Burnside, C. (2016) Identification and inference in linear stochastic discount factor models with excess returns. *Journal of Financial Econometrics*, 14(2), pp. 295-330. (doi:10.1093/jjfinec/nbv018)

Identification and Inference in Linear Stochastic Discount Factor Models with Excess Returns*

Craig Burnside†

June 2015

Abstract

When excess returns are used to estimate linear stochastic discount factor (SDF) models, researchers often adopt a normalization of the SDF that sets its mean to 1, or one that sets its intercept to 1. These normalizations are often treated as equivalent, but they are subtly different both in population, and in finite samples. Standard asymptotic inference relies on rank conditions that differ across the two normalizations, and which can fail to differing degrees. I first establish that failure of the rank conditions is a genuine concern for many well known SDF models in the literature. I also describe how failure of the rank conditions can affect inference, both in population and in finite samples. I propose using tests of the rank conditions not only as a diagnostic device, but also for model reduction. I show that this model reduction procedure has desirable properties in a Monte Carlo experiment with a calibrated model.

J.E.L. Classification: C3, G12
Keywords: GMM, normalizations, rank conditions, spurious factors, macroeconomic factors.

*This is a substantially revised version of an earlier draft entitled "Identification and Inference in Linear Stochastic Discount Factor Models". I am grateful to Jeremy Graveline, Cosmin Ilut, Shakeeb Khan, Frank Kleibergen, Francisco Peñaranda, Cesare Robotti and three anonymous referees for their comments and suggestions.

†Duke University, University of Glasgow and NBER

Standard asset pricing theory implies that if there are no arbitrage opportunities available in a set of assets, then there exists a stochastic discount factor (SDF), m, such that

$$E(R^e m) = 0, \tag{1}$$

where R^e is an $n \times 1$ vector of excess returns on the assets.

In empirical asset pricing, a common approach is to specify the SDF as a linear function of a $k \times 1$ vector of risk factors, f, with $k < n$. To take an example, suppose that we write $m = 1 - (f - \mu_f)'\gamma$, where $\mu_f = E(f)$ and γ is a conformable vector of parameters. Notice that with this specification of the SDF we can rewrite equation (1) as

$$E(R^e) = \text{cov}(R^e, f)\gamma. \tag{2}$$

Since $E(R^e f') = \text{cov}(R^e, f) + E(R^e)\mu_f'$ we can also rewrite equation (1) as

$$E(R^e) = E(R^e f')\delta \qquad \text{with} \qquad \delta \equiv \frac{\gamma}{1 + \mu_f'\gamma}. \tag{3}$$

Another way we can get to equation (3) directly is to start by writing the SDF as $m = 1 - f'\delta$ and then substitute this expression into equation (1).

The simple transformation involved in moving from equation (2) to equation (3) seems innocuous. However, there are important issues involving model validity and identification that emerge upon closer inspection. These arise in population, and have non-trivial consequences for estimation and inference.

To highlight the issues, consider a candidate model with a scalar risk factor, f. Suppose that $E(R^e) \neq 0$ and $\text{cov}(R^e, f) = 0$. Clearly, there is no value of γ such that equation (2) holds. Put differently, not only does the candidate model appear to be false, by reference to equation (2), but the parameter γ is also not identified by equation (2). At the same time, $E(R^e f) = E(R^e)\mu_f$, so, as long as $\mu_f \neq 0$, there is a simple solution to equation (3): $\delta = 1/\mu_f$. In this case, the candidate model appears to be valid, by reference to equation (3), and the parameter δ is well identified.

The scalar example is also useful for illustrating the symmetric case where $E(R^e f) = 0$, combined with $E(R^e) \neq 0$. Here, there is no value of δ such that equation (3) holds. Put differently, not only does the candidate model appear to be false, by reference to equation (3), but the parameter δ is also not identified by equation (3). At the same time, $\text{cov}(R^e, f) = -E(R^e)\mu_f$ and, as long as $\mu_f \neq 0$, there is a simple solution to equation (2): $\gamma = -1/\mu_f$.

In this case, the candidate model appears to be valid, by reference to equation (2), and the parameter γ is well identified.

An obvious question arises in these examples. Is the model valid or not? In each case there is one representation of the model that correctly assigns a zero price to the vector R^e. For this reason, we will say that, strictly speaking, these representations are *valid*. Nonetheless, it is worth digging a little deeper. At first there appears to be perfect symmetry between the two cases, but there is at least one dimension in which the symmetry is imperfect. In the second example, with $E(R^e f) = 0$, the solution to equation (2) implies an SDF with a positive (unit) mean, $m = f/\mu_f$, that has non-zero covariance with the return vector. On the other hand, in the first example, with $\text{cov}(R^e, f) = 0$ the solution to equation (3) implies an SDF, $m = 1 - f/\mu_f$, with mean zero and no covariance with the return vector. There are at least three reasons a researcher may find the zero-mean SDF dissatisfying:

1. Because $\text{cov}(R^e, m) = 0$ the aggregate measure of risk has a relationship with returns that is arguably devoid of economic interpretation.

2. Although equation (1) makes reference to excess returns, it is clear that an SDF with a zero mean cannot price a risk free asset correctly. Therefore, if it prices excess returns correctly it will price all gross returns incorrectly. This flaw cannot be remedied by a linear transformation of the SDF since only a proportional transformation would keep equation (1) intact, and would not change the mean of the SDF.

3. When the SDF is mean zero it implicitly assigns zero or negative prices to contingent claims on some–presumably many–states of the world. Although these claims may not be available to agents, this does raise issues as to whether the SDF can accurately price assets not included in the vector being studied by the researcher. Hansen and Jagannathan (1997) offer a similar reason for sometimes imposing the restriction that the SDF is nonnegative everywhere.

A central question in this paper is: If we want to exclude (reject) models of this type, how might we do so?

When f is a vector we have the more general case. Standard methods of estimating and evaluating linear SDF models rely on $\text{cov}(R^e, f)$ and $E(R^e f')$ having full column rank; i.e. rank k. When $\text{cov}(R^e, f)$ has rank less than k, γ is not identified, and estimators based on

(2) have nonstandard asymptotic distributions. When $E(R^e f')$ has rank less than k, δ is not identified, and estimators based on (3) have nonstandard asymptotic distributions.

Analogous to the scalar case discussed above, when $\text{rank}[\text{cov}(R^e, f)] < \text{rank}[E(R^e f')]$ I prove that there is no solution to (2) but, as in the scalar case, there is always at least one solution to equation (3). Any such solution has the property that $E(m) = 0$. Additionally, m is a linear combination of f that has no covariance with R^e.

In the symmetric case where $\text{rank}[\text{cov}(R^e, f)] > \text{rank}[E(R^e f')]$ I prove that there is no solution to (3) but there is a always a solution to (2) that, as in the scalar case, corresponds to an SDF with no intercept and a unit mean.

To show that failure of the rank conditions is not simply a theoretical curiosity, but is a practical reality, I use rank tests proposed by Cragg and Donald (1997) [discussed in Wright (2003)] and Kleibergen and Paap (2006). Consider Table 1, which shows the results of rank tests for several well-known models in the literature. For the CAPM and Fama and French (1993) three factor model, the null hypothesis of reduced rank is strongly rejected. However, for models with macroeconomic factors, the null hypothesis of reduced rank is not rejected in many cases. This suggests, at a minimum, that these models are poorly identified. But it is also noteworthy that in several cases the rank of $\text{cov}(R^e, f)$ appears to be less than the rank of $E(R^e f')$. For example, consider the model from Yogo (2006). This model includes three factors, but the tests hint that $\text{rank}[\text{cov}(R^e, f)] = 1$ and $\text{rank}[E(R^e f')] = 2$. For the Lustig and Verdelhan (2007) model, which also has three factors, the tests hint that $\text{rank}[\text{cov}(R^e, f)] = 0$ and $\text{rank}[E(R^e f')] = 1$. This evidence suggests that if there are parameterizations of these models that correctly price the assets, it is only because they put all of their emphasis on linear combinations of the risk factors that are uncorrelated with returns.

In Section 5 I use Monte Carlo evidence to show that rank tests can be an effective diagnostic.[1] In my calibrated models, the null hypothesis that $\text{rank}[\text{cov}(R^e, f)]$ is less than its true rank is always rejected. On the other hand, the null hypothesis that $\text{rank}[\text{cov}(R^e, f)]$ is equal to its true rank is rejected with a probability that converges, as the sample size grows, to the size of the test. The latter finding is simply confirmation of known properties of these tests.

[1] Equivalently, Kan and Zhang (1999b) suggest that researchers should test for significant spread among the factor betas, $\text{cov}(R^e, f)\Sigma_f^{-1}$. They point out that while Chen, Roll, and Ross (1986) and Ferson and Harvey (1993) performed such tests, relatively few researchers do so.

But what should an applied researcher do if the diagnostics suggest that $\text{rank}[\text{cov}(R^e, f)] < k$ or $\text{rank}[E(R^e f')] < k$? Is it preferable to use the mean-normalization over the intercept-normalization? My Monte-Carlo evidence suggests no simple advice of this type. Inference with either normalization is affected by identification problems, although often in different ways. Instead of favoring one normalization over another, I propose using rank tests as part of a model reduction procedure. This procedure is designed to eliminate models dominated by "useless" factors. Described intuitively, if a rank test suggests that $\text{rank}[\text{cov}(R^e, f)] = r < k$, the procedure finds r linear combinations of the factors, denoted $\tilde{f}_r$, for which the columns of $\text{cov}(R^e, \tilde{f}_r)$ come closest to spanning the columns of $\text{cov}(R^e, f)$. The appropriate linear combinations are a by-product of the Kleibergen and Paap (2006) rank test. The model is estimated and tested using the vector $\tilde{f}_r$ as risk factors. The economic meaning of the resulting parameter estimates can easily be recovered using the linear transformation that maps from f to $\tilde{f}_r$. My Monte Carlo evidence suggests that this procedure does not induce a significant size distortion when testing valid models.

An extensive literature, related to this paper, examines the properties of asset pricing tests in the presence of spurious risk factors; i.e. factors that are uncorrelated with the returns. Kan and Zhang (1999a) examine the behavior of GMM estimators, when the estimated SDF includes a spurious factor. In some cases, this factor is added to the true model. In other cases, it is added to a model that includes only some or none of the true risk factors. Their results relate to mine for the mean-normalization, because they study risk factors that are mean zero by construction. Kan and Zhang (1999b) study similar issues in the context of the two-pass approach to model evaluation. Kleibergen (2009) finds that standard inference based on the Fama and MacBeth (1973) and GLS two-pass regression methods is inadequate under partial or weak identification of a factor model. He suggests adoption of alternative statistics that lead to inference that is centered around the maximum likelihood estimator of Gibbons (1982). Kleibergen (2009) shows that the basic problem with the traditional methods is the dependence across the moment conditions that identify the betas and the moment conditions that identify risk premia. By rewriting the model in terms of betas of returns relative to a benchmark asset, he gets rid of this dependence, which leads to improved statistical performance. Kleibergen and Zhan (2013) use these novel statistics to show that the confidence sets for a number of models based on macroeconomic factors are, effectively, unbounded. The method proposed here, by contrast, stays within the realm of traditional methods by eliminating spurious linear combinations of factors from the model.

Another closely related paper is Kan and Robotti (2008). They examine the behavior of the Hansen and Jagannathan (1997) (HJ) distance measure under the two model normalizations discussed here. They show that the traditional HJ statistic can be manipulated by affine transformations of the factors when the intercept-normalization is adopted. They suggest a modified statistic that imposes the mean-normalization, and uses the covariance matrix of returns, not the cross moment matrix, as the weighting matrix. They also discuss how estimation of the SDF representation is affected by the choice of normalization, including a discussion of misspecification, but their analysis presumes the model is identified. Related papers by Gospodinov, Kan, and Robotti (2012) and Gospodinov, Kan, and Robotti (2013a) develop a unified framework for model estimation, evaluation and model comparison based on the unconstrained HJ-distance. The effects of model misspecification on estimation and inference are also discussed by Hou and Kimmel (2006), Shanken and Zhou (2007), Kan and Robotti (2009), and Kan, Robotti, and Shanken (2013). Some of this literature proposes using misspecification-robust inference. However, this approach assumes that models are properly identified. Gospodinov, Kan, and Robotti (2013b) extends the results regarding HJ distance into the realm of under-identified models. They also propose a sequential model selection device where individual factors are dropped sequentially based on robust t-statistics. Bryzgalova (2014) proposes a shrinkage-based framework for eliminating individual (or multiple) spurious factors. In her framework, estimation and model reduction is done in one step, rather than sequentially. In contrast to these two papers, the method proposed here focuses on problematic linear combinations of the factors. Gospodinov, Kan, and Robotti (2014) is also related, in that the authors relate model validity and rank conditions.

The paper is organized as follows. Section 1 uses a simple example with one risk factor and one asset to lay the groundwork for the rest of the paper. Section 2 discusses issues of SDF model validity, identification, and misspecification and derives the main theoretical results. Section 3 discusses the approaches I use to estimate the two model normalizations, to test for their identification, and to potentially reduce the dimension of the model. Section 4 discusses empirical findings for some models in the literature and links differences (across normalizations) in qualitative findings to failure of rank conditions. Section 5 performs a series of Monte Carlo simulation exercises that demonstrate the consequences of failure of the rank conditions in samples similar in size to those being studied in the literature. I discuss the behavior of parameter estimates, OIR tests, and my model selection procedure. Section 6 concludes.

1 An Illustrative Scalar Example

Many of the issues discussed in this paper can be understood via the following simple example. In this example, a researcher's goal is to find an SDF for the excess return on some asset, denoted R^e. To make things concrete, we will assume that $E(R^e) > 0$ for this single asset. The researcher proposes the following model of the SDF:

$$m = a - fb, \tag{4}$$

where a and b are scalars and f is a scalar risk factor. The proposed SDF prices the asset if

$$E(R^e m) = E[R^e(a - fb)] = 0. \tag{5}$$

Clearly, any (a, b) on the locus

$$E(R^e f)b = E(R^e)a, \tag{6}$$

works. I refer to this locus as the *pricing locus.* I also refer to any SDF corresponding to a pair (a, b) on this locus as *valid.*

Figures 1–4 illustrate this example under different assumptions about the joint behavior of f and R^e.[2] Each figure illustrates possible parameter pairs $(a, b) \in \mathbb{R}^2$. The parameter space is divided, for illustration, into two regions. The non-shaded region corresponds to those values of (a, b) such that $E(m) > 0$, which we denote $\mathcal{M}_+$. The shaded region corresponds to those values of (a, b) such that $E(m) < 0$, which we denote $\mathcal{M}_-$. Finally, the *boundary* between the regions is the locus $a = b\mu_f$. On this locus, which we denote $\mathcal{M}_0$, $E(m) = 0$. Every example assumes that $\mu_f > 0$ so that this locus is the upward sloping line $b = a/\mu_f$.

I consider two generic and two special cases of the pricing locus, which is illustrated in each figure as a solid line:

1. $E(R^e)/E(R^e f) < 1/\mu_f$. In this case, the pricing locus is either upward sloping, but with a smaller slope than the boundary, or it is downward sloping. The distinction between these cases is not important, so Figure 1 uses the latter case for clarity.

[2] Related graphs appear in the appendix of Peñaranda and Sentana (2015).

2. $E(R^e)/E(R^e f) > 1/\mu_f$. In this case, the pricing locus is upward sloping and steeper than the boundary. This case is illustrated in Figure 2.

3. The knife-edge case where $E(R^e)/E(R^e f) = 1/\mu_f$. This case occurs if $\text{cov}(R^e, f) = 0$, and implies that the pricing locus and the boundary coincide. This case is illustrated in Figure 3.

4. Another interesting knife-edge case happens when $E(R^e f) = 0$. This means the pricing locus is vertical at $a = 0$. This case is illustrated in Figure 4.

Given that the pricing locus is not a single point, our researcher adopts a normalization of the SDF. There are two commonly used alternatives. The first of these, the intercept-normalization, assumes that $a = 1$. To illustrate this normalization, each figure includes a dashed line corresponding to the locus $(1, b)$. Alternatively, the mean-normalization assumes that the mean of the SDF is 1. Consistent with this, each figure includes a dotted line $b = (a - 1)/\mu_f$ along which all candidate SDFs have a unit mean. This line is parallel to the boundary. Having adopted one or other of these normalizations, there is at most one solution to the pricing equation.

Case 1. In Figure 1, the pricing line crosses both the dashed and dotted lines (at A and B). Therefore, there exist valid unit-intercept and unit-mean SDFs, and both lie in $\mathcal{M}_+$. This is a favorable case, for several reasons. In both cases, the model is identified. When we turn to finite sample estimation, inference functions as intended. Parameter estimates are consistent and have standard asymptotic distributions. If multiple assets are available, the OIR test functions as intended. And the SDFs at A and B are proportional by a positive constant.

Case 2. In Figure 2 we see that both normalizations are identified. The pricing line crosses the dotted line (at B), so there exists a valid unit-mean SDF in $\mathcal{M}_+$. It crosses the dashed line (at A), so the unique valid unit-intercept SDF lies in $\mathcal{M}_-$.

While, strictly speaking, the SDF at A is valid, because it solves the pricing equation, it has three potentially undesirable features. First, because it implies $E(m) < 0$ it obviously prices risk free assets incorrectly. Second, since the mean of m is negative, this implies that there are states of the world to which this SDF assigns negative state prices. Third, the relationship between m and f is positive $(b > 0)$ whereas $b < 0$ for all valid SDFs that lie in $\mathcal{M}_+$. Obviously a researcher paying attention to the fact that $E(m) < 0$ could flip the sign of this SDF, and resolve all of these issues. Of course, doing so would, in effect, imply adopting a different normalization.

Case 3. When $\text{cov}(R^e, f) = 0$ we have Figure 3. Here, the pricing line is the same as the boundary, $\mathcal{M}_0$. Therefore, there are no valid (a, b) pairs in $\mathcal{M}_+$ nor in $\mathcal{M}_-$. The pricing line is parallel to the dotted line, so there exists no valid unit-mean SDF in this case. The pricing/boundary line crosses the dashed line at A (where $b = 1/\mu_f$) so the valid unit-intercept SDF has a zero mean.

The fact that there is no solution to the pricing equation for the mean-normalization means that the model is not identified. In population the researcher would be able to reject the model, since nonexistence is an analytic result. However, the identification problem manifests itself in non-standard inference in finite samples, even asymptotically.

On the other hand, if the researcher adopts the intercept-normalization, the model is identified, in the sense that there is a unique solution to the pricing equation. Parameter estimates are consistent and have standard asymptotic distributions. If multiple assets are used to test the model, it is rejected with probability equal to the size of the test. In other words, the model is likely to end up being validated by standard inference.

Should the model be rejected in Case 3? This is a matter of interpretation. Clearly, the model at A is potentially problematic. It implies $E(m) = 0$, so it prices risk free assets incorrectly, and must assign negative state-prices to some states of the world. There is also a sense in which it is economically uninteresting, since it implies zero covariance between the measure of aggregate risk, m, and the returns being priced. At the same time, the model at A does price R^e correctly.

Case 4. When $E(R^e f) = 0$ the pricing equation is the vertical line $a = 0$, as illustrated in Figure 4. The pricing line crosses the dotted line at B (where $b = -1/\mu_f$), so there exists a valid unit-mean SDF. The pricing line does not cross the dashed line, so there is no valid unit-intercept SDF.

Econometrically, Case 4 is symmetric to Case 3 with, in this case, point B being defined, but point A not being defined. In other words, one normalization (in this case the mean-normalization) is identified, while the other (the intercept-normalization) is not.

