additionally, we can without loss of generality impose the restriction that $\|\mathbf{u}\| = 1$. This restriction will ensure that the vector $\mathbf{u}$ is uniquely identified[3]. The vector $\mathbf{u}$ can be interpreted as parameterizing the *direction* of identification failure in the space of factor risk premia. Our approach will then be to test for the existence of a solution $\mathbf{u}$ to the above system.

It will be useful to define the following matrix:

$$\mathbf{\Omega}(\mathbf{u}) \doteq (I_{n\times n} \otimes \mathbf{u}')\, \mathbf{\Omega}^{\mathbf{B}}\, (I_{n\times n} \otimes \mathbf{u})\,. \tag{12}$$

and the matrix $\widehat{\mathbf{\Omega}}_T(\mathbf{u})$ in terms of $\widehat{\mathbf{\Omega}}_T^{\mathbf{B}}$ analogously.

Next, define a statistic $K_T^{(1)}(\mathbf{u})$ as

$$K_T^{(1)}(\mathbf{u}) \doteq T\left(\widehat{\mathbf{B}}_T\mathbf{u}\right)' \widehat{\mathbf{\Omega}}_T^{-1}(\mathbf{u})\left(\widehat{\mathbf{B}}_T\mathbf{u}\right) \tag{13}$$

We define our test statistic as follows:

$$\begin{aligned} K_T^{(1)} &\doteq \min_{\mathbf{u}\in\mathbb{R}^k} K_T^{(1)}(\mathbf{u}) \\ \text{s.t. } \mathbf{u}'\mathbf{u} &= 1 \\ \mathbf{u}'\mathbf{e}_i &= 0 \end{aligned}$$

If the kernel-orthogonality condition holds, this test statistic will have an asymptotic chi-square distribution. If the condition is violated, then the test statistic will approach infinity as the sample size becomes large.

Proposition 3.2. *Assume that rank*$(\mathbf{B}) = k-1$ *and that the kernel-orthogonality condition 2.12 holds. Then, as* $T \to \infty$

$$K_T^{(1)} \xrightarrow{d} \chi^2_{n-k+2}.$$

Then, we explicitly construct a test with asymptotic size α as follows. Write $\chi^2_{df,(1-\alpha)}$ for the $1-\alpha$ quantile of a χ^2 distribution with df degrees of freedom.

TEST PROCEDURE: For $d = 1$, *reject if* $K_T^{(1)} \geqslant \chi^2_{n-k+2,(1-\alpha)}$.

[3]Note that for any non-zero $\mathbf{u} \in \mathbb{R}^d$ satisfying (11) and $\mathbf{u}'\mathbf{e}_i = 0$, $\mathbf{u}/\|\mathbf{u}\|$ is also a solution. Since $\dim(\ker(\mathbf{B})) = 1$ this normalization produces a unique vector $\mathbf{u}$ up to unit multiplication.

Next, we consider the case in which the dimension of rank deficiency d is strictly greater than one, and construct an asymptotically conservative test. Our approach tests for the existence of a d-dimensional orthonormal basis of the kernel of $\mathbf{B}$ which is orthognal to $\mathbf{e}_i$.

Suppose that such an orthonormal basis $\mathbf{u}_1, ..., \mathbf{u}_d$ exists. We can stack the vectors together into a matrix matrix $\mathbf{U} \in \mathbb{R}^{k\times d}$. Observe that we have the restrictions

$$\begin{aligned} \mathbf{U}'\mathbf{U} &= I_{d\times d} \\ \mathbf{U}'\mathbf{e}_i &= \mathbf{0} \in \mathbb{R}^d. \end{aligned}$$

Next, define a matrix

$$\mathbf{\Omega}(\mathbf{U}) \doteq (I_{n\times n} \otimes \mathbf{U}')\mathbf{\Omega}^B(I_{n\times n} \otimes \mathbf{U})$$

and a statstic $K_T^{(d)}(\mathbf{U})$ associated with $\mathbf{U}$ as

$$K_T^{(d)}(\mathbf{U}) \doteq T \text{ vec}\left(\left(\widehat{\mathbf{B}}_T\mathbf{U}\right)'\right)' \widehat{\mathbf{\Omega}}_T(\mathbf{U})^{-1}\text{vec}\left(\left(\widehat{\mathbf{B}}_T\mathbf{U}\right)'\right).$$

Then we define our test statistic as

$$\begin{aligned} K_T^{(d)} \doteq & \min_{\mathbf{U}\in\mathbb{R}^{k\times d}} \ K_T^{(d)}(\mathbf{U}) \\ \text{s.t. } & \mathbf{U}'\mathbf{U} = I_{d\times d} \\ & \mathbf{U}'\mathbf{e}_i = \mathbf{0} \in \mathbb{R}^d \end{aligned}$$

If the kernel-orthogonality condition holds, then test statistic will be asymptotically dominated by a chi-square distribution. If the condition is violated, then the test statistic will approach infinity.

Proposition 3.3. *Assume rank*$(\mathbf{B}) = k-d$ *with* $1 < d < k$ *and that the kernel-orthogonality condition 2.12 holds. Then the limiting distribution of* $K_T^{(d)}$ *is stochastically dominated by a* $\chi^2_{n\times d}$ *random variable, i.e. for any constant* C

$$\lim_{T\to\infty} P\left(K_T^{(d)} \geqslant C\right) \leqslant P\left(\chi^2_{n\times d} \geqslant C\right)$$

We can therefore form a test with asymptotic size α as follows:

TEST PROCEDURE: For $1 < d < k$*, reject if* $K_T^{(d)} \geqslant \chi^2_{n\times d,(1-\alpha)}$.

While theoretically simple, this test has the unfortunate feature that the rank r must be correctly pre-specified. We show how to to relax this assumption in the subsequent section.

3.2 Kernel-Orthogonality Test with Unknown Rank Deficiency

Assume now that $r = \text{rank}(\mathbf{B})$ is not known a priori. Suppose we instead have an estimator $\widehat{r}_T$ with the property that $\widehat{r}_T \xrightarrow{p} r$. Then we can form a test that is asymptotically conservative with asymptotic size α using a plug-in approach as follows:

TEST PROCEDURE: (Plug-in KO test)

Reject if one of the following holds:

- $\widehat{r}_T = k - 1$ *and* $K_T^{(1)} \geqslant \chi^2_{n-k+2,1-\alpha}$.
- $0 < \widehat{r}_T < k - 1$ *and* $K_T^{(k-\widehat{r}_T)} \geqslant \chi^2_{n\times(k-\widehat{r}_T),1-\alpha}$.
- $\widehat{r}_T = 0$.

We refer to this test as a plug-in Kernel-Orthogonality (KO) test. Note that the test never rejects if $\widehat{r}_T = k$, i.e. the economist estimates that the matrix of betas has full rank. This is intuitive since if the population beta matrix has full rank, then all factor risk premia are point identified and the kernel-orthogonality condition holds for all factors. Additionally, note that the test always rejects if $\widehat{r}_T = 0$, i.e. the economist estimates that the matrix of betas has rank zero. Again, this is intuitive because if the population matrix of betas has rank zero, then no factor risk premia are point identified and the kernel-orthogonality condition fails for all factors. The following result shows that our test procedure is asymptotically conservative with size α.

Proposition 3.4. *The plug-in KO test is asymptotically conservative with size α. Furthermore, if $d = 1$, the test is asymptotically exact.*

3.3 Consistent Rank Estimators

We describe two approaches used in the literature to consistently estimate the rank of a matrix whose entries are contaminated by measurement error. The first approach is based on sequential hypothesis testing. The second approach is based on information criteria methods. In our empirical applications we use the first approach.

Potscher (1983) and Cragg and Donald (1997) describe sequential test procedures to consistently estimate the rank of a matrix whose entries are contaminated by measurement error. Starting with the null hypothesis of $r = 1$ a sequence of tests is performed. If the null hypothesis is rejected, r is increased by one and then the appropriate rank test is performed. When the null is not rejected, r is adopted as the etsimate $\widehat{r}_T$. For a fixed size α, this will not be a consistent estimator of the true rank r^*. However, if the size of the test α_T is allowed to depend on the sample size T, then a consistent estimator can be constructed. In particular Hosoya (1989) and Cragg and Donald (1997) show that by letting α_T go to zero at a sufficiently slow rate that $\log(\alpha_T)/T$ goes to zero then the rank estimator obtained from the sequential testing procedure will be a consistent estimator.

An alternate approach is to rely on information criteria methods to estimate the rank of the matrix. These methods construct a rank estimator r that minimizes a criterion function of the form

$$IC(r) = Tl(\widehat{B}_T, r) + f(T)F(r) \tag{14}$$

where $l(\widehat{B}_T, r)$ is the negative pseudo log likelihood evaluated at the restricted MLE with rank restricted to be r and $F(r) = r * (n + k - r)$. Alternative specifications have been proposed for $f(T)$, which correspond to different rank estimators. AIC corresponds to $f(T) = 2$, BIC corresponds to $f(T) = \log(T)$, and the Hannan and Quinn criterion uses $f(T) = 2\log(\log(T))$. Cragg and Donald (1997) show that if $f(T) \to \infty$ and $f(T)/T \to 0$ as $T \to \infty$ then the associated rank estimator will be consistent.

Standard asymptotic theory provides little guidance beyond requiring the consistency of a rank estimator. In our empirical analysis, we will focus on plug-in rank estimation via sequential testing. In simulations, we have found the sequential testing approach to give more accurate rank estimates than standard information criteria methods in the case of one-dimension rank deficiency. This will be the empirically relevant case in our applications.

Remark 3.5. *It is important to distinguish between the estimated rank of the population beta matrix $\widehat{r}_T$ and the rank of the estimated beta matrix rank$\left(\widehat{\mathbf{B}}_T\right)$. The former converges in probability to the rank of the population beta matrix $r^* =$ rank$(\mathbf{B})$ whereas the latter will always be equal to k.*

4 Empirical Results

We apply our methodology to two well-known factor specifications from the asset pricing literature. The first application is based on Lustig and Verdelhan (2007) who argue that the consumption risk of US investors is priced in the cross-section of currency returns. The second application is based on Adrian et al. (2014) who argue that the leverage of financial intermediaries is priced in the cross-section of US equity returns.

4.1 Application: Lustig and Verdelhan (2007)

Lustig and Verdelhan (2007) use the standard Fama-MacBeth methodology to argue that consumption growth risk of US investors is priced in the cross section of currency excess returns. Here, we apply out test methodology to one of their specifications.

Lustig and Verdelhan (2007) form eight currency portfolios of excess returns sorted by interest rate differentials. They motivate several factor specifications based on the Euler equation for US investors. We focus on the specification motivated by an investor with Epstein-Zin preferences with preferences over nondurables consumption. Under an appropriate log-linearization, this implies a two-factor linear model in nondurables consumption growth and log of the US market return. Applying our test methodology to their data, we obtain the results in tables 1 and 2.

Table 1: Results of sequential rank tests based on Lustig and Verdelhan (2007) for two-factor EZ-CCAPM model using eight sorted currency excess returns. Data are annual from 1953-2002. We obtain for reasonable values of α_T that $\text{rank}(\mathbf{B}) = 1$.

	p-value:
rank $= \mathbf{0}$	< 0.01
rank $\leqslant \mathbf{1}$	0.35

Table 1 shows the results on the sequential rank test procedure. We see for reasonable choices of α_T that we reject the null hypothesis that $r = 0$ but fail to reject the hypothesis that $r = 1$. Thus we obtain $\widehat{r}_T = 1$, i.e. we estimate that the matrix of population betas has one-dimensional rank deficiency, and hence the vector of factor risk premia fails to be point identified.

Table 2 shows the results of the plug-in kernel-orthogonality test for each factor risk premia using $\widehat{r}_T = 1$. We strongly reject the null hypothesis that the market risk premium is point identified. Thus we conclude that the market risk premium fails to be point identified

Table 2: Results of plug-in kernel-orthogonality tests based on Lustig and Verdelhan (2007) for two-factor EZ-CCAPM model using eight sorted currency excess returns. Data are annual from 1953-2002. Both tests use plug-in value $\widehat{r}_T = 1$.

	p-value:
KO-test (Market)	< 0.01
KO-test (consumption growth)	0.69

from the cross section of currency returns. We are unable to reject the null hypothesis that the risk premium of consumption growth is point identified. Thus we conclude that the risk premium associated with consumption growth is point identified.

4.2 Application: Adrian et al. (2014)

Our next application is to the Intermediary Leverage (hereafter: Leverage) factor of Adrian et al. (2014). They argue that the outsize importance of financial intermediaries in the US financial system implies that a "financial intermediary SDF" should price the cross-section of average returns. Following arguments from Brunnermeier and Pedersen (2009), they propose using the leverage of financial broker-dealers as a proxy for funding constraints and, therefore, as the factor in their SDF.

As the Leverage factor is measured quarterly, and returns are measured monthly, it is possible to estimate models at either frequency. Adrian et al. (2014) opt to use monthly returns with a quarterly time-series for Leverage. They also project Leverage onto a set of portfolios and use the estimated factor mimicking portfolio to estimate a monthly model. We will employ our methodology on these two models and also on a third one using quarterly returns and the quarterly factor.

The other factor in the specification is the excess return on the market, which we download from Ken French's website. This is also measured monthly. For returns, we use 45 portfolios from French's website: 10 portfolios sorted on size (market capitalization), 10 portfolios sorted on book-to-market, and 25 portfolios double sorted on size and book-to-market. Our date range is 1968 to 2010. When we consider quarterly returns, we simply compound the corresponding monthly returns. Finally, when we create the mimicking portfolio for leverage, we follow Adrian et al. (2014) and project the Leverage factor on the six Fama-French industry portfolios and the momentum portfolio.