However, Case 4 is economically different from Case 3. In Case 4, the identified model at B has a positive mean, whereas, in Case 3 the model at A does not. The model at B in Case 4 does have one perplexing property. Peñaranda and Sentana (2015) argue that an "SDF that is exactly proportional to an orthogonal factor is not very attractive from an economic point of view". They argue that this is because the projection of f onto R^e with no constant—in other words, the uncentered factor mimicking portfolio—is constant and equal to 0. This property follows from equation (5) because $a = 0$ and $m = -fb$, implying that $E(R^e f) = 0$. This property of the SDF may be unattractive, but it must be kept in mind that in every case described above (i.e. cases 1—4), the SDF is orthogonal to R^e so an uncentered *SDF*-mimicking portfolio is *always* constant and equal to 0.[3] Additionally, if the feature that $E(R^e f) = 0$ is viewed as problematic, it can be remedied by a linear transformation. On the other hand, in Case 3, linear transformation of the factor cannot change the fact that the covariance between the factor and the return is zero.

Cases 1—4 illustrate many of the issues dealt with later in this paper, however, the more general setting of a multi-factor model is more complicated. In a setting with k factors, the intercept and mean-normalizations are identified if, respectively, $\text{rank}[E(R^e f')] = k$ and $\text{rank}[\text{cov}(R^e, f)] = k$. If both these rank conditions hold, we have a situation analogous to Cases 1 and 2. When $\text{rank}[E(R^e f')] = k$ and $\text{rank}[\text{cov}(R^e, f)] = k - 1$ we have a situation analogous to Case 3. When $\text{rank}[E(R^e f')] = k - 1$ and $\text{rank}[\text{cov}(R^e, f)] = k$ we have a situation analogous to Case 4. But we also have an additional possibility that both models are under-identified. This possibility can only arise in the scalar case if we have $\text{cov}(R^e, f) = 0$ and $\mu_f = 0$, in which case we also have $E(R^e f) = 0$.

2 Validity, Misspecification and Identification

Let R_t^e be an $n \times 1$ vector whose ith element, R_{it}^e, is the excess return at time t to asset i defined as the difference between the return on asset i and the risk free rate. For notational convenience I suppress time subscripts unless they are strictly needed.

[3] I am grateful to Francisco Peñaranda for pointing out the following fact. It is always possible to mechanically construct an SDF or risk factor as the residual of the least squares projection of any random variable x onto R^e (without a constant). In fact, any such residual will satisfy the pricing equation for returns. I would argue that this is not a problem with the mean-normalization, per se. Rather, it is a problem endemic to working with excess returns. It also does not imply that every factor for which $E(R^e f') = 0$ is constructed in an economically meaningless way, because $E(R^e m) = 0$ holds for *any* valid SDF.

Assumption 1. At least one element of the return vector has a non-zero mean; that is $E(R^e) \neq 0$.

Under the assumption of no-arbitrage there exists a strictly positive random variable, m, such that

$$E(R^e m) = 0. \tag{7}$$

We consider candidate models of the SDF, m, that take the following form:

$$m = a - f'b, \tag{8}$$

where f is a $k \times 1$ vector of risk factors, a is a scalar constant and b is a $k \times 1$ vector of parameters. We will assume, throughout, that $n \geq k$. Given the vector of risk factors, we let $\mathcal{M}$ denote the set of all m of the form given in equation (8). We let $\mathcal{M}_+ = \{m \in \mathcal{M} | E(m) > 0\}$, $\mathcal{M}_- = \{m \in \mathcal{M} | E(m) < 0\}$, and $\mathcal{M}_0 = \{m \in \mathcal{M} | E(m) = 0\}$.

2.1 Valid Models of the SDF

Definition 1. m is a *valid* SDF if equation (7) holds for at least one (a, b).

Clearly, if m is valid for some (a, b) so is κm for any scalar, κ, since this transformation preserves that equation (7) holds. Thus, validity of the SDF does not uniquely determine (a, b). For this reason, it is common to adopt a normalization of the SDF that reduces the dimension of the parameter space. We consider two normalizations that appear frequently in the literature.

To arrive at the first normalization we rewrite equation (8) as

$$m = a(1 - f'\delta),$$

with $\delta = b/a$. I refer to

$$m^\delta \equiv 1 - f'\delta, \tag{9}$$

as the *intercept-normalization* of m since it is a scaled version of m with a unit intercept. Clearly, equation (7) implies that

$$E(R^e) = E(R^e f')\delta \qquad \text{or} \qquad \mu_R = D\delta. \tag{10}$$

where $\mu_R \equiv E(R^e)$ and $D \equiv E(R^e f')$. Given definition 1, m^δ is a valid SDF if equation (10) holds for at least one δ.

Alternatively (8) can be rewritten as

$$m = \xi[1 - (f - \mu_f)'\gamma],$$

where $\mu_f = E(f)$, $\xi = a - \mu_f' b$, and $\gamma = b/\xi$. I refer to

$$m^\gamma \equiv 1 - (f - \mu_f)'\gamma, \tag{11}$$

as the *mean-normalization* since it is a scaled version of m with a unit mean. Given this normalization equation (7) implies that

$$E(R^e) = E[R^e(f - \mu_f)']\gamma \qquad \text{or} \qquad \mu_R = C\gamma, \tag{12}$$

where $C \equiv E[R^e(f - \mu_f)'] = \text{cov}(R^e, f)$. Given definition 1, m^γ is a valid SDF if equation (12) holds for at least one γ.

2.2 Misspecified Models of the SDF

A model, m, is misspecified if there is no (a, b) such that (7) holds. Because the two normalizations, m^δ and m^γ, are nested within m, it follows that

$$m^\delta \,\text{misspecified} \Longleftarrow m \,\text{misspecified} \Longrightarrow m^\gamma \,\text{misspecified}.$$

Notice that the arrows only run in one direction. Case 3, above, is an example in which m^γ is misspecified, but m and m^δ are valid. Similarly, Case 4, above, is an example in which m^δ is misspecified, but m and m^γ are valid.

When a model is misspecified, it is useful to consider linear combinations of the moment restrictions that appear in equations (10) and (12). In particular, we consider the following equations

$$D'W^\delta[E(R^e) - E(R^e f')\delta] = 0, \qquad \text{or} \qquad D'W^\delta(\mu_R - D\delta) = 0, \tag{13}$$

$$C'W^\gamma\{E(R^e) - E[R^e(f - \mu_f)']\gamma\} = 0, \qquad \text{or} \qquad C'W^\gamma(\mu_R - C\gamma) = 0. \tag{14}$$

Here, W^δ and W^γ are assumed to be $n \times n$ positive definite symmetric matrices. Equations (13) and (14) are of interest because they are population equivalents of the equations we use below to define GMM estimators of δ and γ. Solutions to these equations exist even if solutions to equations (10) and (12) do not, but they may or may not be unique.

2.3 Rank Conditions and Identification

In this section, we establish important relationships between identification, which is stated in terms of the ranks of the matrices C and D, the model's validity and the mean of the SDF. We use the following notation: $r_D \equiv \text{rank}(D)$ and $r_C \equiv \text{rank}(C)$. We let $X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}$ denote a $k \times k$ orthogonal matrix whose first r_D columns, X_1, span the rowspace of D [denoted $\mathcal{R}(D)$], and whose remaining columns, X_2, span the nullspace of D [denoted $\mathcal{N}(D)$]. We let $Y = \begin{pmatrix} Y_1 & Y_2 \end{pmatrix}$ denote a $k \times k$ orthogonal matrix whose first r_C columns, Y_1, span the rowspace of C [denoted $\mathcal{R}(C)$], and whose remaining columns, Y_2, span the nullspace of C [denoted $\mathcal{N}(C)$].

Definition 2. The SDF m^δ is *identified* if $r_D = k$. The SDF m^γ is *identified* if $r_C = k$.

Theorem 1. *A solution to equation* (10)*, if it exists, is unique if and only if* m^δ is *identified. A solution to equation* (12)*, if it exists, is unique if and only if* m^γ *is identified.*

Proof. Let δ_0 be a solution to (10) so that $\mu_R = D\delta_0$. First, assume that $r_D = k$. Assume there is another parameter vector $\delta_1 \neq \delta_0$ such that (10) holds. Then $\mu_R = D\delta_1$. But this implies that $D(\delta_0 - \delta_1) = 0$. Because $r_D = k$ this implies that $\delta_1 = \delta_0$, which is a contradiction. Now assume that $r_D < k$ with δ_0 defined as before. We can construct another solution $\delta_1 = \delta_0 + x$ for any $x \neq 0$ such that $Dx = 0$. Because $r_D < k$, we know such an x exists. The same argument applies to the second part of the lemma. □

Theorem 2. *There exists a solution to equation* (13)*, which is unique if and only if* m^δ *is identified. There exists a solution to equation* (14)*, which is unique if and only if* m^γ *is identified.*

Proof. Because W^δ is positive definite, $r_D = \text{rank}(D'W^\delta D)$. So when $r_D = k$, standard arguments imply that $\delta = (D'W^\delta D)^{-1}D'W^\delta\mu_R$ is the unique solution to equation (13). When D has reduced rank, there are many solutions to equation (13). Consider, first, the case where $r_D = 0$. In this case, $D = 0$ and any $\delta \in R^k$ is a solution to (13). If $1 \leq r_D < k$, define $\tilde{D} = DX_1$ and let $\tilde{\delta} = (\tilde{D}'W^\delta\tilde{D})^{-1}\tilde{D}'W^\delta\mu_R$. Then, as shown in the extended proof in the Appendix, $\delta = X_1\tilde{\delta} + x$ is a solution to (13) for any $x \in \mathcal{N}(D)$. A similar argument applies to the second statement in the theorem. □

A consequence of Theorems 1 and 2 is that estimators based on the sample equivalents of (13) and (14) have standard properties if D and C have full rank. If the underlying moment restrictions, (10) and (12), hold, theorems from Hansen (1982) apply. If these moment conditions are not satisfied, theorems for misspecified models from Hall and Inoue (2003) and Hall (2005) apply. I focus, instead, on situations where $r_D < k$ or $r_C < k$.

Lemma 1. *The matrices C and D differ in rank by at most 1, i.e. $|r_C - r_D| \leq 1$.*

Proof. By definition $D = C + \mu_R \mu_f'$. Since μ_R and μ_f are vectors, $\mu_R \mu_f'$ has, at most, unit rank. Consequently, $r_D \leq r_C + 1$, and $r_C \leq r_D + 1$. □

Lemma 2. *$r_C < r_D$ if and only if $\mu_f \notin \mathcal{R}(C)$ and $\mu_f \in \mathcal{R}(D)$. $r_C > r_D$ if and only if $\mu_f \notin \mathcal{R}(D)$ and $\mu_f \in \mathcal{R}(C)$. $r_C = r_D$ if and only if $\mu_f \in \mathcal{R}(C)$ and $\mu_f \in \mathcal{R}(D)$.*

Proof. Let $C = U_C S_C Y'$ and $D = U_D S_D X'$ denote the singular value decompositions of C and D. Allowing for the possibility that C and D have less than full rank we can equivalently write these as

$$C = U_{C1} S_{C1} Y_1' \qquad D = U_{D1} S_{D1} X_1'$$

where U_{C1} represents the first r_C columns of U_C, S_{C1} is the upper left $r_C \times r_C$ corner of S_C and Y_1 is defined as above. U_{D1}, S_{D1} and X_1 are defined similarly. We can then write

$$D = C + \mu_R \mu_f' = U_{C1} S_{C1} Y_1' + \mu_R \mu_f'$$

$$C = D - \mu_R \mu_f' = U_{D1} S_{D1} X_1' - \mu_R \mu_f'.$$

We first establish the ifs. From the first equation it is clear that if $\mu_f \notin \mathcal{R}(C)$ the columns of Y_1 do not span the rows of D, but $(Y_1 \; \mu_f)$ does span the rows of D and has full column rank. Therefore, $r_D > r_C$ and $\mu_f \in \mathcal{R}(D)$. Similarly, if $\mu_f \notin \mathcal{R}(D)$ the columns of X_1 do not span the rows of C, but $(X_1 \; \mu_f)$ does span the rows of C. Therefore, $r_C > r_D$ and $\mu_f \in \mathcal{R}(C)$.

The first two results imply that the only remaining possibility is that $\mu_f \in \mathcal{R}(C)$ and $\mu_f \in \mathcal{R}(D)$. Given the above equations this means that the rows of D are in $\mathcal{R}(C)$ and the rows of C are in $\mathcal{R}(D)$. This implies that C and D have the same rank. Since we have considered all possible statements regarding the location of μ_f the only ifs have also been established. □

Theorem 3. *If m is valid and $m \notin \mathcal{M}_0$ then $r_C \geq r_D$.*

Proof. By definition $D = C + \mu_R \mu_f'$. When an SDF of the general form (8) is valid and has a non-zero mean we can always rewrite it in the form (11), so that equation (12) holds.

Hence, $\mu_R = C\gamma$ and so we can write

$$D = C + \mu_R \mu_f' = C(I_k + \gamma \mu_f').$$

It follows that $r_D \leq r_C$. □

The following corollary follows directly from Theorem 3:

Corollary 1. *If $r_C < r_D$ any valid m must lie in $\mathcal{M}_0$ (has a zero mean).*

Theorem 4. *If m is valid and $a \neq 0$ then $r_C \leq r_D$.* If, additionally, $m \notin \mathcal{M}_0$ *then* $r_C = r_D$.

Proof. By definition $C = D - \mu_R \mu_f'$. When an SDF of the general form (8) is valid and $a \neq 0$ the SDF can be written in the form (9) and equation (10) holds. Hence, $\mu_R = D\delta$ and so we can write

$$C = D - \mu_R \mu_f' = D(I_k - \delta \mu_f').$$

It follows that $r_C \leq r_D$. The final statement in the theorem follows from Theorem 3. □

The following corollary follows directly from Theorem 4:

Corollary 2. *If $r_C > r_D$ any valid m must have a zero intercept.*

Clearly, we can also combine the two theorems to get the following corollary:

Corollary 3. *If m is valid, $m \notin \mathcal{M}_0$ and $a \neq 0$ then $r_C = r_D$.*

Lemma 3. *If $r_C = r_D = 0$ and m is valid then $m \in \mathcal{M}_0$ and $a = 0$.*

Proof. When $r_C = r_D = 0$ this means $C = D = 0$. It follows from validity that $0 = E(R^e m) = \mu_R a - Db = \mu_R a$, implying $a = 0$. Similarly, $0 = E(R^e m) = \mu_R a - Cb - \mu_R \mu_f' b = -\mu_R \mu_f' b$, implying that $\mu_f' b = 0$. Hence, $E(m) = a - \mu_f' b = 0$. □

Figure 5 summarizes our results, so far, regarding the ranks of C and D. Each dot represents a combination of ranks that is mathematically possible; i.e. consistent with Lemma 1. The mean-normalization is identified if $r_C = k$, otherwise it is underidentified. The intercept-normalization is identified if $r_D = k$, otherwise it is underidentified. When $r_C < r_D$ (the shaded gray dots), the mean-normalization is underidentified and any valid m is mean zero. When $r_C > r_D$ (the striped dots), the intercept-normalization is underidentified and any valid m has $a = 0$. If $r_C = r_D = 0$ (the gray striped dots), then any valid m is mean zero and has $a = 0$. The solid black dots represent combinations of rank potentially consistent with validity, $E(m) \neq 0$ and $a \neq 0$. We turn, next, to an exploration of the consequences of underidentification.

2.4 Under-identification and the existence of valid SDFs

There are three cases to consider: (1) $r_D = r_C - 1$ with $r_C \le k$, (2) $r_C = r_D - 1$ with $r_D \le k$ and (3) $r_C = r_D < k$. We show results for each of these cases in turn.

Theorem 5. *If $r_D < r_C$ there exists a valid parameterization of the mean-normalization, but it has a zero intercept.*

Proof. Let $\tilde{\gamma}_2$ be a $(k - r_D) \times 1$ vector and choose it so that $(\mu_f' X_2)\tilde{\gamma}_2 = -1$. We know we can choose $\tilde{\gamma}_2$ in this way because Lemma 2 implies that $\mu_f \notin \mathcal{R}(D)$, so that $\mu_f' X_2 \neq 0$. Then let $\gamma = X_2 \tilde{\gamma}_2$. We then have

$$\mu_R - C\gamma = \mu_R - C X_2 \tilde{\gamma}_2 = \mu_R - (D - \mu_R \mu_f') X_2 \tilde{\gamma}_2.$$

Recall that $DX_2 = 0$ so that we have

$$\mu_R - C\gamma = \mu_R [1 + (\mu_f' X_2)\tilde{\gamma}_2] = 0.$$

The second statement in the theorem follows from Corollary 2. □

When $r_C = k$ the solution for γ described in the theorem is unique, otherwise it is not. There are two ways of getting to this result. First, we already know, from Theorem 1, that when $r_C = k$ any solution to $\mu_R = C\gamma$ is unique. Second, when $r_D < r_C = k$, Lemma 1 implies that $r_D = k - 1$ so that $\tilde{\gamma}_2 = -1/(\mu_f' X_2)$ is a unique scalar.[4]

Theorem 6. *If $r_C < r_D$ there exists a valid parameterization of the intercept-normalization, but it lies in $\mathcal{M}_0$.*

Proof. The first part of the proof mimics the proof of Theorem 5. Let $\tilde{\delta}_2$ be a $(k - r_C) \times 1$ vector and choose it so that $(\mu_f' Y_2)\tilde{\delta}_2 = 1$. We know we can choose $\tilde{\delta}_2$ in this way because Lemma 2 implies that $\mu_f \notin \mathcal{R}(C)$, so that $\mu_f' Y_2 \neq 0$. Then let $\delta = Y_2 \tilde{\delta}_2$. We then have

$$\mu_R - D\delta \;\; = \;\; \mu_R - D Y_2 \tilde{\delta}_2 = \mu_R - (C + \mu_R \mu_f') Y_2 \tilde{\delta}_2.$$

Recall that $CY_2 = 0$ so that we have

$$\mu_R - D\delta = \mu_R [1 - (\mu_f' Y_2)\tilde{\delta}_2] = 0.$$

The second statement in the theorem follows from Corollary 1. □

[4] We can ensure that $\tilde{\gamma}_2$ is unique, in this case, by normalizing the first non-zero element of the vector X_2 to be positive. Regardless, $\gamma = X_2 \tilde{\gamma}_2$ is always unique, in this case, because a sign change in X_2 translates to a sign change in $\tilde{\gamma}_2$ so these sign changes cancel out in γ.

It is worth noting that the solution for δ described in the proof has a counterintuitive economic interpretation: The SDF puts zero weight on the linear combinations of the risk factors that are correlated with returns, and all its weight on linear combinations of the risk factors that are uncorrelated with returns:

$$\operatorname{cov}(R^e, m^\delta) = -\operatorname{cov}(R^e, f)\delta = -\operatorname{cov}(R^e, f)Y_2\tilde{\delta}_2 = -\underbrace{\operatorname{cov}(R^e, fY_1)}_{\neq 0}0 - \underbrace{\operatorname{cov}(R^e, fY_2)}_{=0}\tilde{\delta}_2 = 0.$$

When $r_D = k$ the solution for δ described in the theorem is unique, otherwise it is not. We already know, from Theorem 1, that when $r_D = k$ any solution to $\mu_R = D\delta$ is unique. When $r_C < r_D = k$, Lemma 1 implies that $r_C = k - 1$ so that $\tilde{\delta}_2 = 1/(\mu_f' Y_2)$ is a unique scalar.[5]

Finally we have the case where $r_C = r_D < k$. In this case, neither normalization of the SDF is identified. If $r_C = r_D = 0$ we have $C = D = 0$, and there are no solutions to equations (10) and (12). The only valid m has the form $m = -f'b$ with $\mu_f' b = 0$ so it is both mean zero and has a zero intercept. When $0 < r_C = r_D < k$ there may or may not be solutions to equations (10) and (12).

3 Estimation and Inference using GMM

3.1 The intercept-normalization

Using the n moment restrictions given by (10), δ is estimated using GMM. Define $u_t^\delta(\delta) = R_t^e(1 - f_t'\delta)$ and let $g_T^\delta(\delta) = \frac{1}{T}\sum_{t=1}^T u_t^\delta(\delta) = \bar{R}^e - D_T\delta$ be an $n \times 1$ vector of pricing errors, where $\bar{R}^e = \frac{1}{T}\sum_{t=1}^T R_t^e$, $D_T = \frac{1}{T}\sum_{t=1}^T R_t^e f_t'$ and T is the sample size.