Table 3 shows the results of the sequential rank test procedure. We see that for reasonable values of α_T that we reject the hypothesis that $r = 0$, but fail to reject the hypothesis that $r = 1$. Thus we obtain that $\widehat{r}_T = 1$.

Table 4 shows the results of the plug-in kernel-orthogonality tests for the market and intermediary leverage factors using the plug-in estimator $\widehat{r}_T = 1$. The first test fails to reject the null hypothesis that the market risk premium is point identified. The second test rejects the null hypothesis that the risk premium associated with intermediary leverage is point identified. We conclude that the risk premium associated with intermediary leverage fails to be point identified from the cross-section of equity returns. Thus our test procedure leads to a significantly different conclusion than Adrian et al. (2014) who claim that the intermediary leverage factor is strongly priced in the cross-section of equity returns.

Table 3: Results of sequential rank tests based on Adrian et al. (2014) for two-factor intermediary leverage model using 45 Fama-French portoflios. Data are based on monthly returns. We obtain for reasonable value of α_T that $\widehat{r}_T = 1$.

	p-value:
rank $= \mathbf{0}$	< 0.01
rank $\leqslant \mathbf{1}$	0.90

Table 4: Results of kernel-orthognality tests based on Adrian et al. (2014) for two-factor intermediary leverage model using 45 Fama-French portoflios. Data are based on monthly returns. Both tests use plug-in value $\widehat{r}_T = 1$.

	p-value:
KO-test (Market)	0.16
KO-test (Intermediary leverage)	< 0.01

5 Conclusion

This paper proposed diagnostic procedures to detect whether individual factor risk premia are in fact identified from return data. We illustrated how a regularity condition, which we called the *kernel-orthogonality condition*, is necessary and sufficient for a given factor risk premia to be identified. This condition is weaker than the standard rank condition because individual factor risk premia can be identified even when the full vector of factor risk premia is not. We showed that this condition is mathematically equivalent to the existence of a “true” (population) factor mimicking portfolio. Even if the linear factor model is misspecified, this condition remains informative about the identification of the factor risk premium consistent with minimal pricing error. We then developed a novel statistical test of whether individual factor risk premia are identified, and applied it to two well-known specifications from the empirical asset pricing literature.

Our test procedure can be a useful diagnostic tool for researchers in assessing a given multi factor specification. We would strongly recommend that applied researchers always test the rank condition on the matrix of estimated betas directly, a step often omitted in empirical papers. Our analysis goes beyond this, in that it can be informative to applied researchers about precisely which factor, or groups of factors, are causing rank-deficiency and which factors have strongly identified factor risk premia even in cases when the matrix of betas appears to have rank deficiency.

There are several natural extensions of our analysis. Firstly, since many factor specifications have been proposed in the literature, there are many additional specifications to which one could apply our test methodology. Additionally, the distribution theory for our test procedures was derived under conventional asymptotics treating the number of assets n as fixed while letting the sample length T approach infinity. This may be a poor approximation if we use a large panel of assets so the number of assets is large relative to the sample length. Asymptotics that let both n and T grow at comparable rates, or even holding T fixed and letting n approach infinity may be more appropriate in these settings. Alternatively, it would be potentially revealing to study the behavior of our test procedure under "weak identification" asymptotics, i.e. drifting parameter asymptotics as the identification strength drifts towards rank deficiency.[4] We leave these considerations and others to future research.

[4]This is somewhat complicated in our setting by the fact that there is no unique or obvious rank-deficient beta matrix to drift towards.

References

Adrian, Tobias, Erkko Etula, and Tyler Muir. 2014. Financial Intermediaries and the Cross-Section of Asset Returns. *The Journal of Finance* 69 (6):2557–2596.

Arellano, Manuel, Lars Peter Hansen, and Enrique Sentana. 2012. Underidentification? *Journal of Econometrics* 170 (2):256–280.

Black, Fischer, Michael C Jensen, and Myron Scholes. 1972. The Capital Asset Pricing Model: Some Empirical Tests. *Studies in the Theory of Capital Markets* .

Breeden, Douglas T, Michael R Gibbons, and Robert H Litzenberger. 1989. Empirical Tests of the Consumption-Oriented CAPM. *The Journal of Finance* 44 (2):231–262.

Brunnermeier, Markus K and Lasse Heje Pedersen. 2009. Market Liquidity and Funding Liquidity. *The Review of Financial Studies* 22 (6):2201–2238.

Bryzgalova, Svetlana. 2015. Spurious Factors in Linear Asset Pricing Models .

Burnside, Craig. 2016. Identification and Inference in Linear Stochastic Discount Factor Models with Excess Returns. *Journal of Financial Econometrics* 14 (2):295–330.

Carhart, Mark M. 1997. On Persistence in Mutual Fund Performance. *The Journal of Finance* 52 (1):57–82.

Chen, Nai-Fu, Richard Roll, and Stephen A Ross. 1986. Economic Forces and the Stock Market. *Journal of Business* 383–403.

Cragg, John G and Stephen G Donald. 1997. Inferring the Rank of a Matrix. *Journal of Econometrics* 76 (1-2):223–250.

Fama, Eugene F and Kenneth R French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33:3–56.

Fama, Eugene F and James D MacBeth. 1973. Risk, Return, and Equilibrium: Empirical tests. *Journal of Political Economy* 81 (3):607–636.

Giglio, Stefano and Dacheng Xiu. 2021. Asset Pricing with Omitted Factors. *Journal of Political Economy* 129 (7):1947–1990.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti. 2014. Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors. *The Review of Financial Studies* 27 (7):2139–2170.

———. 2017. Spurious Inference in Reduced-Rank Asset-Pricing Models. *Econometrica* 85 (5):1613–1628.

Hansen, Lars Peter and Ravi Jagannathan. 1997. Assessing Specification Errors in Stochastic Discount Factor Models. *The Journal of Finance* 52 (2):557–590.

Harvey, Campbell R, Yan Liu, and Heqing Zhu. 2016. … and the cross-section of expected returns. *The Review of Financial Studies* 29 (1):5–68.

He, Zhiguo, Bryan Kelly, and Asaf Manela. 2017. Intermediary Asset Pricing: New Evidence from Many Asset Classes. *Journal of Financial Economics* 126 (1):1–35.

Hosoya, Yuzo. 1989. Hierarchical Statistical Models and a Generalized Likelihood Ratio Test. *Journal of the Royal Statistical Society: Series B (Methodological)* 51 (3):435–447.

Huberman, Gur, Shmuel Kandel, and Robert F Stambaugh. 1987. Mimicking portfolios and exact arbitrage pricing. *The Journal of Finance* 42 (1):1–9.

Jagannathan, Ravi and Zhenyu Wang. 1996. The Conditional CAPM and the Cross-Section of Expected Returns. *The Journal of Finance* 51 (1):3–53.