I consider GMM estimators that set $a_T^\delta g_T^\delta = 0$, where a_T^δ is a $k \times n$ matrix and takes the form $a_T^\delta = D_T' W_T^\delta$, where W_T^δ is an $n \times n$ positive definite weighting matrix. It follows that the GMM estimator of δ is

$$\hat{\delta} = \left(D_T' W_T^\delta D_T\right)^{-1} D_T' W_T^\delta \bar{R}^e. \tag{15}$$

In the first GMM step I let $W_T^\delta = I_n$. In step $j + 1$, I let $W_T^\delta = (S_T^\delta)^{-1}$ where $S_T^\delta = \frac{1}{T}\sum_{t=1}^T \hat{u}_{jt}^\delta \hat{u}_{jt}^{\delta\prime}$, $\hat{u}_{jt}^\delta = R_t^e(1 - f_t'\hat{\delta}_j)$ and $\hat{\delta}_j$ represents the jth-step estimator of δ.[6]

Let $\Delta_T^\delta = -D_T$. An OIR, or pricing error, test is based on the statistic

$$J^\delta = T g_T^\delta(\hat{\delta})'(\hat{V}_g^\delta)^+ g_T^\delta(\hat{\delta}), \tag{16}$$

[5] Uniqueness requires an argument analogous to the one provided in footnote (4).

[6] By computing S_T^δ in this way, I impose the theoretical restriction that $E_{t-j}(u_t^\delta) = 0$ for all $j \geq 1$.

where

$$\hat{V}_g^\delta = A_T^\delta S_T^\delta (A_T^\delta)' \quad \text{with} \quad A_T^\delta = I_n - \Delta_T^\delta (a_T^\delta \Delta_T^\delta)^{-1} a_T^\delta, \tag{17}$$

and $(\hat{V}_g^\delta)^+$ indicates the generalized inverse of the matrix $\hat{V}_g^\delta$.

3.2 The mean-normalization

Using the n moment restrictions given by (12) along with the moment condition $E(f_t - \mu_f) = 0$, γ and μ_f are estimated using GMM. I use the notation $\theta = (\begin{array}{cc} \gamma' & \mu_f' \end{array})'$ for the combined parameter vector. Define $u_{1t}^\theta(\theta) = R_t^e[1 - (f_t - \mu_f)'\gamma]$ and let $g_{1T}^\theta(\theta) = \frac{1}{T}\sum_{t=1}^T u_{1t}^\theta(\theta) = \bar{R}^e - (D_T - \bar{R}^e \mu_f')\gamma$. Define $u_{2t}^\theta(\theta) = f_t - \mu_f$ and let $g_{2T}^\theta(\theta) = \frac{1}{T}\sum_{t=1}^T u_{2t}^\theta(\theta) = \bar{f} - \mu_f$, where $\bar{f} = \frac{1}{T}\sum_{t=1}^T f_t$. Define $u_t^\theta = (\begin{array}{cc} u_{1t}^{\theta\prime} & u_{2t}^{\theta\prime} \end{array})'$ and $g_T^\theta = (\begin{array}{cc} g_{1T}^{\theta\prime} & g_{2T}^{\theta\prime} \end{array})'$.

I consider GMM estimators that set $a_T^\theta g_T^\theta = 0$, where a_T^θ is a $2k \times (n+k)$ matrix and takes the form

$$a_T^\theta = \begin{pmatrix} C_T' W_T^\theta & 0 \\ 0 & I_k \end{pmatrix}, \tag{18}$$

where $C_T = D_T - \bar{R}^e \bar{f}'$ and W_T^θ is an $n \times n$ positive definite weighting matrix. It follows that the GMM estimators of γ and μ_f are

$$\begin{aligned} \hat{\gamma} &= (C_T' W_T^\theta C_T)^{-1} C_T' W_T^\theta \bar{R}^e & (19) \\ \hat{\mu}_f &= \bar{f}. & (20) \end{aligned}$$

In the first GMM step I let $W_T^\theta = I_n$. In step $j+1$, I let $W_T^\theta = (P_T S_T^\theta P_T')^{-1}$ where $P_T = [\begin{array}{cc} I_n & \bar{R}^e \hat{\gamma}_j' \end{array}]$, $\hat{\gamma}_j$ represents the jth-step estimator of γ and S_T^θ is a consistent estimator of $S^\theta = \sum_{j=-\infty}^{+\infty} E(u_t^\theta u_{t-j}^{\theta\prime})$.[7] Because u_{2t}^θ may be serially correlated I use a VARHAC estimator, described in more detail in the Appendix, to compute S_T.[8]

Let

$$\Delta_T^\theta = \begin{pmatrix} -C_T & \bar{R}^e \hat{\gamma}' \\ 0 & -I_k \end{pmatrix}. \tag{21}$$

An OIR test is based on

$$J^\theta = T g_T^\theta(\hat{\theta})' (\hat{V}_g^\theta)^+ g_T^\theta(\hat{\theta}), \tag{22}$$

[7] The first step of the GMM procedure is numerically equivalent (in terms of pricing errors) to using the two-pass regression method and running the cross-sectional regression with no constant. In the later GMM steps, Cochrane (2005) suggests using the matrix $(\begin{array}{cc} I_n & 0_{n\times k} \end{array})$ in place of P_T in the expression for W_T^γ. This is less efficient in terms of the covariance matrix of $\hat{\gamma}$, but is asymptotically equivalent in terms of the OIR test.

[8] The VARHAC estimator that I use imposes the theoretical restriction that $E_{t-j}(u_{1t}^\theta) = 0$ for all $j \geq 1$. This restriction does not hold for u_{2t}^θ since the risk factors may be serially correlated. The VARHAC procedure whitens u_{2t}^θ using a VAR, but is otherwise identical to a standard HAC procedure.

where

$$\hat{V}_g^\theta = A_T^\theta S_T^\theta (A_T^\theta)' \quad \text{with} \quad A_T^\theta = I_{n+k} - \Delta_T^\theta (a_T^\theta \Delta_T^\theta)^{-1} a_T^\theta. \tag{23}$$

Yogo (2006) uses a different, optimal, GMM procedure in conjunction with the mean-normalization. Noting that the derivative of g_{1T}^θ with respect to μ_f is non-zero, he uses a variant of a_T^θ that is not block diagonal because this improves asymptotic efficiency. In his case $\hat{\mu}_f$ does not, in general, equal $\bar{f}$. As it turns out, in finite samples, the properties of Yogo's procedure are quite different than the properties of the procedure I have outlined here. Peñaranda and Sentana (2015) discuss Yogo's procedure and normalizations of the SDF in the context of continuously-updated (CU)-GMM estimators. I discuss the optimal-GMM and CU-GMM procedures in more detail in a separately available appendix.

3.3 Asymptotic properties of the estimators

Let r_C and r_D be defined as in Section 2. If $r_D = k$ inference is standard for the intercept-normalization. If there exists a solution, δ, to equation (10), it is unique. Under regularity conditions provided in Hansen (1982), $\hat{\delta} \stackrel{a.s.}{\to} \delta$ where δ is that unique solution. Also $\sqrt{T}(\hat{\delta} - \delta) \stackrel{d}{\to} N(0, V^\delta)$ with $V_\delta = (a^\delta \Delta^\delta)^{-1} a^\delta S^\delta a^{\delta\prime} [(a^\delta \Delta^\delta)']^{-1}$, where a^δ and Δ^δ are the probability limits of a_T^δ and Δ_T^δ. Finally, $J^\delta \stackrel{d}{\to} \chi^2_{n-k}$.

Similarly, if $r_C = k$ inference is standard for the mean-normalization. If there exists a solution, γ, to equation (12), it is unique. Under regularity conditions provided in Hansen (1982), $\hat{\theta} \stackrel{a.s.}{\to} \theta$, where $\theta = (\ \gamma' \quad \mu_f'\ )'$ and γ is the unique solution to equation (12). Also $\sqrt{T}(\hat{\theta} - \theta) \stackrel{d}{\to} N(0, V_\theta)$ with $V_\theta = (a^\theta \Delta^\theta)^{-1} a^\theta S^\theta a^{\theta\prime} [(a^\theta \Delta^\theta)']^{-1}$, where a^θ and Δ^θ are the probability limits of a_T^θ and Δ_T^θ. Finally, $J^\theta \stackrel{d}{\to} \chi^2_{n-k}$.

When the rank conditions are satisfied but there do not exist solutions to equations (10) and (12) the model is misspecified. We can refer to Hall (2005), Chapter 4 for the asymptotic properties of $\hat{\delta}$, J^δ, $\hat{\theta}$ and J^θ in this case.

When $r_D < k$ the intercept-normalization is under-identified. Therefore, $\hat{\delta}$ and J^δ have non-standard asymptotic distributions. Similarly, when $r_C < k$ the mean-normalization is under-identified. Therefore, $\hat{\gamma}$ and J^θ have non-standard asymptotic distributions.

Kan and Zhang (1999a) consider SDF models with factors that are assumed to be mean zero, so their findings are directly relevant for the mean-normalization. They derive the asymptotic properties of Wald tests applied to the elements of $\hat{\gamma}$ and the OIR test for the case where $r_C < k$. Assume that there is a valid SDF model with risk factors f_t. Suppose that a researcher estimates a model with (f_t, x_t) as the conjectured factors, where x_t is completely "useless" and mean zero.[9] This model is correctly specified, in the sense that it nests the true model, and it actually doesn't matter what coefficient ends up attached to the spurious factor, x_t. Because x_t is uncorrelated with R_t^e it has no effect on the pricing equation. On the other hand, suppose a researcher estimates a model with (f_{1t}, x_t) as the conjectured factors, where f_{1t} is an incomplete subset of f_t. In this case, the model is misspecified. The asymptotic properties of the tests studied by Kan and Zhang (1999a) depend on whether the model is correctly specified or misspecified. Additionally, the extent of the misspecification matters and how many steps are taken over the GMM weighting matrix also matters. For example, using a model design with two true factors, Kan and Zhang (1999a) show that a model that uses only the spurious factor x_t will be rejected with probability less than the size of the test when a 2nd-step GMM test is used, but with probability greater than 0.9 when a 3rd-step GMM test is used. On the other hand, when one of the true factors is included along with the spurious factor, the power of the 2nd-step GMM test improves, while the power of the 3rd-step GMM test does not.

One interesting case is when $r_C = k - 1$ and $r_D = k$. As we saw in Section 2, when $r_C = k - 1$ and $r_D = k$ there is always a solution, δ, to equation (10). So, in large samples, tests of the model based on the intercept-normalization will reject it with the same probability as the size of the test, say 5% of the time. But, in this situation, as discussed in Section 2, the solution to equation (10) implies a zero-mean SDF. As mentioned above, this solution puts no weight on economically relevant risk factors, and, instead, puts all its weight on a "useless" linear combination of the factors. The OIR test is not useful for detecting this problematic feature of the solution, because, strictly speaking, the zero-mean SDF is valid in this case. Kan and Zhang's (1999a) results tell us that the mean-normalization can also run into trouble, especially when the model is tested in the first or second step of GMM.

Symmetrically when $r_C = k$ and $r_D = k - 1$ there is always a solution, γ, to equation (12). So, in large samples, tests of the model based on the mean-normalization will reject it with the same probability as the size of the test, say 5% of the time. It is not entirely clear what the properties of the intercept-normalization will be in this case, but they are likely non-standard given that δ is not identified.

[9] Kan and Zhang (1999a) define a useless factor as one that is independent of R_t^e and f_t at all leads and lags.

3.4 Testing identification

In the introduction I discussed tests of the rank conditions proposed by Cragg and Donald (1997) and Kleibergen and Paap (2006). The Kleibergen and Paap (2006) statistic has a computational advantage because it does not involve nonlinear optimization whereas the Cragg and Donald (1997) test does. Instead, it only involves forming the singular value decomposition (SVD) of a matrix, giving it a computational advantage in Monte Carlo experiments. For this reason, I focus the rest of my discussion on the Kleibergen and Paap (2006) statistic.

Suppose we have an $n \times k$ matrix Π, and the null hypothesis is that $\text{rank}(\Pi) = r < k$. Rather than directly test the rank of Π using some estimate, $\hat{\Pi}$, Kleibergen and Paap (2006) suggest forming the scaled matrix $\Theta = G\Pi F'$ where $G_{n\times n}$ and $F_{k\times k}$ are invertible matrices that make Θ invariant to invertible transformations of the data. When $\Pi = \text{cov}(R^e, f)$ natural choices are $G = \Sigma_R^{-1/2}$, the Cholesky decomposition of the inverse of the covariance matrix of R^e, and $F = \Sigma_f^{-1/2}$, the Cholesky decomposition of the inverse of the covariance matrix of f.

The next step is to form the SVD $\Theta = USV'$, where the upper $k \times k$ block of S is a diagonal matrix with the singular values of Θ arranged from largest to smallest. When $\text{rank}(\Pi) = \text{rank}(\Theta) = r < k$, the $k - r$ smallest singular values Θ are zero. Therefore, the rank test forms the SVD of an estimate $\hat{\Theta}$ and finds its $k - r$ smallest singular values. The test statistic measures the size of these singular values, in a statistical sense, relative to zero. The Appendix provides details of the calculation of the statistic, denoted $\text{rk}(r)$, which converges asymptotically to a $\chi^2_{(n-r)(k-r)}$ distribution.

3.5 A procedure for model reduction

Suppose that a researcher has a candidate model based on the k-dimensional risk factor, f. I propose the following sequential procedure for testing the model's rank, possibly reducing the dimension of the model, estimating the model, and performing an OIR test. The procedure assumes that a researcher has a favored size, denoted α, to be used in the OIR test. Model reduction, estimation and evaluation takes place after a separate procedure determines the dimension of the model the researcher will actually estimate. The model dimension procedure can be described as a loop as follows.

Step 0. Test whether $r_C = 0$ using the rk(0) statistic.

If the p-value associated with the statistic is greater than α, discard the model and proceed no further. If the p-value associated with the statistic is less than or equal to α continue to step 2.

Step 1. Test whether $r_C = 1$ using the rk(1) statistic.

If the p-value associated with the statistic is greater than α, break out of the loop and invoke the model reduction procedure described below. If the p-value associated with the statistic is less than or equal to α continue to step 3.

The procedure continues, like this, until ...

Step $k-1$. Test whether $r_C = k-1$ using the rk$(k-1)$ statistic.

If the p-value associated with the statistic is greater than α, break out of the loop and invoke the model reduction procedure described below. If the p-value associated with the statistic is less than or equal to α proceed to model estimation without reducing the dimension of the model.

The model reduction procedure is invoked if the p-value associated with rk(r) is greater than α, and is a by-product of the construction of the statistic. To see how it works consider the following linear combination of the risk factors: $\tilde{f} = Af$ with $A = V'F$, where V is the the matrix from the SVD of Θ, defined above, and $F = \Sigma_f^{-1/2}$. The covariance matrix of $\tilde{f}$ is $\tilde{\Sigma}_f = V'V = I_k$, implying that $\tilde{f}$ is a vector of mutually orthogonal factors with unit variance. Also $\tilde{C} \equiv \text{cov}(R^e, \tilde{f}) = CF'V$. It follows that $\tilde{\Theta} = G\tilde{C}(\tilde{\Sigma}^{-1/2})' = GCF'V = USV'V = US$ implying that $\tilde{\Theta}$ and Θ have the same singular values. When $r_C = r$, only the first r elements of $\tilde{f}$ are relevant factors. Hence the procedure suggests reducing the model to one in which $\tilde{f}_r = A_r f$ is a $r \times 1$ vector of risk factors. Here A_r is a $r \times k$ matrix representing the first r rows of the matrix A, or, equivalently, $V_r'F$, where V_r represents the first r columns of V. In the actual procedure all of the matrices are replaced with their sample equivalents.[10]

Model estimation proceeds in the standard way using either the mean-normalization or the intercept-normalization. The OIR test is conducted in a standard way but has $n - r$ degrees of freedom. If the researcher wishes to recover the parameters associated with the original factors this is straightforward. The SDFs for the two normalizations can be written

[10] I discuss the implications of the non-uniqueness of the SVD of Θ in the Appendix.

as

$$
\begin{aligned}
m^{\delta} &= 1 - \tilde{f}_r'\tilde{\delta} = 1 - f'(A_r'\tilde{\delta}), \\
m^{\gamma} &= 1 - (\tilde{f}_r - \tilde{\mu}_f)'\tilde{\gamma} = 1 - (f - \mu_f)'(A_r'\tilde{\gamma}),
\end{aligned}
$$

so the parameters attached to the original factors are $\delta = A_r'\tilde{\delta}$ and $\gamma = A_r'\tilde{\gamma}$.

This is obviously a simple procedure, but one might ask whether it has reasonable size and power properties. Consider size first. The size calculation depends on the true value of r_C, which I treat as an unknown.

If $r_C = k$ the p-value associated with the test in steps 0 through $k-1$ should converge to zero 0 in sufficiently large samples. This is because the null hypotheses being tested are not local to the alternative, which is that $\text{cov}(R^e, f)$ has full column rank. Hence, if $r_C = k$ the probability of rejecting the model should be α, the size of the OIR test.

By the same argument, if $r_C = k-1$ the p-value associated with the test in steps 0 through $k-2$ should converge to zero 0 in sufficiently large samples. At step $k-1$, however, $r_C = k-1$ will be rejected with a probability that limits to α. So, with probability $1-\alpha$ the model reduction procedure will be invoked, and the probability of rejecting the model based on the OIR test should be α. On the other hand, with probability α the model reduction procedure is not invoked. In this case, the probability of rejecting the model based on the OIR test will be some $\tilde{\alpha}_{k-1} \neq \alpha$ because inference is non-standard for models with spurious factors. Therefore, overall, the probability of rejecting the model is $\alpha_{k-1} = (1-\alpha)\alpha + \alpha\tilde{\alpha}_{k-1}$. Propositions 3 and 4 in Kan and Zhang (1999a) suggest that for the mean-normalization the probability of rejecting a correctly specified model that includes a spurious factor is less than α. This suggests that $0 \leq \tilde{\alpha}_{k-1} < \alpha$ and that $(1-\alpha)\alpha \leq \alpha_{k-1} < \alpha$ but it must certainly be the case that $\alpha_{k-1} < \alpha(2-\alpha)$.

A similar argument can be used to establish that if $r_C = r$, the probability of invoking the model reduction procedure at step r is $1-\alpha$, in which case the probability of rejecting the model based on the OIR test would be α. But with probability α it would not be invoked and the researcher would move on to step $r+1$. Again, this suggests that the probability of rejecting the model is $\alpha_r = (1-\alpha)\alpha + \alpha\tilde{\alpha}_r$, where $\tilde{\alpha}_r$ is the probability of rejecting the larger model at a later step. Again, the results in Kan and Zhang (1999a) are suggestive that $0 \leq \tilde{\alpha}_r < \alpha$ but certainly $\alpha_r < \alpha(2-\alpha)$.

In terms of power, consider a misspecified model. If the C matrix associated with this model has full rank this should be revealed with probability one in large samples. Therefore, the model reduction procedure will not be invoked and the model should be rejected with the same probability as it would be using standard GMM procedures. The main advantage of the procedure is that it should improve the detection of invalid models for which C has less than full rank because the model will often be reduced in dimension to one where the parameters are fully identified and inference is standard. Additionally, the procedure is designed to discard spurious risk factors, or reject models based entirely on them.

In Section 5 I use Monte Carlo experiments to explore the size and power properties of the model reduction procedure.