———. 1998. An Asymptotic Theory for Estimating Beta-Pricing Models using Cross-Sectional Regression. *The Journal of Finance* 53 (4):1285–1309.

Kan, Raymond and Cesare Robotti. 2009. Model Comparison using the Hansen-Jagannathan Distance. *The Review of Financial Studies* 22 (9):3449–3490.

Kazemi, Maziar. 2019. Semi-Parametric Estimation of Factor Risk-Premia.

Kleibergen, Frank. 2009. Tests of risk premia in linear factor models. *Journal of econometrics* 149 (2):149–173.

Kleibergen, Frank and Richard Paap. 2006. Generalized Reduced Rank Tests using the Singular Value Decomposition. *Journal of Econometrics* 133 (1):97–126.

Kleibergen, Frank and Zhaoguo Zhan. 2018. Identification-Robust Inference on Risk Premia of Mimicking Portfolios of Non-Traded Factors. *Journal of Financial Econometrics* 16 (2):155–190.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. 2020. Shrinking the Cross-Section. *Journal of Financial Economics* 135 (2):271–292.

Lettau, Martin and Sydney Ludvigson. 2001. Resurrecting the (C) CAPM: A Cross-Sectional Test when Risk Premia are Time-Varying. *Journal of Political Economy* 109 (6):1238–1287.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken. 2010. A Skeptical Appraisal of Asset Pricing Tests. *Journal of Financial Economics* 96 (2):175–194.

Lintner, John. 1965. Security Prices, Risk, and Maximal Gains from Diversification. *The Journal of Finance* 20 (4):587–615.

Luenberger, David G. 1997. *Optimization by Vector Space Methods*. John Wiley & Sons.

Lustig, Hanno and Adrien Verdelhan. 2007. The Cross Section of Foreign Currency Risk Rremia and Consumption Growth Risk. *American Economic Review* 97 (1):89–117.

Manresa, Elena, Francisco Peñaranda, and Enrique Sentana. 2017. Empirical Evaluation of Overspecified Asset Pricing Models .

Potscher, BM. 1983. Order Estimation in ARMA-Models by Lagrangian Multiplier Tests. *The Annals of Statistics* 872–885.

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

Sharpe, William F. 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance* 19 (3):425–442.

Windmeijer, Frank et al. 2018. Testing Over-and Underidentification in Linear Models, with Applications to Dynamic Panel Data and Asset-Pricing Models .

A Appendix: Derivations and Proofs

Proposition A.1. *The set $I_\Lambda^i \subseteq \mathbb{R}$ is either a singleton or the whole real line.*

Proof. For any $\boldsymbol{\lambda} \in I_\Lambda$ we can write equation (7) as

$$\mathbf{B}\boldsymbol{\lambda} = \boldsymbol{\mu}.$$

Assume I_Λ^i is not a singleton. Then there exist distinct vectors $\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)} \in I_\Lambda$ such that $\boldsymbol{\lambda}_i^{(1)} \neq \boldsymbol{\lambda}_i^{(2)}$. Observe that for any $\alpha \in \mathbb{R}$, we see that

$$(1-\alpha)\mathbf{B}\boldsymbol{\lambda}^{(1)} + \alpha\mathbf{B}\boldsymbol{\lambda}^{(2)} = \boldsymbol{\mu}.$$

Thus $(1-\alpha)\boldsymbol{\lambda}^{(1)} + \alpha\boldsymbol{\lambda}^{(2)} \in I_\Lambda$. Write $l_1 = \boldsymbol{\lambda}_i^{(1)}$ and $l_2 = \boldsymbol{\lambda}_i^{(2)}$. Now for any $c \in \mathbb{R}$, take $\alpha = \frac{c-l_1}{l_2-l_1}$ so that $(1-\alpha)l_1 + \alpha l_2 = c$. It follows that $c \in I_\Lambda^i$. □

Proposition A.2. *The following are equivalent:*

(1) $\boldsymbol{\lambda}_i$ *is point identified.*

(2) $\ker(\mathbf{B}) \perp \boldsymbol{e}_i \in \mathbb{R}^k$

(3) $\boldsymbol{e}_i \in \text{Image}(\mathbf{B}')$.

Proof. By assumption 2.3, we know that there exists a solution $\boldsymbol{\lambda}^*$ to equation (7), which can be obtained as $\boldsymbol{\lambda}^* = \boldsymbol{\Sigma}_f \overline{\boldsymbol{\Psi}}_1^+ \overline{\boldsymbol{\Psi}}_2$. Furthermore, the set of solutions $\boldsymbol{\lambda}$ to equation (7) can be expressed as

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}^* + \boldsymbol{\eta}$$

for $\boldsymbol{\eta} \in \ker(\overline{\boldsymbol{\Psi}}_1)$. It is clear therefore that all solutions $\boldsymbol{\lambda}$ satisfy $\boldsymbol{\lambda} \cdot \mathbf{e}_i = \boldsymbol{\lambda}_i^*$ if and only if $\boldsymbol{\eta} \cdot \mathbf{e}_i = 0$ for all $\boldsymbol{\eta} \in \ker(\overline{\boldsymbol{\Psi}}_1)$. Thus (1) and (2) are equivalent. The equivalence of (2) and (3) follows from the fact that $\ker(\mathbf{B})^\perp = \text{Image}(\mathbf{B}') \in \mathbb{R}^k$. □

Lemma A.3. $I_{\Lambda,W}$ *can be equivalently expressed as*

$$I_{\Lambda,W} = \left\{\boldsymbol{\lambda} \in \mathbb{R}^k : \mathbf{B}'W\mathbf{B}\boldsymbol{\lambda} = \mathbf{B}'W\boldsymbol{\mu}\right\}$$

Proof. This follows immediately from the fact that the objective function is a convex differentiable function of $\boldsymbol{\lambda}$. □

Proof of proposition 2.16:

To prove necessecity, suppose condition 2.12 is violated. Then there exists a vector $\boldsymbol{\eta} \in \ker(\mathbf{B})$ with $\boldsymbol{\eta}_i \neq 0$. Take $\boldsymbol{\lambda}^* \in I_{\Lambda,W}$. From lemma A.3 we see that $\boldsymbol{\lambda}^* + \boldsymbol{\eta} \in I_{\Lambda,W}$. Thus, $\boldsymbol{\lambda}_i^* \neq \boldsymbol{\lambda}_i^* + \boldsymbol{\eta}_i \in I_{\Lambda,W}^i$. Thus $I_{\Lambda,W}^i$ is not a singleton.

To prove sufficiency, assume that W is positive definite and suppose $I_{\Lambda,W}^i$ is not a singleton. Then there exist two vectors $\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)}$ in $I_{\Lambda,W}$ with $\boldsymbol{\lambda}_i^{(1)} \neq \boldsymbol{\lambda}_i^{(2)}$. Then consider the vector $\boldsymbol{\eta} = \boldsymbol{\lambda}_i^{(1)} - \boldsymbol{\lambda}_i^{(2)}$. From Lemma A.3 we see that

$$\mathbf{B}'W\mathbf{B}\boldsymbol{\eta} = \mathbf{0}$$

and hence

$$\boldsymbol{\eta}'\mathbf{B}'W\mathbf{B}\boldsymbol{\eta} = 0.$$

Since W is positive definite, it must be the case that $\mathbf{B}\boldsymbol{\eta} = \mathbf{0}$. Thus condition 2.12 is violated. $\square$

Proof of proposition 2.17:

The HJ distance $\delta(\boldsymbol{\lambda})$ solves the convex minimization problem:

$$\begin{aligned} \frac{1}{2}\delta^2(\boldsymbol{\lambda}) = \min_{m \in L^2} \ & \frac{1}{2}\mathbb{E}\left[\left(m - R_f^{-1}(1 - \mathbf{f}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda})\right)^2\right] \\ \text{s.t. } & \mathbb{E}[m\mathbf{r}^e] = \mathbf{0} \\ & \mathbb{E}[mR_f] = 1 \end{aligned}$$

which can be solved by standard duality methods for infinite-dimensional problems (See Luenberger (1997) for discussion). Form the Lagrangian

$$L(m, \boldsymbol{\gamma}_1, \gamma_2) = \frac{1}{2}\left(m - R_f^{-1}(1 - \mathbf{f}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda})\right)^2 - \gamma_1'\mathbf{r}^e - \gamma_2(mR_f - 1).$$

Any solution to the primal problem must be a saddle-point of the Lagrangian. Taking first-order conditions with respect to m we obtain

$$m = R_f^{-1}\left(1 - \mathbf{f}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}\right) + \boldsymbol{\gamma}_1'\mathbf{r}^e + \gamma_2 R_f.$$

Next, write $\mathbf{S} = \mathbb{E}[(\mathbf{r}^e)(\mathbf{r}^e)']$. The first-order conditions with respect to the multipliers imply that the constraints must bind. From this we obtain,

$$0 = R_f^{-1}(\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda}) + \mathbf{S}\boldsymbol{\gamma}_1 + \gamma_2 R_f \boldsymbol{\mu}$$
$$0 = R_f \boldsymbol{\gamma}_1' \boldsymbol{\mu} + \gamma_2 R_f^2$$

which together simplify to

$$0 = R_f^{-1}(\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda}) + \boldsymbol{\Sigma}\boldsymbol{\gamma}_1.$$

It follows that we can write

$$\boldsymbol{\gamma} = -\boldsymbol{\Sigma}^{+}(\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda}) + \boldsymbol{\xi}$$

where $\boldsymbol{\xi} \in \ker(\boldsymbol{\Sigma})$. Hence we must have

$$m = R_f^{-1}\left(1 - \mathbf{f}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}\right) - R_f^{-1}(\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda})'\boldsymbol{\Sigma}^{+}(\mathbf{r}^e - \boldsymbol{\mu}) + \boldsymbol{\xi}'(\mathbf{r}^e - \boldsymbol{\mu}).$$

Substituting this expression for m into the objective function and evaluating gives

$$\frac{1}{2}\delta(\boldsymbol{\lambda})^2 = \frac{1}{2}R_f^{-2}(\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda})'\boldsymbol{\Sigma}^{+}(\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\lambda})$$

as claimed. □

Proposition A.4. *Let $U \in \mathbb{R}^{m \times p}$ be a random matrix and let $\boldsymbol{\Theta} \in \mathbb{R}^{p \times q}$ be a matrix of scalars such that*

$$\mathbb{E}[U]\boldsymbol{\Theta} = \mathbf{0}_{m \times q}.$$

Then the covariance matrix $\Omega(\boldsymbol{\Theta}) \doteq \mathbb{E}[\text{vec}(U\boldsymbol{\Theta})\text{vec}(U\boldsymbol{\Theta})']$ associated with the vectorized system

$$\text{vec}(\mathbb{E}[U]\boldsymbol{\Theta}) = \mathbf{0} \in \mathbb{R}^{mq}$$

can be equivalently expressed as

$$\Omega(\boldsymbol{\Theta}) = (\boldsymbol{\Theta}' \otimes I_{m \times m})\,\mathbb{E}[\text{vec}(U)\text{vec}(U)']\,(\boldsymbol{\Theta} \otimes I_{m \times m})$$

Proof. Observe that

$$\text{vec}(U\boldsymbol{\Theta}) = (\boldsymbol{\Theta}' \otimes I_{m \times m})\,\text{vec}(U).$$

Therefore

$$\Omega(\boldsymbol{\Theta}) = \mathbb{E}[\text{vec}(U\boldsymbol{\Theta})\text{vec}(U\boldsymbol{\Theta})']$$

$$
\begin{aligned}
&= \mathbb{E}\left[\left(\boldsymbol{\Theta}' \otimes I_{m\times m}\right) \text{vec}(U)\text{vec}(U)' \left(\boldsymbol{\Theta}' \otimes I_{m\times m}\right)'\right] \\
&= \left(\boldsymbol{\Theta}' \otimes I_{m\times m}\right) \mathbb{E}\left[\text{vec}(U)\text{vec}(U)'\right] \left(\boldsymbol{\Theta}' \otimes I_{m\times m}\right)' \\
&= \left(\boldsymbol{\Theta}' \otimes I_{m\times m}\right) \mathbb{E}\left[\text{vec}(U)\text{vec}(U)'\right] \left(\boldsymbol{\Theta} \otimes I_{m\times m}\right)
\end{aligned}
$$

as claimed. □

Lemma A.5. *Let $M = [M_1\ M_2]$ where $M_1 \in \mathbb{R}^{q\times m}$ and $M_2 \in \mathbb{R}^{q\times q}$ is symmetric. Then*

$$
\frac{\partial\ \text{vec}\left(M_2^{-1}M_1\right)}{\partial\ \text{vec}\left(M\right)} = \left[M_2^{-1} \otimes I_{m\times m} \quad -\left(M_2^{-1}M_1\right) \otimes M_2^{-1}\right].
$$

Proof. Observe that $\text{vec}(M) = \text{vec}\left([M_1\ M_2]\right) = \left(\text{vec}(M_1), \text{vec}(M_2)\right)$ and compute

$$
\begin{aligned}
\frac{\partial\ \text{vec}\left(M_2^{-1}M_1\right)}{\partial\ \text{vec}\left(M_1\right)} &= \frac{\partial\ \text{vec}\left(M_2^{-1}M_1 I_{m\times m}\right)}{\partial\ \text{vec}\left(M_1\right)} = M_2^{-1} \otimes I_{m\times m} \\
\frac{\partial\ \text{vec}\left(M_2^{-1}M_1\right)}{\partial\ \text{vec}\left(M_2\right)} &= \frac{\partial\ \text{vec}\left(I_{q\times q} M_2^{-1}M_1\right)}{\partial\ \text{vec}\left(M_2\right)} = -\left(M_2^{-1}M_1\right) \otimes M_2^{-1}.
\end{aligned}
$$