4 Empirical Findings

In Table 1 I present the results of rank tests for a variety of models taken from the literature. In this section I discuss two of these models in more detail: the Fama and French (1993) Three Factor model and the Yogo (2006) model. These models are also used as the inspiration for the Monte-Carlo experiments that I run later in the paper. Estimates of the model parameters and the results of OIR tests are presented in Table 2. I use quarterly data over the period 1949Q1–2012Q4 to estimate both models. For R^e I use the excess returns of the 25 portfolios of U.S. stocks sorted on size and the book-to-market value ratio introduced by Fama and French (1993).[11]

4.1 The Fama-French Three Factor Model

The Fama-French model uses three factors: (1) the excess return on the value-weighted U.S. stock market (Mkt-Rf), (2) the return differential between a portfolio of small firms and a portfolio of large firms (SMB) and (3) the return differential between a portfolio of high-value firms and a portfolio of low-value firms (HML), where a firm's value is measured as the ratio of its book value to its market capitalization.

As we saw in Table 1, the rank tests strongly reject the null hypotheses that this model has reduced rank. It appears that $r_C = r_D = 3$. Table 2 indicates that the statistical significance of δ and γ and the results of the OIR tests (which are all strong rejections) are quite similar across the two normalizations. Additionally, Table 2 shows that if the first step GMM estimates of δ obtained with intercept-normalization are mapped to equivalent

[11] The data are described in more detail in the Appendix.

values of γ these estimates are quite similar to the estimates of γ obtained with the mean-normalization. However, this is not true for iterated GMM.

4.2 Yogo (2006)

Yogo (2006) proposes a consumption-based model in which agents have recursive preferences over a consumption bundle of nondurable and durable goods. When this model is approximated with a linear SDF, the risk factors are the log-growth rate of real per capita consumption of nondurables and services (Δc_{ns}), the log-growth rate of the real per capita consumption of durables (Δc_d) and the return on wealth, which is proxied with the real return on the value-weighted U.S. stock market (Mkt).

From the results of the rank tests shown in Table 1 it would be tempting to conclude that $r_C = 1$ and $r_D = 2$ for Yogo's model. If this were true in population, then any valid parameterization of the model would have a zero mean, and would put all its weight on spurious linear combinations of the factors.

As Table 3 shows, if we rely on the intercept-normalization we get statistically significant estimates of δ for Δc_d (for first step and iterated GMM) and Δc_{ns} (for iterated GMM). The model also passes the OIR test with flying colors.

On the other hand, if we use the mean-normalization we draw different conclusions. For first step GMM, none of the parameter estimates are statistically significant, but the model is not rejected by the OIR test. For iterated GMM, the γ parameters for Δc_d and Δc_{ns} are statistically insignificant, but the coefficient associated with Mkt is marginally significant. The model is sharply rejected by the OIR test. It is tempting to attribute the qualitatively different inference across normalizations to the fact that $r_C < r_D$.

This model appears to be a candidate for reduction to a single factor. Using the procedure outlined in Section 3.5, we end up with the following single factor: $\tilde{f}_1 = -1.47 \cdot \Delta c_{ns} + 2.72 \cdot \Delta c_d + 12.1 \cdot \text{Mkt}$. This factor has a correlation of 0.9999 with Mkt.[12] The two consumption factors are, roughly speaking, excluded by the model reduction procedure, and we end up with a model very similar to the CAPM, except that we have Mkt instead of Mkt-Rf as the risk factor. As Table 3 indicates, once the model is reduced in dimension the model is sharply rejected based on the OIR test for both normalizations.

[12] As described above, the model reduction procedure uses $V_1'F$, where V_1 is the first column of V in the SVD of Θ. As it turns out, if we form $V_{2:k}'F$, using the remaining columns of V these linear combinations put almost no weight on Mkt, and are almost uncorrelated with it.

5 A Monte Carlo Experiment

To further demonstrate the sensitivity of empirical results to the choice of normalization in the presence of failure of the rank conditions, I conduct Monte Carlo experiments. These experiments also provide evidence on the performance of the model reduction procedure I proposed in Section 3.5.

5.1 The True Model, Some Misspecified Models, and an Over-specified Model

Here, the true model is calibrated to resemble the Fama-French three-factor model estimated using U.S. data over the sample period 1949:Q1–2012:Q4. The model is used to generate an $n \times 1$ (with $n = 25$) vector of artificial excess returns R_t^e with $E(R^e)$ equalling the model-predicted expected returns in the data. Details on how the simulated data are generated are provided in the Appendix.

Using the data generated from the model, we estimate several test models in 10000 artificial samples of 256 observations (the size of our U.S. data sample):

1. The true model, which includes the artificial Mkt-Rf, SMB and HML factors. This model has full rank and is correctly specified, by construction.

2. A model that includes the artificial Mkt-Rf factor. This factor is relevant, in the sense that it has non-zero covariance with the returns generated within the experiment. So it satisfies the rank conditions for identification. However, the model is misspecified because the Mkt-Rf factor, alone, cannot price returns accurately.

3. A model with a single spurious (uncorrelated with returns) factor, S_t, whose behavior somewhat mimics durable consumption growth in U.S. data. I assume that $S_t \sim Niid(\mu_s, \sigma_s^2)$ and set μ_S and σ_S^2 equal to the sample mean and variance of durable consumption growth in the data. In this model, C has reduced rank (0), but D has full rank (1).

4. A model with two factors: Mkt-Rf and S. In this model, C has reduced rank (1), but D has full rank (2).

5. The last model is an over-specified model which includes the Mkt-Rf, SMB and HML factors, as well as the spurious factor, S. In this model C and D are $n \times 4$ but *both*

have reduced rank (3). The model is not misspecified, because it nests the true model. The presence of S, however, affects estimation and inference.

5.2 Rank Tests

I first explore the performance of the Kleibergen and Paap (2006) rank tests. Table 4 shows that the rank tests never understate the rank of C. For models 1 and 2, which have full rank, the tests reject the null hypothesis of reduced rank 100% of the time. For model 4, where $r_C = 1$, the test always rejects the null that $r_C = 0$. For model 5, where $r_C = 3$, the test always rejects the null that $r_C \leq 2$. The tests do sometimes overstate the rank of C in the sense that for model 3, where $r_C = 0$, the test rejects the null that $r_C = 0$ 12.1% of the time (when size is set at 5%). For model 4, where $r_C = 1$, the test rejects the null that $r_C = 1$ 11.9% of the time (when size is set at 5%). For model 5, where $r_C = 3$, the test rejects the null that $r_C = 3$ 11.4% of the time (when size is set at 5%).

For models 1 and 2, where D has full rank, the tests reject the null hypothesis of reduced rank 100% of the time. For model 3, where D also has full rank the rank test rejects the null 92.5 of the time (when size is set at 5%). For model 4, where D also has full rank the rank test rejects the null 69.8% of the time (when size is set at 5%).[13] For model 5, where $r_D = 3$ the rank tests reject the null that $r_D = 3$ 10.9% of the time (when size is set at 5%).

In extended results presented in the Appendix, I show that the rank tests become close to perfect in 10000 simulated samples each of which has 10000 observations. When the null hypothesis of reduced rank is false, the tests reject the null 100% of the time, as I conjectured in Section 3.5. When the null hypotheses of reduced rank are true, I find the tests have almost exactly their asymptotic size.

5.3 Estimating the True Model

Table 5 shows that for both normalizations the parameter estimates are centered near the true values of the parameters. For the factors that play the biggest role in pricing the returns in the model (the pseudo Mkt-Rf and HML factors) the parameters are statistically significant in almost all samples. The parameter associated with the pseudo-SMB factor is usually not significant, consistent with it playing a very small role in pricing the returns. The OIR test usually does not reject the model, with size being slightly excessive for the

[13] Rank tests on D play no role in the model reduction procedure that I described earlier.

mean normalization and very slightly reduced for the intercept normalization. In extended results presented in the Appendix, I show that OIR tests have almost exactly the correct size in 10000 simulated samples each of which has 10000 observations. The qualitative difference between the results for the two normalizations disappears.

5.4 Estimating Model 2

Model 2 is misspecified. The proposed risk factor, Mkt-Rf, is relevant but insufficient to price the assets. Not surprisingly, as Table 6 shows, for both normalizations we find parameter estimates to be significant in a very large fraction of the samples. Additionally, for both normalizations the model is rejected roughly 70% of the time using an OIR test with size set at 5%. In extended results presented in the Appendix, I show that these rejection rates rise to 100% in 10000 simulated samples each of which has 10000 observations.

5.5 Estimating Model 3

Model 3 uses a single spurious factor. The intercept and mean-normalizations behave differently because the intercept-normalization is identified and (strictly speaking) valid, while the mean-normalization is neither. Standard asymptotics apply to the intercept-normalization, but not to the mean-normalization. This is reflected in the results shown in Table 7.

In the simulations the inverse of the mean of the spurious factor, S, is $\mu_S^{-1} = 102$. As expected, therefore, the estimates of δ are centered near this value, and are frequently (in this case, always) statistically significant. Additionally, because the pricing equations for the intercept-normalization are satisfied at μ_S^{-1}, the model is rejected very infrequently, with small sample size being close to the asymptotic size of the OIR test.

On the other hand, when the mean-normalization is used, the parameter γ_S is statistically significant in a much smaller fraction of the samples. At the first and second steps of GMM the OIR test has very weak power to reject the model. Paradoxically, the power of the OIR tests rises sharply with further iterations over the weighting matrix, while, at the same time, the tendency of $\hat{\gamma}_S$ to be statistically significant rises to 22% (when size is set at 5%). The remarkably different behavior of $\hat{\gamma}_S$ and its associated t-statistic across the GMM steps can be better understood with reference to Figure 6. When the identity matrix is used to weight the pricing equations the distribution of $\hat{\gamma}_S$ is very wide and bimodal. This bimodality is shared by the distribution of the t-statistic but its tails are not that thick. In later GMM steps, where a non-identity weighting matrix is used, the distribution of $\hat{\gamma}_S$ narrows and is unimodal. But with sufficient iterations over the weighting matrix it has very fat tails. Figure 7 shows that the distribution of $\hat{\gamma}_S$ widens when the sample size in the Monte Carlo experiments is increased to 10000. This reflects the fact that $\hat{\gamma}_S$ is unidentified.[14] As it turns out, the behavior of the t-statistic also worsens, with it exhibiting very fat tails at the first GMM step and after many iterations over the weighting matrix.

The model reduction procedure I described in Section 3.5 suggests discarding the model if the null hypothesis that $r_C = 0$ cannot be rejected at the 5% level of significance. The results in Table 10 show that this null is rejected in only 12.1% of the repeated samples. So the model reduction procedure would yield a 87.9% rejection rate of the model based on the rank test alone. For the procedure as a whole—in which the model is estimated and tested if the model passes the rank test—the rejection rate rises to 88.4% for first and second step GMM and 98.5% for iterated GMM if the mean-normalization is used to estimate the model. The corresponding rejection rates for the intercept-normalization are 89.2% and 93.2%. When I increase the sample size in the simulations to 10000 observations, these rejection rates rise to 95.1% and 100% for the mean-normalization, and 98.5% and 99.0% for the intercept-normalization.

5.6 Estimating Model 4

Model 4 uses the Mkt-Rf factor and the spurious factor, S. Strictly speaking, the intercept-normalization is valid. However it is valid when no weight is put on the relevant factor, Mkt-Rf, and all the weight in the SDF is on the spurious factor S. The mean normalization is misspecified.

In Table 8 we see that, as for Model 3, the estimates of δ_S are centered near $\mu_S^{-1} = 102$, and are always statistically significant. The typical estimate of $\delta_{\text{Mkt}-\text{Rf}}$, on the other hand, is small and significant in a much smaller fraction of the samples. Additionally, because the pricing equations for the intercept-normalization are satisfied at $(0, \mu_S^{-1})$, the model is rejected very infrequently, with small sample size being close to the asymptotic size of the OIR test.

On the other hand, when the mean-normalization is used, $\hat{\gamma}_{\text{Mkt}-\text{Rf}}$ is typically quite large (similar to the values we obtained for Model 2) and significant in a large fraction of the

[14] In an earlier draft I showed that with the data generating process used in my Monte Carlo simulations, $T^{-1/2}\hat{\gamma}_S$ has an asymptotic distribution so, in a sense, we expect the domain of $\hat{\gamma}_S$ to widen at the rate $T^{1/2}$.

samples, while γ_S is statistically significant in a smaller fraction of the samples. Once again, however, at the first and second steps of GMM the OIR test has quite weak power to reject the model. Paradoxically, the power of the OIR tests rises sharply with further iterations over the weighting matrix, while, at the same time, the tendency of $\hat{\gamma}_S$ to be statistically significant rises to 16.5% (when size is set at 5%).

The model reduction procedure I described in Section 3.5 suggests rejecting the model if the null hypothesis that $r_C = 0$ cannot be rejected at the 5% level of significance. However, as the results in Table 4 show this null is rejected in every repeated sample. The model reduction procedure also suggests reducing the model to a single factor if the null hypothesis that $r_C = 1$ cannot be rejected at the 5% level of significance. This occurs in 88.1% of the samples, as indicated in Table 10. For the procedure as a whole the model rejection rate rises to 64.3% for first and second step GMM and 70.8% for iterated GMM. This is an improvement over the 16.8% and 62.4% rejection rates for the procedure without model reduction. The corresponding rejection rates for the intercept-normalization are 60.7% and 63.8%. When I increase the sample size in the simulations to 10000 observations, these rejection rates rise to 97.0% and 100% for the mean-normalization, and 98.8% and 99.3% for the intercept-normalization.

When the model reduction procedure is invoked, it chooses linear combinations of the two factors that put most of their weight on Mkt-Rf and are highly correlated with it. As the sample size increases, the procedure limits to choosing Mkt-Rf as the single factor, since S is uncorrelated with returns, by construction.

5.7 Estimating Model 5

Model 5 uses the factors from the true model (Mkt-Rf, SMB and HML) as well as the spurious factor, S. This model is not misspecified because it nests the true model, but it has reduced rank for both normalizations: $r_C = r_D = 3$.

In Table 9 we see that adding the spurious factor changes the small sample size of the OIR test. The under-rejection we observed in Table 5 for the intercept-normalization becomes more exaggerated here. The over-rejection we observed for the mean-normalization almost vanishes. But, as in the other cases where a spurious factor is included in the model, the normalizations differ sharply in terms of coefficient estimates. For the intercept-normalization estimates of $\delta_{\text{Mkt}-\text{Rf}}$ and δ_{HML} are not centered near their true values, and they are less often found to be statistically significant compared to the case where the spurious factor is excluded from the model. On the other hand, estimates and inference regarding $\gamma_{\text{Mkt}-\text{Rf}}$ and γ_{HML} are almost unaffected by the inclusion of the spurious factor.

The model reduction procedure I described in Section 3.5 suggests reducing the model to three factors when the null hypothesis that $r_C = 3$ cannot be rejected at the 5% level of significance. As the results in Table 10 show, this occurs in 88.6% of the samples. In these samples, the OIR test is performed using the model with three factors instead of four. As a consequence, for the mean-normalization the model rejection rate rises to 7.0% for first and second step GMM and 6.9% for iterated GMM. This compares to 5.1% and 5.6% rejection rates for the procedure without model reduction. The corresponding rejection rates for the intercept-normalization are 5.5% and 5.4%. When I increase the sample size in the simulations to 10000 observations, these rejection rates are 5.0% and 5.0% for the mean-normalization, and 6.1% and 6.2% for the intercept-normalization.

5.8 Indirect Estimates of Model Parameters

If we use the intercept-normalization, GMM produces an estimate, $\hat{\delta}$. This estimate can be mapped into an indirect estimate of γ if we use the formula $\gamma(\hat{\delta}) = \hat{\delta}/(1 - \bar{f}'\hat{\delta})$. Similarly, if we use the mean-normalization, GMM produces an estimate, $\hat{\gamma}$. This estimate can be mapped into an equivalent estimate of δ if we use the formula $\delta(\hat{\gamma}) = \hat{\gamma}/(1 + \bar{f}'\hat{\gamma})$. Is the asymptotic distribution of $\hat{\delta}$ similar to that of $\delta(\hat{\gamma})$? How about $\hat{\gamma}$ and $\gamma(\hat{\delta})$?

The Appendix provides extensive evidence on this question for Monte Carlo experiments with a very large number of observations (10000). These experiments suggest that when the model is true and identified (i.e., Model 1), the direct and indirect estimates have very similar distributions in large repeated samples.

When the model is misspecified (Model 2) the direct and indirect estimates, in general, have different probability limits. They also have different distributions around these limits. This is confirmed by the Monte Carlo evidence, although the distributions look quite similar at the first GMM step.

When the model includes a spurious factor (Models 3, 4 and 5), or, equivalently, the mean normalization is poorly identified (both normalizations are underidentified for Model 5), the direct and indirect estimates of δ and γ have very different looking distributions, even for factors that belong in the model. The presence of spurious factors matters.

6 Conclusion

When excess returns are used to estimate linear SDFs, GMM estimation requires that a normalization of the SDF be adopted. Two standard normalizations of the SDF, the intercept-normalization and the mean-normalization are equivalent in population, in the sense that they are proportional up to a constant, when the model is valid and fully identified. However, the conditions under which these normalizations are identified are, in general, different.

In practice, different normalizations sometimes lead to very different qualitative inferences about a model. Estimates of the slope coefficients of the SDF can differ sharply in terms of magnitude and statistical significance. OIR tests can differ sharply in outcome. I have demonstrated this, here, for several factor models fit to U.S. data. I interpret these differences as the consequence of failures of the rank conditions for identification.

It is well known that lack of identification affects inference. But here, there is an additional problem that the rank conditions for identification can fail differentially across model normalizations. In particular, I establish that when the mean-normalization is less identified than the intercept-normalization, there is always at least one valid parameterization of the proposed SDF, and any valid SDF is mean zero and uncorrelated with returns. Also, when the intercept-normalization is less identified than the mean-normalization, there is always at least one valid parameterization of the proposed SDF, and any valid SDF has zero intercept.

I propose model diagnostics and a model reduction procedure based on Kleibergen and Paap (2006)'s rank test. I argue that these tests should have desirable asymptotic size and power properties. In a Monte Carlo experiment I show that these rank tests are, indeed, a powerful diagnostic. My preliminary evidence also suggests that the model reduction method increases the power of asset pricing tests, while having little undesired effect on their size.

References

Bryzgalova, Svetlana, 2014, Spurious factors in linear asset pricing models, Working Paper, London School of Economics.

Burnside, Craig, 2011, The cross section of foreign currency risk premia and consumption growth risk: Comment, *American Economic Review* 101, 3456–3476.

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *The Journal of Business* 59, 383–403.

Cochrane, John H., 2005, *Asset Pricing* (Princeton University Press) revised edn.

Cragg, John G., and Stephen G. Donald, 1997, Inferring the rank of a matrix, *Journal of Econometrics* 76, 223–250.

den Haan, Wouter J., and Andrew T. Levin, 2000, Robust covariance matrix estimation with data-dependent VAR prewhitening order, NBER Technical Working Paper No. 255.

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, 607–636.

Ferson, Wayne E., and Campbell R. Harvey, 1993, The risk and predictability of international equity returns, *The Review of Financial Studies* 6, 527–566.

Gibbons, Michael R., 1982, Multivariate tests of financial models: A new approach, *Journal of Financial Economics* 10, 3–27.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2012, Further results on the limiting distribution of gmm sample moment conditions, *Journal of Business & Economic Statistics* 30, 494–504.

———, 2013a, Chi-squared tests for evaluation and comparison of asset pricing models, *Journal of Econometrics* 173, 108–125.

———, 2013b, Misspecification-robust inference in linear asset pricing models with irrelevant risk factors, Federal Reserve Bank of Atlanta Working Paper 2013-9.

———, 2014, Spurious fit in unidentified asset-pricing models, Working Paper, Imperial College, London.

Hall, Alastair R., 2005, *Generalized Method of Moments* (Oxford University Press).

———, and Atsushi Inoue, 2003, The large sample behaviour of the generalized method of moments estimator in misspecified models, *Journal of Econometrics* 114, 361–394.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

——— , and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *The Journal of Finance* 52, 557–590.