The result follows. □

Lemma A.6. *Under assumption 3.1 we have that*

$$
\sqrt{T}\left(\text{vec}\left(\widehat{\mathbf{B}}_T' - \mathbf{B}'\right)\right) \xrightarrow{d} \text{Normal}\left(\mathbf{0}, \boldsymbol{\Omega}^{\mathbf{B}}\right)
$$

where

$$
\boldsymbol{\Omega}^{\mathbf{B}} \doteq \mathbf{D}'\boldsymbol{\Omega}^{(a)}\mathbf{D}
$$

with

$$
\mathbf{D} \doteq \left[\boldsymbol{\Sigma}_f^{-1} \otimes I_{n\times n} \quad -\mathbf{B}' \otimes \boldsymbol{\Sigma}_f^{-1}\right]
$$

Proof. Observe that the function $f(\overline{\boldsymbol{\Psi}}^{(a)}) = \text{vec}\left(\boldsymbol{\Sigma}_f^{-1}\left(\overline{\boldsymbol{\Psi}}^{(1)}\right)'\right) = \text{vec}\left(\mathbf{B}'\right)$ is a differentiable function of $\overline{\boldsymbol{\Psi}}^{(a)}$ and hence $\text{vec}\left(\overline{\boldsymbol{\Psi}}^{(a)}\right)$. Applying the delta method together with lemma A.5 gives the desired result. □

An alternative approach would be to think of $\text{vec}\left(\widehat{\mathbf{B}}_T'\right)$ as a Z-estimator. In particular, if we consider the population problem

$$
\min_{\mathbf{B}\in\mathbb{R}^{n\times k}} \mathbb{E}\left[\left(\mathbf{r}^e - \mathbf{B}\mathbf{f}\right)'\left(\mathbf{r}^e - \mathbf{B}\mathbf{f}\right)\right] \tag{15}
$$

with $\widehat{\mathbf{B}}_T$ as the sample analogue to $\mathbf{B}$, we see that the first-order conditions can be expressed as

$$\mathbb{E}\left[\left(\mathbf{r}^e - \mathbf{B}\mathbf{f}\right)\mathbf{f}'\right] = \mathbf{0} \in \mathbb{R}^{n\times k}$$

or in vectorized form as

$$\mathbb{E}\left[\operatorname{vec}\left(\left(\boldsymbol{\Psi}^{(1)}\right)'\right) - \left(I_{k\times k} \otimes \left(\mathbf{f}\ \mathbf{f}'\right)\right)\operatorname{vec}\left(\mathbf{B}'\right)\right].$$

Applying the standard result on asymptotic normality of Z-estimators will give the same result as lemma A.6.

Proof of proposition 3.2:

Recall that the solution to the system defined by equation (11) is unique up to unit multiplication, and that furthermore the criterion $K_T^{(1)}(\mathbf{u})$ is symmetric in $\mathbf{u}$. Thus we can guarantee point identification of $\mathbf{u}$ by restricting $\mathbf{u}$ to lie in a half-space of $\mathbb{R}^d$ without changing the criterion. The resulting sequence $\hat{\mathbf{u}}_T$ will then be a consistent extremum estimator of the unique population solution $\mathbf{u}^*$.

To obtain the asymptotic chi-square distribution of the test statistic, we first make the re-parameterization $\mathbf{u} = \mathbf{u}(\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \mathbb{R}^{k-1}$ and the function $\mathbf{u}(\cdot)$ is defined by $\mathbf{u}(\boldsymbol{\theta}) \doteq (\boldsymbol{\theta}', \sqrt{1-\boldsymbol{\theta}'\boldsymbol{\theta}})'$. We naturally restrict $\boldsymbol{\theta}$ to the compcat set

$$\boldsymbol{\Theta} = \{\boldsymbol{\theta} \in \mathbb{R}^{k-1} : \boldsymbol{\theta}'\boldsymbol{\theta} \leqslant 1\}$$

Observe that the gradient of $\mathbf{u}$ with respect to $\boldsymbol{\theta}$ is given by

$$\mathbf{D}(\boldsymbol{\theta}) = \begin{bmatrix} I_{(k-1)\times(k-1)} \\ (1-\boldsymbol{\theta}'\boldsymbol{\theta})^{-1/2}\boldsymbol{\theta}') \end{bmatrix} \in \mathbb{R}^{k\times(k-1)}.$$

Write $\boldsymbol{\theta}^*$ for the unique $\boldsymbol{\theta}$ such that $\mathbf{u}^* = \mathbf{u}(\boldsymbol{\theta}^*)$, and $\widehat{\boldsymbol{\theta}}_T$ analogously.

From proposition A.4 we see that

$$\sqrt{T}\widehat{\mathbf{B}}_T\mathbf{u}(\boldsymbol{\theta}^*) \xrightarrow{d} \text{Normal}\left(\mathbf{0}, \boldsymbol{\Omega}(\mathbf{u}^*)\right).$$

For notational simplicity write $\widehat{\boldsymbol{\Omega}}_T$ for $\widehat{\boldsymbol{\Omega}}_T(\widehat{\boldsymbol{u}}_T)$. Form the Cholesky factorization

$$\widehat{\boldsymbol{\Omega}}_T = \widehat{C}_T\widehat{C}_T'.$$

Then we see that

$$Z_T \doteq \sqrt{T}\widehat{C}_T' \left(\widehat{\boldsymbol{B}}_T \mathbf{u}^*\right) \xrightarrow{d} \text{Normal}\left(\mathbf{0}, I_{n\times n}\right).$$

Next, we use the standard first-order expansion for the extremum estimator $\widehat{\boldsymbol{\theta}}_T$. Write $\mathbf{D} = \mathbf{D}(\boldsymbol{\theta}^*)$, and write $\widehat{G}_T = \widehat{C}_T'\widehat{\mathbf{B}}_T\mathbf{D}$. Then

$$\begin{aligned}
\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^*\right) &= -\left(\mathbf{D}'\widehat{\mathbf{B}}_T'\widehat{\boldsymbol{\Omega}}_T^{-1}\widehat{\mathbf{B}_T\mathbf{D}}\right)^{-1}\mathbf{D}'\widehat{\mathbf{B}}_T'\widehat{\boldsymbol{\Omega}}_T^{-1}\sqrt{T}\widehat{\mathbf{B}}_T\mathbf{u}^* + o_P(1) \\
&= -\left(\widehat{G}_T'\widehat{G}_T\right)^{-1}\widehat{G}_T'\sqrt{T}\widehat{C}_T'\widehat{\mathbf{B}}_T\mathbf{u}^* + o_P(1)
\end{aligned}$$