Hou, Kewei, and Robert L. Kimmel, 2006, On estimation of risk premia in linear factor models, Working Paper, Ohio State University.

Jagannathan, Ravi, and Yong Wang, 2007, Lazy investors, discretionary consumption, and the cross-section of stock returns, *The Journal of Finance* 62, 1623–1661.

Kan, Raymond, and Cesare Robotti, 2008, Specification tests of asset pricing models using excess returns, *Journal of Empirical Finance* 15, 816–838.

——— , 2009, Model comparison using the Hansen-Jagannathan distance, *The Review of Financial Studies* 22, 3449–3490.

——— , and Jay A. Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance* 68, 2617–2649.

Kan, Raymond, and Chu Zhang, 1999a, GMM tests of stochastic discount factor models with useless factors, *Journal of Financial Economics* 54, 103–127.

——— , 1999b, Two-pass tests of asset pricing models with useless factors, *The Journal of Finance* 54, 203–235.

Kleibergen, Frank, 2009, Tests of risk premia in linear factor models, *Journal of Econometrics* 149, 149–173.

——— , and Richard Paap, 2006, Generalized reduced rank tests using the singular value decomposition, *Journal of Econometrics* 133, 97–126.

Kleibergen, Frank, and Zhaoguo Zhan, 2013, Unexplained factors and their effects on second pass r-squared's and t-tests, Manuscript, Brown University.

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, 1238–1287.

Lustig, Hanno, and Adrien Verdelhan, 2007, The cross section of foreign currency risk premia and consumption growth risk, *The American Economic Review* 97, 89–117.

Peñaranda, Francisco, and Enrique Sentana, 2015, A unifying approach to the empirical evaluation of asset pricing models, *The Review of Economics and Statistics* 97, 412–435.

Shanken, Jay, and Guofu Zhou, 2007, Estimating and testing beta pricing models: Alternative methods and their performance in simulations, *Journal of Financial Economics* 84, 40–86.

Wright, Jonathan H., 2003, Detecting lack of identification in GMM, *Econometric Theory* 19, 322–330.

Yogo, Motohiro, 2006, A consumption-based explanation of expected stock returns, *The Journal of Finance* 61, 539–580.

TABLE 1: **Tests for Failure of Rank Conditions (p-values)**

Model	# of Factors k	Test rank r	$\text{cov}(R^e, f)$ CD	$\text{cov}(R^e, f)$ KP	$E(R^e f')$ CD	$E(R^e f')$ KP	Sample period
CAPM	1	0	0.000	0.000	0.000	0.000	49Q1–12Q4
Fama-French 3-Factor	3	2	0.000	0.000	0.000	0.000	49Q1–12Q4
		1	0.000	0.000	0.000	0.000	
		0	0.000	0.000	0.000	0.000	
CCAPM	1	0	0.076	0.076	0.000	0.000	49Q1–12Q4
Durables-CCAPM	2	1	0.961	0.958	0.024	0.025	49Q1–12Q4
		0	0.017	0.017	0.000	0.000	
Yogo (2006)	3	2	0.977	0.931	0.521	0.513	49Q1–12Q4
		1	0.767	0.770	0.000	0.000	
		0	0.000	0.000	0.000	0.000	
Lettau and Ludvigson (2001)	3	2	0.239	0.167	0.240	0.166	52Q1–12Q3
		1	0.040	0.004	0.004	0.000	
		0	0.000	0.000	0.000	0.000	
Jagannathan and Wang (2007)	1	0	0.005	0.005	0.000	0.000	1949–2012
Lustig and Verdelhan (2007)	3	2	0.732	0.670	0.790	0.708	1953–2002
		1	0.818	0.728	0.640	0.629	
		0	0.758	0.758	0.000	0.000	

Note: For eight models from the literature, the table presents p-values for tests of the null hypothesis $H_0 : \text{rank}(M) = r$ where M is an $n \times k$ matrix [$\text{cov}(R^e, f)$ or $E(R^e f')$]. CD and KP indicate tests based on Cragg and Donald (1997), and Kleibergen and Paap (2006). Every case, except Lustig and Verdelhan (2007), uses the real excess returns to the Fama and French (1993) 25 portfolios sorted on the basis of size and value. For risk factors, the CAPM uses the real market excess return (Mkt-Rf). The Fama-French 3 factor model uses Mkt-Rf, SMB and HML (expressed in real terms). The CCAPM model uses nondurable & service consumption growth. The Durables-CCAPM case adds durable consumption growth to the CCAPM. The Yogo (2006) model adds the real market return (Mkt) to the Durables-CCAPM. The Lettau and Ludvigson (2001) model uses consumption growth, *cay*, and the product of consumption growth and *cay*. The Jagannathan and Wang (2007) model uses data on a Q4-Q4 annual basis, and uses nondurable consumption growth as the risk factor. The Lustig and Verdelhan (2007) case uses a different set of returns: Eight currency portfolios sorted by interest rate, and measured on an annual basis. The risk factors are the same as the ones from Yogo (2006), measured at the annual frequency.

TABLE 2: **GMM Estimates of the Fama & French Three Factor Model**

	First GMM Step					Iterated GMM				
Factor	δ	$\gamma(\delta)$	J^{δ}	γ	J^{γ}	δ	$\gamma(\delta)$	J^{δ}	γ	J^{γ}
Mkt-Rf	3.27	3.69	50.3	3.67	54.9	5.14	6.16	47.9	3.79	54.4
	(0.84)	(1.06)	(0.001)	(1.06)	(0.000)	(0.81)	(1.19)	(0.001)	(1.01)	(0.000)
SMB	−0.17	−0.19		−0.17		−1.21	−1.45		−0.76	
	(1.19)	(1.34)		(1.34)		(1.10)	(1.32)		(1.28)	
HML	5.02	5.68		5.57		6.87	8.24		5.98	
	(1.04)	(1.34)		(1.33)		(0.95)	(1.36)		(1.18)	

Note: This table presents first step GMM estimates and GMM estimates after iterating to convergence over the weighting matrix. For the intercept normalization δ is the SDF parameter, with standard errors in parentheses [$\gamma(\delta)$ is given by $\gamma = \delta/(1 - \mu_f'\delta)$ with standard errors computed using the delta method]. J^{δ} is the OIR test statistic with p-values in parentheses. For the mean normalization γ is the SDF parameter, with standard errors in parentheses. The returns used to estimate the model are the excess returns of the Fama and French (1993) 25 portfolios sorted on the basis of size and value. The risk factors, are Mkt-Rf, SMB and HML. Detailed data descriptions are in the Appendix. Sample period is 1949Q1–2012Q4.

TABLE 3: **GMM Estimates of the Yogo Model**

Factor	First GMM Step δ	$\gamma(\delta)$	J^{δ}	γ	J^{γ}	Iterated GMM δ	$\gamma(\delta)$	J^{δ}	γ	J^{γ}
Δc_{ns}	23.7	288	17.2	234	6.68	45.0	387	18.0	43.5	46.9
	(33.5)	(619)	(0.749)	(204)	(0.999)	(17.4)	(377)	(0.709)	(42.1)	(0.002)
Δc_{d}	81.1	987		249		67.6	581		50.0	
	(19.8)	(1881)		(236)		(9.78)	(526)		(36.2)	
Mkt	0.50	6.04		2.15		0.41	3.55		1.94	
	(0.55)	(12)		(2.91)		(0.43)	(4.76)		(1.03)	
Model Reduction										
$\tilde{f}_1$	0.25	0.27	66.4	0.27	69.9	0.43	0.49	65.6	0.22	70.3
	(0.06)	(0.08)	(0.000)	(0.08)	(0.000)	(0.06)	(0.09)	(0.000)	(0.07)	(0.000)
Implied Parameters										
Δc_{ns}	−0.37	−0.40		−0.39		−0.63	−0.71		−0.32	
Δc_{d}	0.68	0.74		0.73		1.18	1.33		0.60	
Mkt	3.06	3.29		3.26		5.21	5.91		2.69	

Note: This table presents first step GMM estimates and GMM estimates after iterating to convergence over the weighting matrix. For the intercept normalization δ is the SDF parameter, with standard errors in parentheses [$\gamma(\delta)$ is given by $\gamma = \delta/(1-\mu_f'\delta)$ with standard errors computed using the delta method]. J^{δ} is the OIR test statistic with p-values in parentheses. For the mean normalization γ is the SDF parameter, with standard errors in parentheses. The returns used to estimate the model are the real excess returns of the Fama and French (1993) 25 portfolios sorted on the basis of size and value. The risk factors are nondurable & service consumption growth (Δc_{ns}), durable consumption growth, Δc_d, and the real market return (Mkt). Detailed data descriptions are in Appendix. Sample period is 1949Q1–2012Q4.

TABLE 4: **Rank Tests in the Monte Carlo Experiment**

	% rejected at			% rejected at	
	10% level	5% level		10% level	5% level
1. True Model (Factors: Mkt-Rf, SMB, HML) $r_D = r_C = 3$					
$H_0:\ r_D = 2$	100%	100%	$H_0:\ r_C = 2$	100%	100%
$H_0:\ r_D = 1$	100%	100%	$H_0:\ r_C = 1$	100%	100%
$H_0:\ r_D = 0$	100%	100%	$H_0:\ r_C = 0$	100%	100%
2. Single Relevant Factor (Factor: Mkt-Rf) $r_D = r_C = 1$					
$H_0:\ r_D = 0$	100%	100%	$H_0:\ r_C = 0$	100%	100%
3. Single Spurious Factor (Factor: S) $r_D = 1$, $r_C = 0$					
$H_0:\ r_D = 0$	96.3%	92.5%	$H_0:\ r_C = 0$	20.7%	12.1%
4. Two Factors (Factors: Mkt-Rf, S) $r_D = 2$, $r_C = 1$					
$H_0:\ r_D = 1$	80.0%	69.8%	$H_0:\ r_C = 1$	20.2%	11.9%
$H_0:\ r_D = 0$	100%	100%	$H_0:\ r_C = 0$	100%	100%
5. Over-specified Model (Factors: Mkt-Rf, SMB, HML, S) $r_D = r_C = 3$					
$H_0:\ r_D = 3$	19.0%	10.9%	$H_0:\ r_C = 3$	19.5%	11.4%
$H_0:\ r_D = 2$	100%	100%	$H_0:\ r_C = 2$	100%	100%
$H_0:\ r_D = 1$	100%	100%	$H_0:\ r_C = 1$	100%	100%
$H_0:\ r_D = 0$	100%	100%	$H_0:\ r_C = 0$	100%	100%

Notes: The table reports results of Kleibergen and Paap (2006) rank tests from 10000 Monte Carlo experiments with sample size $T = 256$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. Rank tests are performed for five test models that use different factors. $r_C = \text{rank}[\text{cov}(R^e, f)]$ and $r_D = \text{rank}[E(R^e f')]$, where R^e are the returns generated in the experiment, and f is the conjectured vector of factors. The table reports the fraction of the samples in which these tests reject the null hypothesis when the size of the test is set to 10% and 5%. Details of the Monte Carlo experiments are provided in the Appendix and main text. Results for the case where $T = 10000$ are in Appendix Table 1.

TABLE 5: **Monte Carlo Experiment: Estimation of Model 1 (The True Model)**

		GMM Step 1				GMM Step 2				Iterated GMM			
				% Significant				% Significant				% Significant	
	True	Mean	Median	10%	5%	Mean	Median	10%	5%	Mean	Median	10%	5%
Intercept Normalization													
$\delta_{\text{Mkt-Rf}}$	3.25	3.28	3.29	99.2	98.3	3.67	3.68	99.7	99.3	3.74	3.75	99.7	99.3
δ_{SMB}	-0.15	-0.17	-0.17	11.3	5.9	-0.14	-0.16	17.0	10.3	-0.14	-0.15	18.0	11.4
δ_{HML}	4.92	4.97	4.99	99.8	99.4	5.54	5.55	99.9	99.7	5.63	5.65	99.9	99.7
OIR test				9.6	4.3			9.6	4.3			9.3	4.0
Mean Normalization													
$\gamma_{\text{Mkt-Rf}}$	3.70	3.77	3.75	99.2	98.1	3.77	3.74	99.4	98.5	3.79	3.75	99.4	98.6
γ_{SMB}	-0.17	-0.19	-0.19	10.4	5.1	-0.19	-0.19	12.1	6.3	-0.19	-0.19	12.3	6.6
γ_{HML}	5.61	5.70	5.65	99.8	99.4	5.71	5.65	99.8	99.4	5.74	5.68	99.8	99.4
OIR test				13.5	7.1			13.5	7.1			13.4	7.0

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 256$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The true values of the parameters of the SDF are indicated in the first column. The table reports mean and median estimates of the parameters of the intercept and mean-normalizations, as well as the frequency with which the estimated parameters are found to be statistically significant at the 10% and 5% levels. The table also reports the frequency with which the model is rejected at the 10% and 5% levels based on the OIR test. The test statistic is numerically identical at the first and second GMM steps. Details of the Monte Carlo experiments are provided in the Appendix and main text. Results for the case where $T = 10000$ are in Appendix Table 2.

TABLE 6: **Monte Carlo Experiment: Estimation of Model 2**

	GMM Step 1				GMM Step 2				Iterated GMM			
			% Significant				% Significant				% Significant	
	Mean	Median	10%	5%	Mean	Median	10%	5%	Mean	Median	10%	5%
Intercept Normalization												
$\delta_{\mathrm{Mkt-Rf}}$	3.06	3.07	99.0	98.0	3.45	3.46	99.3	98.8	3.60	3.62	99.3	98.9
OIR test			79.1	67.7			79.1	67.7			78.8	66.9
Mean Normalization												
$\gamma_{\mathrm{Mkt-Rf}}$	3.27	3.25	98.8	97.6	2.85	2.82	97.8	95.3	2.78	2.76	97.3	94.5
OIR test			81.2	70.3			81.2	70.3			82.2	71.6

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 256$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The test model includes only the Mkt-Rf factor. The table reports mean and median estimates of the parameters of the intercept and mean-normalizations, as well as the frequency with which the estimated parameters are found to be statistically significant at the 10% and 5% levels. The table also reports the frequency with which the model is rejected at the 10% and 5% levels based on the OIR test. The test statistic is numerically identical at the first and second GMM steps. Details of the Monte Carlo experiments are provided in the Appendix and main text. Results for the case where $T = 10000$ are in Appendix Table 3.

TABLE 7: **Monte Carlo Experiment: Estimation of Model 3**

	GMM Step 1				GMM Step 2				Iterated GMM			
			% Significant				% Significant				% Significant	
	Mean	Median	10%	5%	Mean	Median	10%	5%	Mean	Median	10%	5%
Intercept Normalization												
δ_S	103	101	100.0	100.0	90.4	90.2	100.0	100.0	88.7	88.5	100.0	100.0
OIR test			3.6	1.3			3.6	1.3			12.4	5.6
Mean Normalization												
γ_S	5.63	43.3	11.9	4.4	0.24	0.23	3.3	1.7	0.47	0.85	30.3	21.7
OIR test			3.5	2.8			3.5	2.8			92.2	86.3

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 256$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The test model includes only a spurious factor, S, that is uncorrelated with returns. The table reports mean and median estimates of the parameters of the intercept and mean-normalizations, as well as the frequency with which the estimated parameters are found to be statistically significant at the 10% and 5% levels. The table also reports the frequency with which the model is rejected at the 10% and 5% levels based on the OIR test. The test statistic is numerically identical at the first and second GMM steps. Details of the Monte Carlo experiments are provided in the Appendix and main text. Results for the case where $T = 10000$ are in Appendix Table 4.

TABLE 8: **Monte Carlo Experiment: Estimation of Model 4**

	GMM Step 1				GMM Step 2				Iterated GMM			
			% Significant				% Significant				% Significant	
	Mean	Median	10%	5%	Mean	Median	10%	5%	Mean	Median	10%	5%
Intercept Normalization												
$\delta_{\mathrm{Mkt-Rf}}$	0.08	0.11	10.3	5.4	0.39	0.39	26.2	17.8	0.43	0.44	34.7	25.5
δ_{S}	99.3	98.1	100.0	100.0	87.1	86.9	100.0	100.0	85.5	85.3	100.0	100.0
OIR test			3.2	1.1			3.2	1.1			11.5	5.3
Mean Normalization												
$\gamma_{\mathrm{Mkt-Rf}}$	3.14	3.14	73.8	64.3	2.81	2.79	77.1	65.6	2.79	2.76	95.6	91.7
γ_{S}	4.68	6.77	54.7	29.1	0.19	0.60	10.1	4.9	0.00	0.79	24.8	16.5
OIR test			22.3	16.8			22.3	16.8			74.6	62.4

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 256$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The test model includes Mkt-Rf and a spurious factor, S, that is uncorrelated with returns. The table reports mean and median estimates of the parameters of the intercept and mean-normalizations, as well as the frequency with which the estimated parameters are found to be statistically significant at the 10% and 5% levels. The table also reports the frequency with which the model is rejected at the 10% and 5% levels based on the OIR test. The test statistic is numerically identical at the first and second GMM steps. Details of the Monte Carlo experiments are provided in the Appendix and main text. Results for the case where $T = 10000$ are in Appendix Table 5.

TABLE 9: **Monte Carlo Experiment: Estimation of Model 5 (The Over-specified Model)**

		GMM Step 1				GMM Step 2				Iterated GMM			
				% Significant				% Significant				% Significant	
	True	Mean	Median	10%	5%	Mean	Median	10%	5%	Mean	Median	10%	5%
Intercept Normalization													
$\delta_{\text{Mkt-Rf}}$	3.25	0.89	0.89	44.0	32.2	0.89	0.90	55.2	44.3	0.89	0.90	55.8	45.7
δ_{SMB}	-0.15	-0.04	-0.05	8.8	4.2	-0.03	-0.03	14.0	8.1	-0.03	-0.03	15.3	9.0
δ_{HML}	4.92	1.35	1.34	47.2	35.2	1.35	1.35	58.7	47.8	1.34	1.35	59.4	48.9
δ_S		74.2	74.3	99.7	99.4	74.9	74.9	100.0	100.0	75.0	75.0	100.0	100.0
OIR test				5.5	2.2			5.5	2.2			6.0	2.4
Mean Normalization													
$\gamma_{\text{Mkt-Rf}}$	3.70	3.77	3.74	98.4	96.2	3.77	3.74	99.0	97.6	3.79	3.76	99.1	97.7
γ_{SMB}	-0.17	-0.19	-0.20	9.1	4.3	-0.19	-0.20	11.2	5.9	-0.19	-0.18	11.7	6.3
γ_{HML}	5.61	5.70	5.64	99.2	98.6	5.71	5.66	99.4	98.9	5.74	5.69	99.6	99.0
γ_S		-0.06	0.47	5.7	1.8	-0.11	0.23	10.6	5.3	-0.19	-0.14	11.3	5.9
OIR test				10.0	5.1			10.0	5.1			11.2	5.6

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 256$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data, but the test model also includes a spurious factor, S, that is uncorrelated with returns. The true values of the parameters of the SDF are indicated in the first column. The table reports mean and median estimates of the parameters of the intercept and mean-normalizations, as well as the frequency with which the estimated parameters are found to be statistically significant at the 10% and 5% levels. The table also reports the frequency with which the model is rejected at the 10% and 5% levels based on the OIR test. The test statistic is numerically identical at the first and second GMM steps. Results for the case where $T = 10000$ are in Appendix Table 6.