Finally, write

$$\begin{aligned}
H_T \doteq \sqrt{T}\widehat{C}_T'\widehat{\mathbf{B}}_T\mathbf{u}(\widehat{\boldsymbol{\theta}}_T) &= \sqrt{T}\widehat{C}_T'\widehat{\mathbf{B}}_T\mathbf{u}(\boldsymbol{\theta}^*) + \sqrt{T}\widehat{C}_T'\widehat{\mathbf{B}}_T\mathbf{D}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^*\right) + o_P(1) \\
&= \sqrt{T}\widehat{C}_T'\widehat{\mathbf{B}}_T\mathbf{u}(\boldsymbol{\theta}^*) - \widehat{G}_T\left(\widehat{G}_T'\widehat{G}_T\right)^{-1}\widehat{G}_T'\sqrt{T}\widehat{C}_T'\widehat{\mathbf{B}}_T\mathbf{u}(\boldsymbol{\theta}^*) \\
&= \underbrace{\left[I_{n\times n} - \widehat{G}\left(\widehat{G}'\widehat{G}\right)^{-1}\widehat{G}'\right]}_{\widehat{P}}\sqrt{T}\widehat{C}'\widehat{\mathbf{B}}_T\mathbf{u}^* + o_P(1) \\
&= \widehat{P}_T Z_T + o_P(1)
\end{aligned}$$

We make the observation that $\widehat{P}$ is symmetric and idempotent with rank $k-2$, and that $H_T'H_T = K_T^{(1)}$. It follows immediately that $K_T^{(1)}$ converges in distribution to a χ^2 random variable with $n-(k-2)$ degrees of freedom. □

Proof of proposition 3.3:
Since $\dim\left(\ker\left(\mathbf{B}\right)\right) = d$ for $1 < d < k$ and the kernel-orthogonality condition holds, we know that there exists a matrix $\mathbf{U}^* \in \mathbb{R}^{k\times d}$ such that $\left(\mathbf{U}^*\right)'\left(\mathbf{U}^*\right) = I_{d\times d}$, $\left(\mathbf{U}^*\right)'\mathbf{e}_i = \mathbf{0} \in \mathbb{R}^d$, and $\mathbf{B}\mathbf{U}^* = \mathbf{0} \in \mathbb{R}^{n\times d}$.

Write the Cholesky factorization $\widehat{\boldsymbol{\Omega}}_T^{-1}(\mathbf{U}^*) = \widehat{C}\widehat{C}'$. We see that

$$\widehat{Z} \doteq \sqrt{T}\widehat{C}'\text{vec}\left(\left(\widehat{\mathbf{B}}_T\mathbf{U}^*\right)'\right) \xrightarrow{d} \text{Normal}\left(\mathbf{0}, I_{(nd)\times(nd)}\right).$$

Now, observe by definition of $K_T^{(d)}$ that

$$K_T^{(d)} \leqslant K_T^{(d)}(\mathbf{U}^*) = \widehat{Z}'\widehat{Z} \xrightarrow{d} \chi^2_{n\times d}$$

$\square$

Proof of proposition 3.4:

We break the proof in cases.

Case I: $d = 0$

Observe that $P(\widehat{r}_T = k) \to 1$ and that $P(\text{“reject”}|\widehat{r}_T = k) = 0$. Therefore applying the law of total probability, we see that

$$\begin{aligned} P(\text{“reject”}) &= P(\text{“reject”}|\widehat{r}_T = k)P(\widehat{r}_T = k) + P(\text{“reject”}|\widehat{r}_T < k)P(\widehat{r}_T < k) \\ &\leqslant 0 \cdot P(\widehat{r}_T = k) + 1 \cdot P(\widehat{r}_T < k) \\ &\leqslant P(\widehat{r}_T < k) \end{aligned}$$

which converges to 0 as $T \to \infty$. Therefore the test is asymptotically conservative at level α.

Case II: $d = 1$

We have that

$$P(\text{“reject”}) = P(\text{“reject”}|\widehat{r}_T = k-1)P(\widehat{r}_T = k-1) + P(\text{“reject”}|\widehat{r}_T \neq k-1)P(\widehat{r}_T \neq k-1)$$

We can apply the dominated convergence theorem to both terms to compute their respective limits as $T \to \infty$. Note that the second term converges to zero by consistency of $\widehat{r}_T$. Additionally, the first term converges to α by proposition 3.2 and consistency of $\widehat{r}_T$. Thus the test has asymptotic size α.

Case III: $1 < d < k$.

Applying the law of total probability, we see that

$$\begin{aligned} P(\text{“reject”}) &= P(\text{“reject”}|\widehat{r}_T = k-d)P(\widehat{r}_T = k-d) + P(\text{“reject”}|\widehat{r}_T \neq k-d)P(\widehat{r}_T \neq k-d) \\ &\leqslant P(\text{“reject”}|\widehat{r}_T = k-d) \cdot 1 + 1 \cdot P(\widehat{r}_T \neq k-d). \end{aligned}$$

In view of proposition 3.3 we see that the first term is bounded above by α. By consistency of $\widehat{r}_T$ the second term converges to zero. Thus the test is asymptotically conservative at level α.

Case IV: $d = k$.

If $d = k$, we see that $\text{rank}(\mathbf{B}) = 0$ and hence $\mathbf{B} = \mathbf{0} \in \mathbb{R}^{n\times k}$. Thus condition 2.12 cannot be satisfied. □

B Appendix: Time-varying Risk-free Rates and Betas

We extend our analysis in section 2 to allow for both the risk-free rate R_f and covariances (i.e. β's) to be time-varying. Suppose that the conditional one-period stochastic discount factor $m_{t,t+1}$ has the form

$$m_{t,t+1} = R_{f,t}^{-1}(1 - \mathbf{f}_{t+1}'\mathbf{\Sigma}_f^{-1}\boldsymbol{\lambda})$$

where $R_{f,t}$ is the gross risk-free rate from time t to $t+1$. Assume that $\mathbf{f}_{t+1}$ has conditional mean zero. Then we see that

$$\mathbb{E}_t[m_{t,t+1}R_{f,t}] = 1$$

so $m_{t,t+1}$ correctly prices the one-period risk-free asset. Additionally, to correctly price a vector of excess returns $\mathbf{r}_{t,t+1}^e$ we must have that

$$\mathbf{0} = \mathbb{E}_t\left[m_{t,t+1}\mathbf{r}^e\right] = R_{f,t}^{-1}\mathbb{E}\left[\mathbf{r}_{t,t+1}^e\right] - R_{f,t}^{-1}\mathbb{E}_t\left[\mathbf{r}_{t,t+1}^e\mathbf{f}_{t+1}'\right]\mathbf{\Sigma}_f^{-1}\boldsymbol{\lambda}$$

It follows that

$$\mathbb{E}_t\left[\mathbf{r}_{t,t+1}^e\right] = \mathbb{E}_t\left[\mathbf{r}_{t,t+1}^e\mathbf{f}_{t+1}'\right]\mathbf{\Sigma}_f^{-1}\boldsymbol{\lambda}. \tag{16}$$

Observe that in equation (16), both expected returns and covariances (i.e. β's) can be time-varying. Projecting down equation (16), we obtain that

$$\begin{aligned}\boldsymbol{\mu} = \mathbb{E}[\mathbf{r}^e] &= \overline{\mathbf{\Psi}}^{(1)}\mathbf{\Sigma}_f^{-1}\boldsymbol{\lambda} \\ &= \mathbf{B}\boldsymbol{\lambda}\end{aligned}$$

which is exactly equation (7).