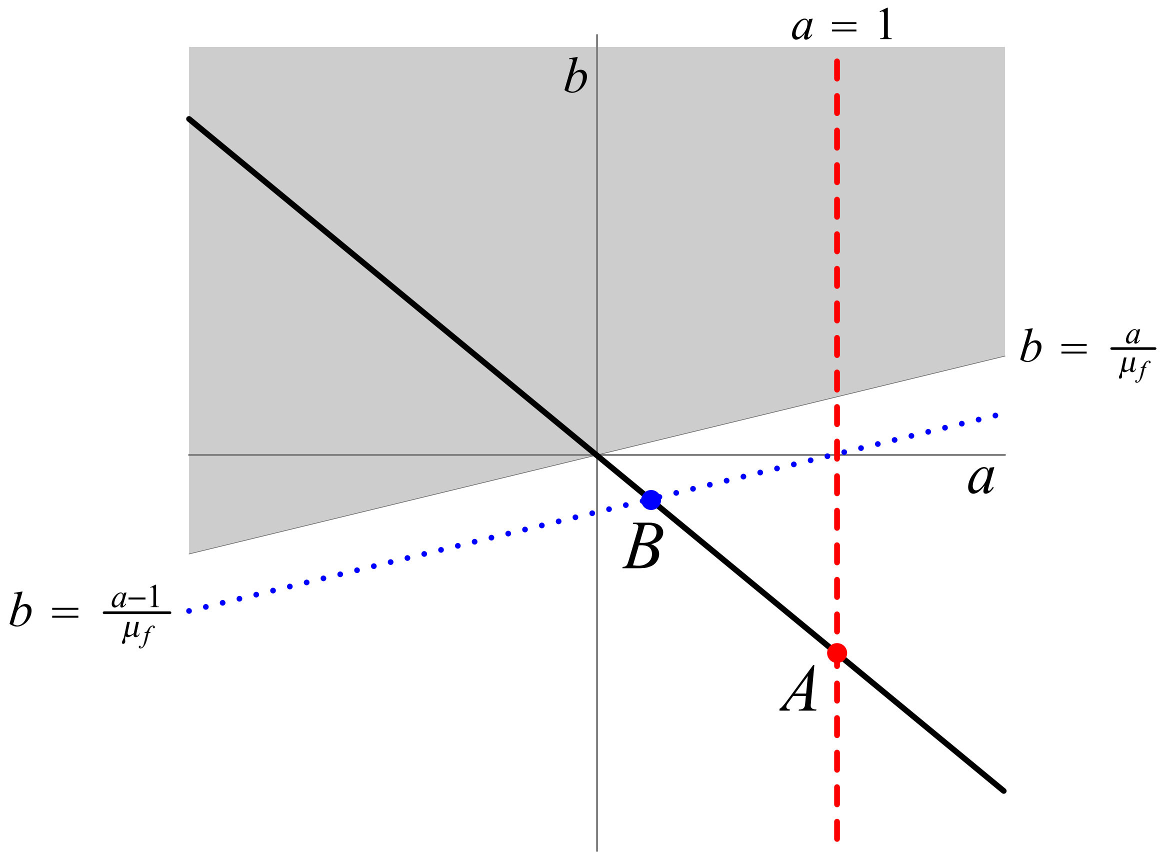
TABLE 10: **Model Reduction using the Mean-Normalization in the Monte Carlo Experiment**

	Model 3		Model 4		Model 5	
	1st/2nd step	Iterated	1st/2nd step	Iterated	1st/2nd step	Iterated
Frequency rank is not reduced	0.121	0.121	0.119	0.119	0.114	0.114
Rejection rate in these cases	3.7%	87.1%	19.5%	64.5%	5.6%	6.0%
Frequency rank is reduced	0.879	0.879	0.881	0.881	0.886	0.886
Rejection rate in these cases	100%	100%	70.3%	71.7%	7.1%	7.1%
Overall rejection rate with model reduction	88.4%	98.5%	64.3%	70.8%	7.0%	6.9%
Rejection rate without model reduction	2.8%	86.3%	16.8%	62.4%	5.1%	5.6%

Notes: The table reports results of the rank reduction procedure from 10000 Monte Carlo experiments with sample size $T = 256$. The nominal size of the OIR tests for the procedure is set to 5%. Details of the Monte Carlo experiments are provided in the Appendix and main text. Results for sample size $T = 10000$ are presented in Appendix Table 7.

FIGURE 1

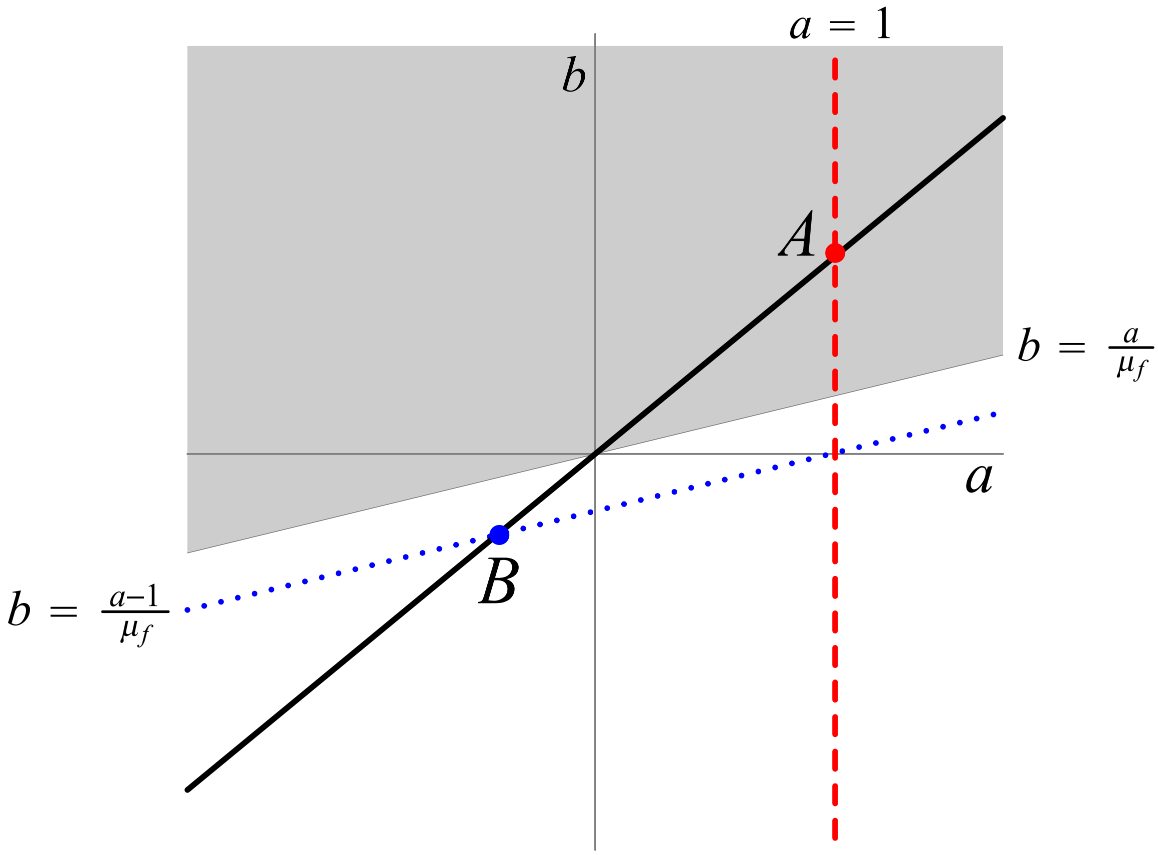
Simple Example with $E(R^e)/E(R^e f) < 1/\mu_f$



Notes: The shaded area denotes the region in which $E(m) < 0$. The heavy solid line denotes the pricing line, $b = [E(R^e)/E(R^e f)]a$. The dashed line denotes all possible parameter pairs where the SDF has a unit intercept. The dotted line denotes all possible parameter pairs where $E(m) = 1$. Point A lies at the intersection of the solid and dashed lines indicating that it corresponds to the unique unit-intercept SDF consistent with the pricing equation. Point B lies at the intersection of the solid and dotted lines indicating that it is the unique unit-mean SDF consistent with the pricing equation.

FIGURE 2

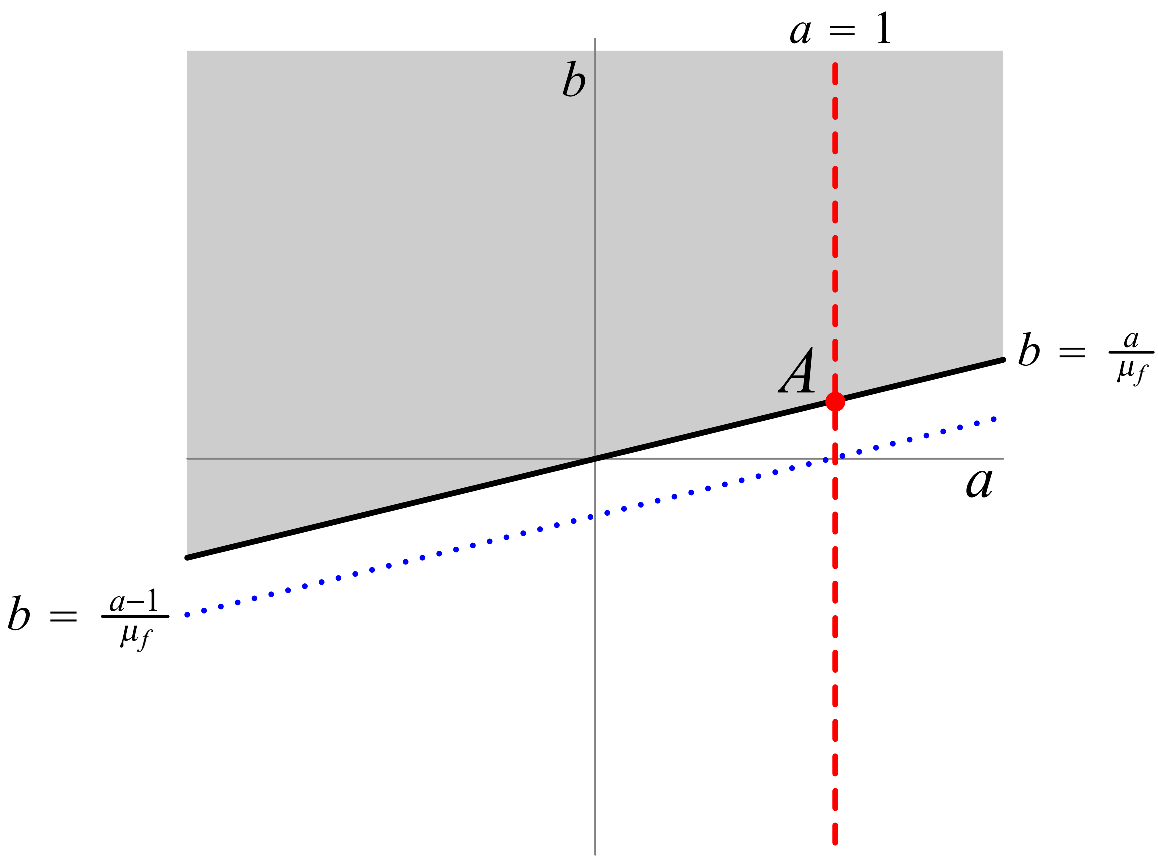
Simple Example with $E(R^e)/E(R^e f) > 1/\mu_f$



Notes: The shaded area denotes the region in which $E(m) < 0$. The heavy solid line denotes the pricing line, $b = [E(R^e)/E(R^e f)]a$. The dashed line denotes all possible parameter pairs where the SDF has a unit intercept. The dotted line denotes all possible parameter pairs where $E(m) = 1$. Point A lies at the intersection of the solid and dashed lines indicating that it corresponds to the unique unit-intercept SDF consistent with the pricing equation. It implies an SDF with a negative mean. Point B lies at the intersection of the solid and dotted lines indicating that it is the unique unit-mean SDF consistent with the pricing equation.

FIGURE 3

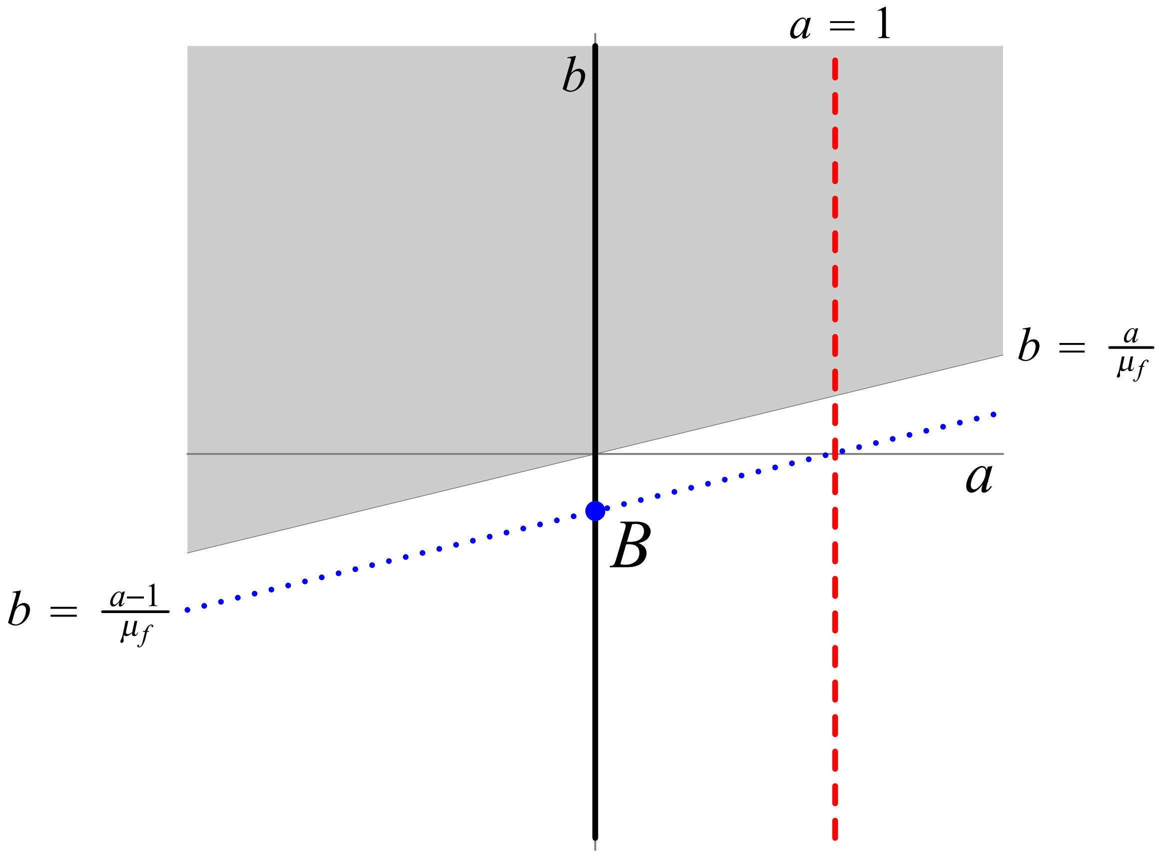
Simple Example with $E(R^e)/E(R^e f) = 1/\mu_f$ or $\mathrm{cov}(R^e, f) = 0$



Notes: The shaded area denotes the region in which $E(m) < 0$. The heavy solid line denotes the pricing line, $b = [E(R^e)/E(R^e f)]a$, which is also the locus where $E(m) = 0$. The dashed line denotes all possible parameter pairs where the SDF has a unit intercept. The dotted line denotes all possible parameter pairs where $E(m) = 1$. Point A lies at the intersection of the solid and dashed lines indicating that it corresponds to the unique unit-intercept SDF consistent with the pricing equation. It implies an SDF with $E(m) = 0$.

FIGURE 4

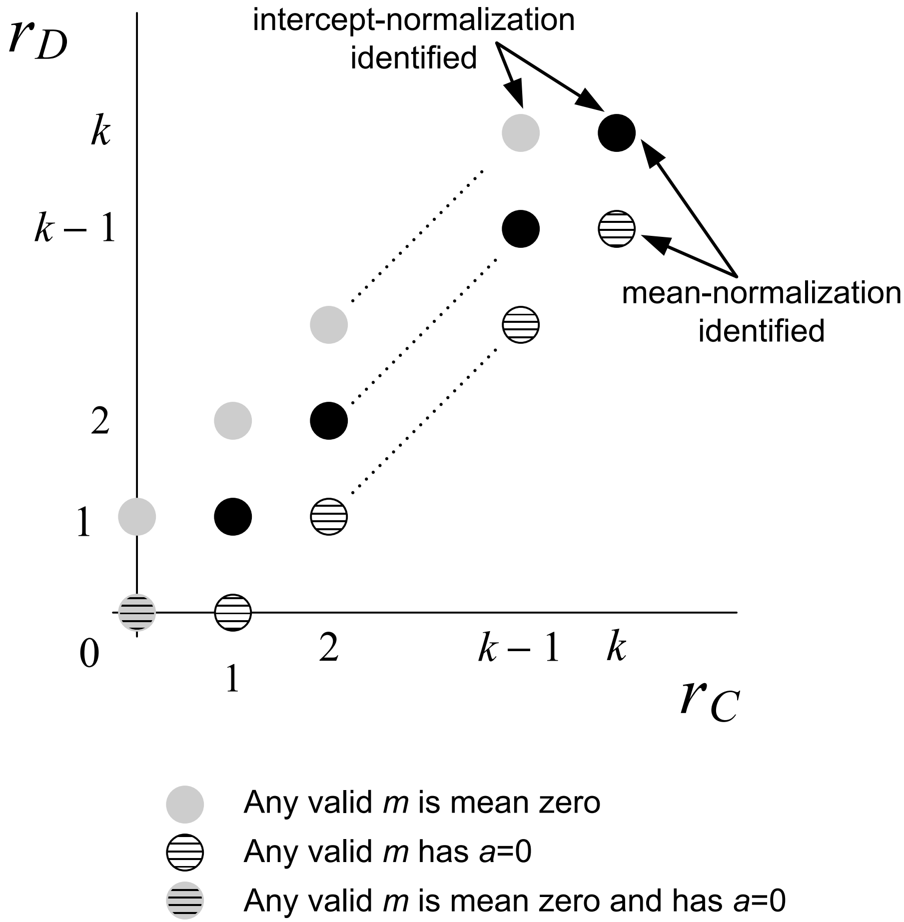
Simple Example with $E(R^e f) = 0$



Notes: The shaded area denotes the region in which $E(m) < 0$. The heavy solid line denotes the pricing line, $b = [E(R^e)/E(R^e f)]a$, which is also the vertical axis in this case. The dashed line denotes all possible parameter pairs where the SDF has a unit intercept. The dotted line denotes all possible parameter pairs where $E(m) = 1$. Point B lies at the intersection of the solid and dotted lines indicating that it is the unique unit-mean SDF consistent with the pricing equation.

FIGURE 5

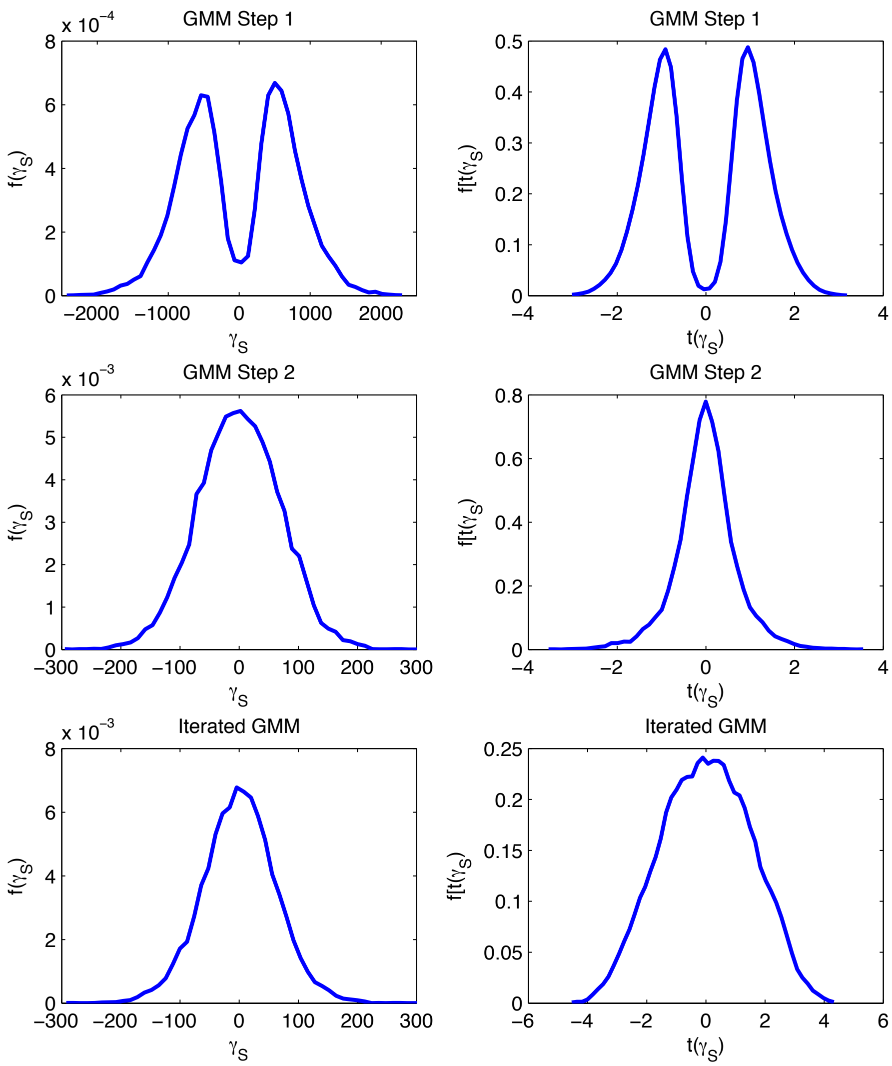
Model Identification and Validity



Notes: Here $r_C = \text{rank}(C)$ and $r_D = \text{rank}(D)$.

FIGURE 6

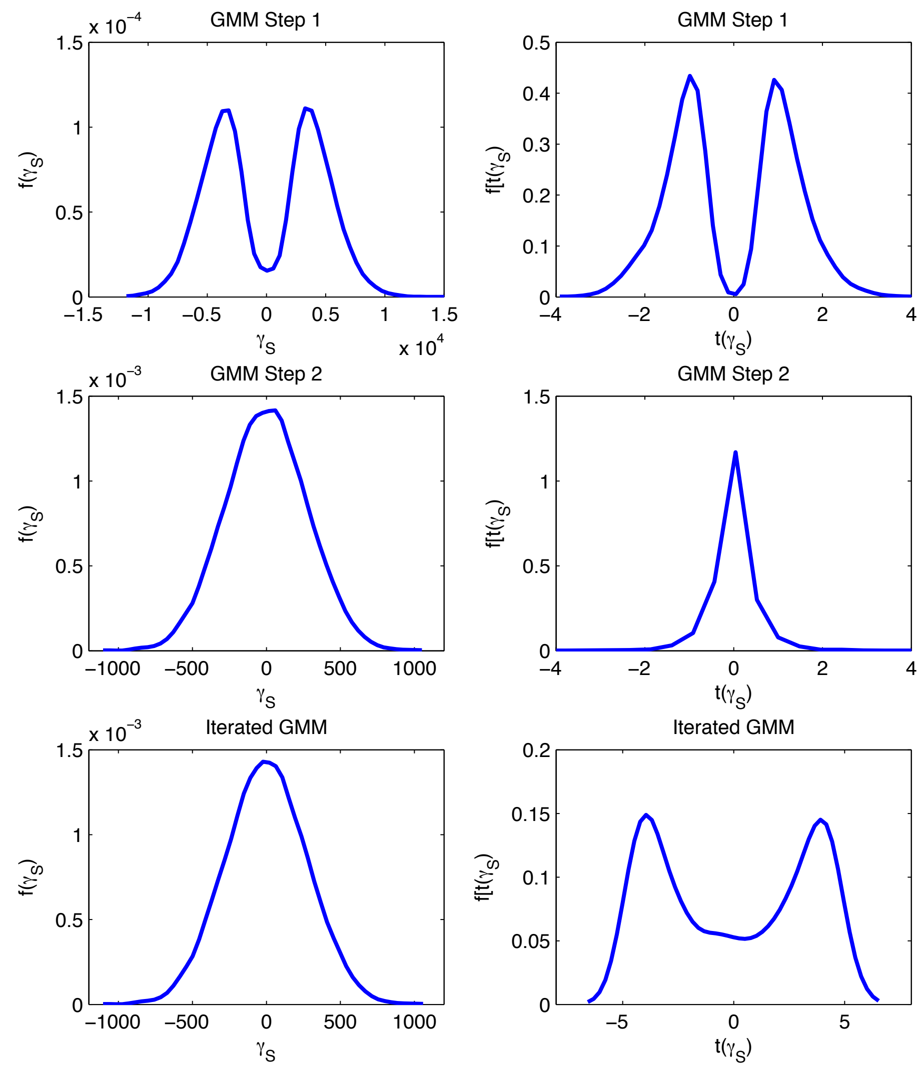
Small Sample Distributions of $\hat{\gamma}_S$ and its Associated t-statistic



Note: The figure reports results from 10000 Monte Carlo experiments with sample size $T = 256$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The test model includes only a spurious factor, S, that is uncorrelated with returns. The graphs report distributions of parameter estimates for the mean-normalization, and the associated t-statistics. Details of the Monte Carlo experiments are provided in the Appendix and main text. Results for the case where $T = 10000$ are in Figure 7.

FIGURE 7

Large Sample Distributions of $\hat{\gamma}_S$ and its Associated t-statistic



Note: The figure reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The test model includes only a spurious factor, S, that is uncorrelated with returns. The graphs report distributions of parameter estimates for the mean-normalization, and the associated t-statistics. Details of the Monte Carlo experiments are provided in the Appendix and main text.

7 Appendix

7.1 Extended Proof of Theorem 2

As in Section 2.3 we have $X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}$ with the columns of X_1 lying in $\mathcal{R}(D)$ and the columns of X_2 lying in $\mathcal{N}(D)$. The matrix X is invertible and $X' = X^{-1}$. We also have $\tilde{D} = DX_1$.

First we write

$$\begin{aligned} D'W^\delta(\mu_R - D\delta) &= XX'D'W^\delta(\mu_R - D\delta) \\ &= X \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} D'W^\delta(\mu_R - D\delta) \\ &= X \begin{pmatrix} \tilde{D}'W^\delta \\ 0 \end{pmatrix} (\mu_R - D\delta). \end{aligned}$$

Then we note that

$$\begin{aligned} \mu_R - D\delta &= \mu_R - DXX'\delta \\ &= \mu_R - D \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \delta \\ &= \mu_R - \begin{pmatrix} \tilde{D} & 0 \end{pmatrix} \begin{pmatrix} X_1'\delta \\ X_2'\delta \end{pmatrix} \\ &= \mu_R - \tilde{D}X_1'\delta \end{aligned}$$

Now let $\delta = X_1\tilde{\delta} + x$ for any $x \in \mathcal{N}(D)$. Then we have

$$\mu_R - D\delta = \mu_R - \tilde{D}\tilde{\delta}.$$

Therefore,

$$D'W^\delta(\mu_R - D\delta) = X \begin{pmatrix} \tilde{D}'W^\delta(\mu_R - \tilde{D}\tilde{\delta}) \\ 0 \end{pmatrix}.$$

Clearly this is zero if and only if $\tilde{\delta} = (\tilde{D}'W^\delta\tilde{D})^{-1}\tilde{D}'W^\delta\mu_R$.$\blacksquare$

7.2 Estimating Long-Run Covariance Matrices

The intercept-normalization

I define $S_T^\delta = \frac{1}{T}\sum_{t=1}^T \hat{u}_t^\delta\hat{u}_t^{\delta\prime}$. When the model is true $E_{t-1}[u_t^\delta(\delta)] = 0$ when evaluated at the true value of δ. Therefore $E[u_t^\delta(\delta)u_{t-j}^\delta(\delta)'] = 0$ for $j \neq 0$. It follows that S_T^δ is a consistent estimate of S^δ.

The mean-normalization

To compute S_T^θ I use a VARHAC procedure proposed by den Haan and Levin (2000). The pricing equations for the assets imply that $E_{t-1}[u_{1t}^\theta(\theta)] = 0$ when evaluated at the true value of θ. Therefore $E[u_{1t}^\theta(\delta)u_{t-j}^\theta(\delta)'] = 0$ for $j > 0$ and $E[u_{1t}^\theta(\delta)u_{1t-j}^\theta(\delta)'] = 0$ for $j < 0$. When I use the VARHAC procedure I make the more restrictive assumption that lagged variables do not appear in the equations for u_{1t}^θ (the errors corresponding to the asset pricing conditions) but allow for lags in the equations for u_{2t}^θ (the GMM errors corresponding to $f_t - \mu$).

7.3 Data

Fama-French 25 Portfolios

Each Fama and French (1993) portfolio represents the intersection of one of 5 groups of stocks sorted according to their market capitalization with one of 5 groups of stocks sorted according to their book equity to market capitalization ratio. The returns are equally weighted. I obtained raw monthly returns from Kenneth French's website http://mba.tuck.dartmouth.edu/ pages/ faculty/ ken.french/ data_library.html. To obtain quarterly returns I compounded monthly returns within each quarter. To obtain quarterly excess returns I subtract the quarterly risk free rate defined as the compounded monthly risk free rate from Fama/French Research Data Factor file. Real excess returns are defined by dividing the nominal excess return by one plus the inflation rate, which I define below.

Consumption Data

To compute real consumption of nondurables and services I proceed as follows. Let C_t^N be the consumption of nondurables and C_t^S be the consumption of services in nominal dollars, and let c_t^N and c_t^S be the corresponding series in constant chained dollars, as published by the Bureau of Economic Analysis. To obtain nominal consumption of nondurables and services I simply set $C_t = C_t^N + C_t^S$. However, because real chained series are not summable, I proceed as follows to create real consumption of nondurables and services, which I denote c_t. First define $s_t = (C_t^N/C_t + C_{t-1}^N/C_{t-1})$, $g_t^N = c_t^N/c_{t-1}^N - 1$ and $g_t^S = c_t^S/c_{t-1}^S - 1$. Then define the growth rate of c_t as $g_t = s_t g_t^N + (1 - s_t)g_t^S$. Notice that a real levels series can then be generated by forward and backward induction relative to a base period. I convert the real levels series into per capita terms by dividing by the quarterly population series published in the National Income and Product Accounts by the BEA. I construct an inflation series using a similar method. Letting π_t^N and π_t^S be the inflation rates for nondurables and services, I let the combined inflation rate be $\pi_t = s_t\pi_t^N + (1 - s_t)\pi_t^S$.

I assume that households derive utility in quarter $t+1$ from the stock of durables at the end of quarter t. To compute the real quarterly stock of durable goods I proceeded as follows. The Bureau of Economic Analysis publishes end-of-year real stocks of durables goods. Let k_t denote the real stock of durables at the end of some year, and let k_{t+4} be the same stock a year (four quarters) later. We observe quarterly real purchases of consumer durables, which I denote c_t^D. I assume that within each year the model

$$k_{t+1} = c_{t+1}^D + (1-\delta)k_t \tag{24}$$

holds, with δ allowed to vary by year. I solve for the value of δ such that the beginning and end-of-year stocks are rationalized by purchases series. This is the δ such that

$$k_{t+4} = c_{t+4}^D + (1-\delta)c_{t+3}^D + (1-\delta)^2 c_{t+2}^D + (1-\delta)^3 c_{t+1}^D + (1-\delta)^4 k_t. \tag{25}$$

Once I identify the value of δ that applies within a year using (25), I use (24) to calculates the within year stocks. I convert the real stocks to per capita terms by dividing by the same population series used for the consumption series.

Fama and French Factors

These series are taken from the Fama/French Research Data Factor file. I define the monthly market return as the sum of the market premium series (Mkt-Rf) and the risk free rate series (Rf). I convert this to a quarterly return by compounding the monthly series geometrically within each quarter. Denoting the resulting series, R_t^M, I convert it to a real return as follows: $r_t^M = (R_t^M - \pi_t)/(1+\pi_t)$.

To create quarterly versions of the Fama-French factors (Mkt-Rf, SMB and HML) I proceed as described in Burnside (2011). To convert them to real excess returns I divide these series by $1+\pi_t$.

Lettau and Ludvigson Factors

Lettau and Ludvigson (2001) propose a scaled CCAPM model, which uses three factors: consumption growth, the *cay* factor (a cointegrating residual between the logarithms of consumption, asset wealth and labor income), and the product of consumption growth and *cay*. I downloaded the factor data directly from the Martin Lettau's web page for the sample period 1952Q1–2012Q3.

Jagannathan and Wang Factors

Jagannathan and Wang (2007) propose a Q4–Q4 CCAPM model. This is simply the CCAPM estimated using annual, rather than quarterly, equity returns, and using annual consumption growth measured from the fourth quarter of one year to the fourth quarter of the year in which the returns are realized. I construct the relevant series from the quarterly data set described above, while constructing annual real excess returns for the FF25 portfolios in similar fashion as to what was described above for quarterly data.

7.4 Lustig and Verdelhan Portfolios and Factors

Lustig and Verdelhan (2007) consider the annual real US dollar excess returns to portfolios of short-term foreign government securities denominated in foreign currency. The sample period is 1953–2002. They form these portfolios on the basis of the interest rates on the underlying securities. In particular the real excess returns on a large number of countries' treasury securities are sorted into eight bins in each period according to the nominal interest rates on the securities, from lowest to highest. The returns to holding equally-weighted portfolios of each bin are then calculated. Lustig and Verdelhan use three risk factors to explain these returns: consumption growth, durables growth and the market return [their model is equivalent to Yogo (2006)'s model]. I take the data for the returns and factors directly from the AEA data repository for their paper.

7.5 Rank Tests

Cragg and Donald (1997)

The Cragg and Donald (1997) and Wright (2003) test for whether B has rank $r < k$ is based on measuring the distance between B and the set of matrices of the same dimension with rank r. Let $\hat{B}$ be a consistent estimator for B and assume that $\sqrt{T}\,\mathrm{vec}(\hat{B} - B_0) \xrightarrow{d} N(0, V_B)$, where B_0 is the true value of B. Let $\hat{V}_B$ be a consistent estimator for V_B. To test the null hypothesis that $\mathrm{rank}(B) = r < k$ I form the statistic

$$L(r) = \min_{P \in \Omega_r} T\,\mathrm{vec}(\hat{B} - P)' V(\hat{B})^{-1}\,\mathrm{vec}(\hat{B} - P)$$

where Ω_r is the set of all $n \times k$ matrices with rank r. If the true rank of B_0 is r, $L(r) \xrightarrow{d} \chi^2_{(n-r)(k-r)}$. I construct tests of the rank conditions for the two normalizations by letting B be C or D and estimating the elements of these matrices by GMM.

To take an example, when the null hypothesis is that $r = 0$, the test is analogous to a simple F-test for $B_{ij} = 0$ for all i, j. In the case the test statistic can be computed with ease and has a chi-squared distribution with nk degrees of freedom. However, when the null hypothesis is that $r_C = 1$, computing $L(1)$ can be computationally burdensome, especially when n and k are large. It involves optimization over $nk - (n-1)(k-1) = k + n - 1$

free parameters since the single vector that forms the basis for the rows of B has $k-1$ free parameters and there are n rows.

Kleibergen and Paap (2006)

We start with an $n \times k$ matrix Π, and the null hypothesis that $\operatorname{rank}(\Pi) = r < k$. We also define $\pi = \operatorname{vec}(\Pi)$ and assume that we have some estimate $\hat{\pi}$ with the property that $\sqrt{T}(\hat{\pi} - \pi) \xrightarrow{d} N(0, V_\pi)$ as well as a consistent estimate of V_π, denoted $\hat{V}_\pi$.

We form the scaled matrix $\Theta = G\Pi F'$ where $G_{n\times n}$ and $F_{k\times k}$ are invertible matrices that make Θ invariant to invertible transformations of the data. As described in the main text, when $\Pi = \operatorname{cov}(R^e, f)$ natural choices are $G = \Sigma_{R^e}^{-1/2}$ and $F = \Sigma_f^{-1/2}$, the Cholesky decompositions of the inverses of the covariance matrices of R^e and f.

If we define $\theta = \operatorname{vec}(\Theta)$ then $\theta = (F \otimes G)\pi$. Now let $\hat{\Theta} = \hat{G}\hat{\Pi}\hat{F}'$ where $\hat{G}$ and $\hat{F}$ are standard sample analogs of G and F. Then, under standard assumptions, $\hat{\theta} = \operatorname{vec}(\hat{\Theta})$ has the property that $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta)$ with $V_\theta = (F \otimes G)V_\pi(F \otimes G)'$.

Now let $USV' = \Theta$ be the SVD of Θ and let $U_{n\times n}$, $S_{n\times k}$ and $V_{k\times k}$ be partitioned as follows:

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \qquad S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \qquad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

where U_{12} is $r \times (n-r)$, U_{22} is $(n-r)\times(n-r)$, S_2 is $(n-r)\times(k-r)$, V_{12} is $r\times(k-r)$ and V_{22} is $(k-r)\times(k-r)$. Similarly let $\hat{U}\hat{S}\hat{V}' = \hat{\Theta}$ be the SVD of $\hat{\Theta}$. Let

$$\hat{\Lambda}_r = (\hat{U}_{22}\hat{U}_{22}')^{-1/2}\hat{U}_{22}\hat{S}_2\hat{V}_{22}'(\hat{V}_{22}\hat{V}_{22}')^{-1/2},$$

$$A_r = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} U_{22}^{-1}(U_{22}U_{22}')^{1/2},$$

$$B_r = (V_{22}V_{22}')^{1/2}(V_{22}')^{-1}\begin{pmatrix} V_{12}' & V_{22}' \end{pmatrix},$$

and $\hat{\lambda}_r = \operatorname{vec}(\hat{\Lambda}_r)$.

Under the null hypothesis that $\operatorname{rank}(\Pi) = r$, $\sqrt{T}\hat{\lambda}_r$ converges in distribution to a normal distribution with mean 0 and covariance matrix $\Omega_r = (B_r \otimes A_r')V_\theta(B_r \otimes A_r')'$. Therefore, $T\hat{\lambda}_r'\Omega_r^{-1}\hat{\lambda}_r$ converges to a $\chi^2_{(n-r)(k-r)}$. In practice, a consistent estimator for Ω_r, $\hat{\Omega}_r = (\hat{B}_r \otimes \hat{A}_r')\hat{V}_\theta(\hat{B}_r \otimes \hat{A}_r')'$, is used to form the test statistic $\operatorname{rk}(r) = T\hat{\lambda}_r'\hat{\Omega}_r^{-1}\hat{\lambda}_r$ with $\hat{A}_r$ and $\hat{B}_r$ being sample analogs of A_r and B_r.

7.6 Monte Carlo Experiment Details

The true SDF is given by $m_t^\gamma = 1 - (f_t - \mu_f)'\gamma$ with μ_f and γ being 3×1 vectors. The vector of risk factors, f_t, is assumed to follow the law of motion $f_t \sim Niid(\mu_f, \Sigma_f)$. I set $\gamma = \begin{pmatrix} 3.70 & -0.17 & 5.61 \end{pmatrix}'$. This parameter vector corresponds to first step GMM estimates of the Fama-French three factor model obtained using the mean-normalization using real, rather than nominal, excess returns and factors to estimate the model. I estimate the model in real terms because some of the test models in the Monte Carlo experiments use synthetic consumption factors. I set μ_f and Σ_f equal to the sample mean and covariance matrix of the Mkt-Rf, SMB and HML factors.

I generate an $n \times 1$ (with $n = 25$) vector of artificial excess returns $R_t^e = \mu_R + \beta(f_t - \mu_f) + \Psi\xi_t$ where μ_R is an $n \times 1$ vector, β is an $n \times k$ matrix, Ψ is an $n \times n$ lower triangular matrix, $\xi_t \sim Niid(0, I_n)$ and is independent of f_t. Given this law of motion for R_t^e, it follows that the covariance matrix of R_t^e is $\Sigma_R = \beta\Sigma_f\beta' + \Psi\Psi'$.