C Appendix: Factor Mimicking Portfolios

Consider any vector of portfolio weights $\boldsymbol{\phi} \in \mathbb{R}^n$. Observe that

$$\text{cov}(\boldsymbol{\phi}'\mathbf{r}^e, \mathbf{f}) = \boldsymbol{\phi}'\mathbb{E}[\mathbf{r}^e\mathbf{f}'] = \boldsymbol{\phi}'\overline{\mathbf{\Psi}}^{(1)}.$$

It follows that the vector of factor betas associated with the portfolio $\boldsymbol{\psi}$ satisfies

$$\begin{aligned}\boldsymbol{\beta}'_{\boldsymbol{\phi}} &= \mathrm{cov}(\boldsymbol{\phi}'\mathbf{r}^e, \mathbf{f})\boldsymbol{\Sigma}_f^{-1} \\ &= \boldsymbol{\phi}'\overline{\boldsymbol{\Psi}}^{(1)}\boldsymbol{\Sigma}_f^{-1} \\ &= \boldsymbol{\phi}'\mathbf{B}\end{aligned}$$

Thus $\boldsymbol{\phi}$ is a mimicking portfolio for factor $\mathbf{f}_i$ if and only if

$$\mathbf{B}'\boldsymbol{\phi} = \mathbf{e}_i. \tag{17}$$

Next, assume the model is correctly specified and that $\boldsymbol{\phi}$ is a mimicking portfolio. Using the fact that $m = R_f^{-1}(1 - \mathbf{f}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda})$

$$\begin{aligned}\mathbb{E}[\boldsymbol{\phi}'\mathbf{r}_e] &= \boldsymbol{\phi}'\mathbb{E}[\mathbf{r}^e] \\ &= -\boldsymbol{\phi}' R_f \mathrm{cov}(\mathbf{r}^e, m) \\ &= \boldsymbol{\phi}'\mathbb{E}[\mathbf{r}^e\mathbf{f}']\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda} \\ &= \boldsymbol{\phi}'\mathbf{B}\boldsymbol{\lambda} \\ &= \mathbf{e}_i'\boldsymbol{\lambda} \\ &= \boldsymbol{\lambda}_i.\end{aligned}$$

D Appendix: Simulations

Simulations for Figure 1:

We perform a simulation based on example 2.10. We take $n = 25$, $T = 50$, $k = 2$. Each period, returns and factors are simulated as

$$\begin{aligned}\mathbf{r}_i^e &= \boldsymbol{\beta}\left(f^* + \lambda^*\right) + \boldsymbol{\epsilon}_i \\ \mathbf{f}_1 &= \frac{1}{2}f^* + \boldsymbol{\eta}_1 \\ \mathbf{f}_2 &= \frac{1}{2}f^* + \boldsymbol{\eta}_2\end{aligned}$$

where the random variables $\{\boldsymbol{\epsilon}_i\}_{i=1}^n$, f^*, $\{\boldsymbol{\eta}_j\}_{j=1}^2$ are generated as iid Normal$(0, (0.1)^2)$. We take $\lambda^* = 0.07$, and calibrate $\boldsymbol{\beta}$ to the market betas for the Fama-French 25 portfolios sorted on size and book-to-market. We assume that the factor f^* is latent, and that the econometrician only directly observes the factors $\mathbf{f} = (\mathbf{f}_1, \mathbf{f}_2)$. Estimated risk premia based on cross-sectional regression. t-statistics computed using Shanken adjustment. Note in view of example 2.10 that only one of $(\mathbf{f}_1, \mathbf{f}_2)$ is necessary. Results based on $N_{\text{sim}} = 10,000$ simulation runs.

Despite being redundant, the distribution of Shanken-adjusted t-statistics for factor 2 is not centered at zero. This can be seen in figure 1. Furthermore, the t-test rejects at a probability significantly higher than the nominal size α of the test for standard values of α. This can be seen in table D1.

Nominal size α:	Rejection probability:
10%	36.5%
5%	26.0 %
1 %	11.3%

Table D1: Rejection probabilities of the second factor in the two-factor simulation for factor $\mathbf{f}_2$.

Simulated distribution of test statistic $K_T^{(1)}$:

We perform simulations of the kernel-orthogonality test based on example 2.9. We assume $k = 2$, $\boldsymbol{\beta} = \mathbf{1} \in \mathbb{R}^n$, $T = 200$, and that both factors and idiosyncratic shocks to returns are generated as iid standard Normal random variables. Note that in this setting, $\operatorname{rank}\left(\overline{\boldsymbol{\Psi}}^{(1)}\right) = \operatorname{rank}(\mathbf{B}) = 1$ so we have one-dimensional underidentification. We apply our test to the risk premium of factor 1 which is point identified.

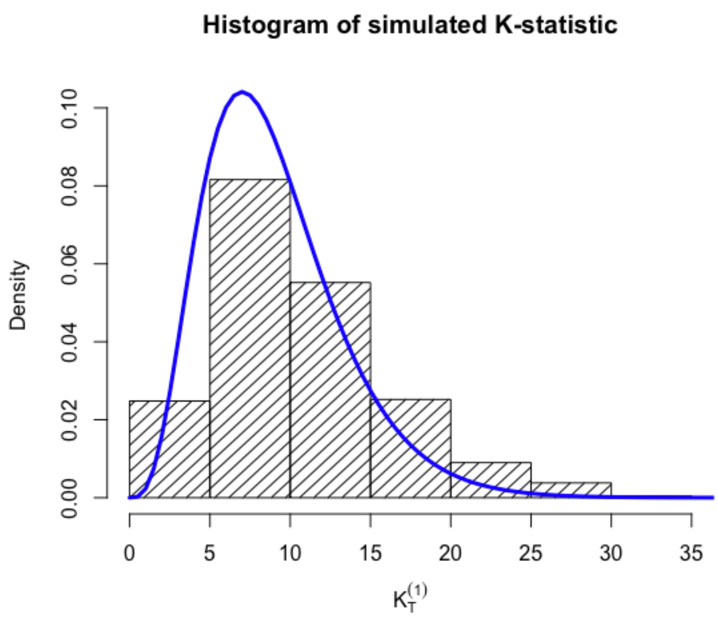

Figure D1: Simulated distribution of test statistic $K_T^{(1)}$ for the kernel-orthogonality test with one-dimensional underidentification under null hypothesis. $T = 200$, $n = 9$, $k = 2$, $N_{sim} = 1000$. Asymptotic χ^2 density curve shown in blue. $\mathbf{\Omega}^B$ estimated via bootstrapping with $N_b = 1000$.