So that the model shares some characteristics of actual data, I set Σ_R equal to its sample equivalent for the Fama-French 25 portfolios sorted on size and value. I set β equal to the matrix of factor betas for these returns regressed on Mkt-Rf, SMB and HML. I set Ψ equal to the Cholesky decomposition of the covariance matrix of the residuals from those regressions.

Given the assumptions above we have

$$\begin{aligned} E(R_t^e m_t) &= E\left\{[\mu_R + \beta(f_t - \mu_f) + \Psi\xi_t][1 - (f_t - \mu_f)'\gamma]\right\} \\ &= \mu_R - \beta\Sigma_f\gamma. \end{aligned} \tag{26}$$

To ensure that the right hand side of equation (26) is zero, I set $\mu_R = \beta\Sigma_f\gamma$. This means that expected returns of the model correspond to the model-predicted expected returns for the GMM estimates described above.

7.7 Non-Uniqueness of the SVD and the Model Reduction Procedure

As mentioned in footnote 10, the model reduction procedure relies on identifying linear combinations of factors using the SVD of the matrix $\Theta = USV'$. This might be viewed as problematic given that V in the SVD is sometimes non-unique.

There are two types of non-uniqueness. Neither causes difficulty with the procedure as long as the researcher's primary interest is in recovering the parameters associated with the original factors. Recall that the procedure identifies $\tilde{f}_r = A_r f$ as the reduced set of risk factors, where $V_r'F$ and V_r represents the first r columns of V.

The first type of non-uniqueness is that it is usually possible to swap the signs of columns in V and corresponding columns in U. So, we could swap the signs of the elements of V_r. Of course, this reverses the signs of the factors, and the resulting coefficients, $\tilde{\delta}$ and $\tilde{\gamma}$. Therefore, if one interested in backing out the implied δ and γ one ends up with the same values as before. This is because the transformations $\delta = F'V_r\tilde{\delta}$ and $\gamma = F'V_r\tilde{\gamma}$ negate the change of sign.

The second type of non-uniqueness arises if there are repeat singular values. In this the columns of V associated with these repeat singular values can be rotated using any orthogonal matrix. So, now, imagine that this issue applies to two (or more) of the columns in V_r. This means that we could find one SVD and define $\tilde{f}_r = (V_r' F) f$ and, using an arbitrary unitary rotation matrix $Q_{r\times r}$ define $\bar{V}_r = V_r Q$ and let $\bar{f}_r = (\bar{V}_r' F) f$.

For the model based on the $\tilde{f}_r$ we have

$$
\begin{aligned}
\tilde{D} &= E(R^e \tilde{f}_r') = E(R^e f') F' V_r = D F' V_r \\
\tilde{C} &= \operatorname{cov}(R^e, \tilde{f}_r) = \operatorname{cov}(R^e, f) F' V_r = C F' V_r.
\end{aligned}
$$

For the model based on the $\bar{f}_r$ we have

$$
\begin{aligned}
\bar{D} &= E(R^e \bar{f}_r') = E(R^e f') F' \bar{V}_r = D F' V_r Q = \tilde{D} Q \\
\bar{C} &= \operatorname{cov}(R^e, \bar{f}_r) = \operatorname{cov}(R^e, f) F' \bar{V}_r = C F' V_r Q = \tilde{C} Q.
\end{aligned}
$$

Now consider the population version of first step GMM with the mean normalization for these reduced factors. We have

$$
\bar{\gamma} = (\bar{C}'\bar{C})^{-1}(\bar{C}'\mu_R) = (Q'\tilde{C}'\tilde{C}Q)^{-1}(Q'\tilde{C}'\mu_R) = Q'(\tilde{C}'\tilde{C})^{-1}(\tilde{C}'\mu_R) = Q'\tilde{\gamma}.
$$

Similarly

$$
\bar{\delta} = Q'\tilde{\delta}.
$$

Thus the parameters associated with the original factors are the same regardless of which rotation we use. That is $\gamma \equiv A_r'\tilde{\gamma}$ is the same as

$$
\gamma \equiv \bar{A}_r'\bar{\gamma} = (\bar{V}_r' F)'(Q'\tilde{\gamma}) = F'\bar{V}_r Q'\tilde{\gamma} = F' V_r \tilde{\gamma} = A_r'\tilde{\gamma},
$$

and $\delta \equiv A_r'\tilde{\delta}$ is the same as $\delta \equiv \bar{A}_r'\bar{\delta}$.

Since the parameters associated with the different rotations are equivalent for the original factors, the GMM errors at the first step are the same. So we get identical weighting matrices at the next step of GMM. This makes the estimators at all GMM steps equivalent, regardless of the rotation used.

APPENDIX TABLE 1: **Rank Tests in the Monte Carlo Experiment**

	% rejected at 10% level	% rejected at 5% level		% rejected at 10% level	% rejected at 5% level
True Model (Factors: Mkt-Rf, SMB, HML) $r_D = r_C = 3$					
$H_0: \ r_D = 2$	100%	100%	$H_0: \ r_C = 2$	100%	100%
$H_0: \ r_D = 1$	100%	100%	$H_0: \ r_C = 1$	100%	100%
$H_0: \ r_D = 0$	100%	100%	$H_0: \ r_C = 0$	100%	100%
Single Relevant Factor (Factor: Mkt-Rf) $r_D = r_C = 1$					
$H_0: \ r_D = 0$	100%	100%	$H_0: \ r_C = 0$	100%	100%
Single Spurious Factor (Factor: S) $r_D = 1$, $r_C = 0$					
$H_0: \ r_D = 0$	100%	100%	$H_0: \ r_C = 0$	10.4%	5.1%
Two Factors (Factors: Mkt-Rf, S) $r_D = 2$, $r_C = 1$					
$H_0: \ r_D = 1$	100%	100%	$H_0: \ r_C = 1$	10.3%	4.9%
$H_0: \ r_D = 0$	100%	100%	$H_0: \ r_C = 0$	100%	100%
5. Over-specified Model (Factors: Mkt-Rf, SMB, HML, S) $r_D = r_C = 3$					
$H_0: \ r_D = 3$	10.1%	5.3%	$H_0: \ r_C = 3$	10.2%	5.2%
$H_0: \ r_D = 2$	100%	100%	$H_0: \ r_C = 2$	100%	100%
$H_0: \ r_D = 1$	100%	100%	$H_0: \ r_C = 1$	100%	100%
$H_0: \ r_D = 0$	100%	100%	$H_0: \ r_C = 0$	100%	100%

Notes: The table reports results of Kleibergen and Paap (2006) rank tests from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. Rank tests are performed for five test models that use different factors. $r_C = \text{rank}[\text{cov}(R^e, f)]$ and $r_D = \text{rank}[E(R^e f')]$, where R^e are the returns generated in the experiment, and f is the conjectured vector of factors. The table reports the fraction of the samples in which these tests reject the null hypothesis when the size of the test is set to 10% and 5%. Details of the Monte Carlo experiments are provided in the Appendix and main text.

APPENDIX TABLE 2: **Monte Carlo Experiment: Estimation of Model 1 (The True Model)**

		GMM Step 1				GMM Step 2				Iterated GMM			
				% Significant				% Significant				% Significant	
	True	Mean	Median	10%	5%	Mean	Median	10%	5%	Mean	Median	10%	5%
Intercept Normalization													
$\delta_{\text{Mkt}-\text{Rf}}$	3.28	3.28	3.27	100.0	100.0	3.29	3.29	100.0	100.0	3.29	3.29	100.0	100.0
δ_{SMB}	-0.15	-0.15	-0.15	18.4	10.8	-0.15	-0.15	18.9	10.9	-0.15	-0.15	18.9	10.9
δ_{HML}	4.97	4.97	4.97	100.0	100.0	4.98	4.98	100.0	100.0	4.98	4.98	100.0	100.0
OIR test				9.7	5.0			9.7	5.0			9.7	5.0
Mean Normalization													
$\gamma_{\text{Mkt}-\text{Rf}}$	3.70	3.70	3.70	100.0	100.0	3.70	3.70	100.0	100.0	3.70	3.70	100.0	100.0
γ_{SMB}	-0.17	-0.17	-0.17	18.4	10.8	-0.17	-0.17	18.8	10.9	-0.17	-0.17	18.8	10.9
γ_{HML}	5.61	5.62	5.62	100.0	100.0	5.62	5.62	100.0	100.0	5.62	5.62	100.0	100.0
OIR test				9.7	5.0			9.7	5.0			9.7	5.1

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The true values of the parameters of the SDF are indicated in the first column. The table reports mean and median estimates of the parameters of the intercept and mean-normalizations, as well as the frequency with which the estimated parameters are found to be statistically significant at the 10% and 5% levels. The table also reports the frequency with which the model is rejected at the 10% and 5% levels based on the OIR test. The test statistic is numerically identical at the first and second GMM steps. Details of the Monte Carlo experiments are provided in the Appendix and main text.

APPENDIX TABLE 3: **Monte Carlo Experiment: Estimation of Model 2**

	GMM Step 1				GMM Step 2				Iterated GMM			
			% Significant				% Significant				% Significant	
	Mean	Median	10%	5%	Mean	Median	10%	5%	Mean	Median	10%	5%
Intercept Normalization												
$\delta_{\text{Mkt}-\text{Rf}}$	3.06	3.06	100.0	100.0	3.03	3.04	100.0	100.0	3.03	3.03	100.0	100.0
OIR test			100.0	100.0			100.0	100.0			100.0	100.0
Mean Normalization												
$\gamma_{\text{Mkt}-\text{Rf}}$	3.24	3.24	100.0	100.0	2.78	2.78	100.0	100.0	2.74	2.74	100.0	100.0
OIR test			100.0	100.0			100.0	100.0			100.0	100.0

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The test model includes only the Mkt-Rf factor. The table reports mean and median estimates of the parameters of the intercept and mean-normalizations, as well as the frequency with which the estimated parameters are found to be statistically significant at the 10% and 5% levels. The table also reports the frequency with which the model is rejected at the 10% and 5% levels based on the OIR test. The test statistic is numerically identical at the first and second GMM steps. Details of the Monte Carlo experiments are provided in the Appendix and main text.

APPENDIX TABLE 4: **Monte Carlo Experiment: Estimation of Model 3**

	GMM Step 1				GMM Step 2				Iterated GMM			
			% Significant				% Significant				% Significant	
	Mean	Median	10%	5%	Mean	Median	10%	5%	Mean	Median	10%	5%
Intercept Normalization												
δ_S	102	102	100.0	100.0	101	101	100.0	100.0	101	101	100.0	100.0
OIR test			9.4	4.4			9.4	4.4			10.5	5.0
Mean Normalization												
γ_S	9.22	-185	18.7	10.4	-0.39	-0.47	2.2	1.6	-0.52	-1.71	82.3	78.4
OIR test			3.6	3.4			3.6	3.4			99.9	99.8

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The test model includes only a spurious factor, S, that is uncorrelated with returns. The table reports mean and median estimates of the parameters of the intercept and mean-normalizations, as well as the frequency with which the estimated parameters are found to be statistically significant at the 10% and 5% levels. The table also reports the frequency with which the model is rejected at the 10% and 5% levels based on the OIR test. The test statistic is numerically identical at the first and second GMM steps. Details of the Monte Carlo experiments are provided in the Appendix and main text.

APPENDIX TABLE 5: **Monte Carlo Experiment: Estimation of Model 4**

	GMM Step 1				GMM Step 2				Iterated GMM			
			% Significant				% Significant				% Significant	
	Mean	Median	10%	5%	Mean	Median	10%	5%	Mean	Median	10%	5%
Intercept Normalization												
$\delta_{\text{Mkt}-\text{Rf}}$	0.00	0.00	9.9	4.9	0.03	0.03	11.6	6.1	0.03	0.03	11.9	6.3
δ_{S}	102	102	100.0	100.0	101	101	100.0	100.0	101	101	100.0	100.0
OIR test			9.1	4.1			9.1	4.1			10.9	5.4
Mean Normalization												
$\gamma_{\text{Mkt}-\text{Rf}}$	3.11	3.18	83.3	77.1	2.75	2.75	91.1	79.1	2.74	2.74	100.0	100.0
γ_{S}	-18.1	-39.7	91.5	76.8	-1.17	-4.35	4.6	2.2	-1.19	-3.55	77.4	72.4
OIR test			31.5	30.0			31.5	30.0			99.8	99.8

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The test model includes Mkt-Rf and a spurious factor, S, that is uncorrelated with returns. The table reports mean and median estimates of the parameters of the intercept and mean-normalizations, as well as the frequency with which the estimated parameters are found to be statistically significant at the 10% and 5% levels. The table also reports the frequency with which the model is rejected at the 10% and 5% levels based on the OIR test. The test statistic is numerically identical at the first and second GMM steps. Details of the Monte Carlo experiments are provided in the Appendix and main text.

APPENDIX TABLE 6: **Monte Carlo Experiment: Estimation of Model 5 (The Over-specified Model)**

		GMM Step 1				GMM Step 2				Iterated GMM			
				% Significant				% Significant				% Significant	
	True	Mean	Median	10%	5%	Mean	Median	10%	5%	Mean	Median	10%	5%
Intercept Normalization													
$\delta_{\text{Mkt-Rf}}$	3.25	0.89	0.89	69.4	58.9	0.89	0.89	83.6	75.2	0.89	0.89	84.1	76.0
δ_{SMB}	-0.15	-0.04	-0.04	10.5	5.7	-0.04	-0.04	10.8	5.9	-0.04	-0.04	11.2	6.2
δ_{HML}	4.92	1.35	1.36	69.7	59.1	1.35	1.35	83.8	75.8	1.35	1.35	84.2	76.7
δ_S		74.1	74.1	99.8	99.5	74.2	74.2	100.0	100.0	74.2	74.2	100.0	100.0
OIR test				7.0	3.1			7.0	3.1			8.2	3.7
Mean Normalization													
$\gamma_{\text{Mkt-Rf}}$	3.70	3.70	3.70	100.0	100.0	3.70	3.70	100.0	100.0	3.70	3.70	100.0	100.0
γ_{SMB}	-0.17	-0.17	-0.17	15.9	9.0	-0.17	-0.17	16.6	9.4	-0.16	-0.17	17.1	9.8
γ_{HML}	5.61	5.62	5.62	100.0	100.0	5.62	5.62	100.0	100.0	5.62	5.62	100.0	100.0
γ_S		-0.21	0.21	6.6	1.9	-0.36	-0.59	7.3	3.1	-0.36	-0.55	7.8	3.0
OIR test				6.5	2.8			6.5	2.8			7.6	3.4

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data, but the test model also includes a spurious factor, S, that is uncorrelated with returns. The true values of the parameters of the SDF are indicated in the first column. The table reports mean and median estimates of the parameters of the intercept and mean-normalizations, as well as the frequency with which the estimated parameters are found to be statistically significant at the 10% and 5% levels. The table also reports the frequency with which the model is rejected at the 10% and 5% levels based on the OIR test. The test statistic is numerically identical at the first and second GMM steps.

APPENDIX TABLE 7: **Model Reduction using the Mean-Normalization in the Monte Carlo Experiment**

	Model 3		Model 4		Model 5	
	1st/2nd step	Iterated	1st/2nd step	Iterated	1st/2nd step	Iterated
Frequency rank is not reduced	0.051	0.051	0.049	0.049	0.052	0.052
Rejection rate in these cases	3.2%	100%	38.1%	100%	3.5%	4.3%
Frequency rank is reduced	0.949	0.949	0.951	0.951	0.948	0.948
Rejection rate in these cases	100%	100%	100%	100%	5.1%	5.1%
Overall rejection rate with model reduction	95.1%	100%	97.0%	100%	5.0%	5.0%
Rejection rate without model reduction	3.4%	99.8%	30.0%	99.8%	3.4%	3.8%

Notes: The table reports results of the rank reduction procedure from 10000 Monte Carlo experiments with sample size $T = 10000$. The nominal size of the OIR tests for the procedure is set to 5%. Details of the Monte Carlo experiments are provided in the Appendix and main text.

APPENDIX TABLE 8: **Distributions of Direct and Indirect Parameter Estimates: Model 1**

	GMM Step 1			GMM Step 2			Iterated GMM		
	P10	Median	P90	P10	Median	P90	P10	Median	P90
Intercept Normalization									
$\delta_{\mathrm{Mkt-Rf}}$	3.12	3.27	3.43	3.13	3.29	3.44	3.13	3.29	3.44
δ_{SMB}	-0.41	-0.15	0.12	-0.40	-0.15	0.11	-0.40	-0.15	0.11
δ_{HML}	4.76	4.97	5.18	4.78	4.98	5.20	4.78	4.98	5.20
$\delta(\gamma)_{\mathrm{Mkt-Rf}}$	3.12	3.27	3.43	3.12	3.27	3.43	3.12	3.27	3.43
$\delta(\gamma)_{\mathrm{SMB}}$	-0.41	-0.15	0.12	-0.40	-0.15	0.11	-0.40	-0.15	0.11
$\delta(\gamma)_{\mathrm{HML}}$	4.76	4.97	5.18	4.76	4.97	5.18	4.76	4.97	5.18
Mean Normalization									
$\gamma_{\mathrm{Mkt-Rf}}$	3.51	3.70	3.90	3.51	3.70	3.89	3.51	3.70	3.89
γ_{SMB}	-0.47	-0.17	0.13	-0.45	-0.17	0.12	-0.45	-0.17	0.12
γ_{HML}	5.34	5.62	5.89	5.35	5.62	5.89	5.35	5.62	5.89
$\gamma(\delta)_{\mathrm{Mkt-Rf}}$	3.51	3.70	3.90	3.52	3.72	3.91	3.52	3.72	3.91
$\gamma(\delta)_{\mathrm{SMB}}$	-0.47	-0.17	0.13	-0.45	-0.17	0.12	-0.45	-0.17	0.12
$\gamma(\delta)_{\mathrm{HML}}$	5.34	5.62	5.89	5.37	5.64	5.91	5.37	5.64	5.91

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The table reports 10th percentiles, medians, and 90th percentiles of the distributions of the parameter estimates for the true model. Details of the Monte Carlo experiments are provided in the Appendix and main text.

APPENDIX TABLE 9: **Distributions of Direct and Indirect Parameter Estimates: Model 2**

	GMM Step 1			GMM Step 2			Iterated GMM		
	P10	Median	P90	P10	Median	P90	P10	Median	P90
Intercept Normalization									
$\delta_{\text{Mkt}-\text{Rf}}$	2.91	3.06	3.21	2.87	3.04	3.20	2.87	3.03	3.19
$\delta(\gamma)_{\text{Mkt}-\text{Rf}}$	2.90	3.05	3.20	2.50	2.64	2.78	2.46	2.61	2.75
Mean Normalization									
$\gamma_{\text{Mkt}-\text{Rf}}$	3.06	3.24	3.41	2.61	2.78	2.95	2.58	2.74	2.91
$\gamma(\delta)_{\text{Mkt}-\text{Rf}}$	3.07	3.25	3.43	3.03	3.22	3.41	3.02	3.22	3.41

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The table reports 10th percentiles, medians, and 90th percentiles of the distributions of the parameter estimates for Model 2, which includes only Mkt-Rf as a risk factor. Details of the Monte Carlo experiments are provided in the Appendix and main text.

APPENDIX TABLE 10: **Distributions of Direct and Indirect Parameter Estimates: Model 3**

	GMM Step 1			GMM Step 2			Iterated GMM		
	P10	Median	P90	P10	Median	P90	P10	Median	P90
Intercept Normalization									
δ_S	99.0	102	105	99.1	101	103	99.1	101	103
$\delta(\gamma)_S$	97.9	102	106	-43.2	75.7	236	-42.1	75.7	235
Mean Normalization									
γ_S	-5617	-185	5617	-354	0	357	-355	-1.71	356
$\gamma(\delta)_S$	-19720	-1502	17535	-22426	4657	24897	-22385	4674	24884

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The table reports 10th percentiles, medians, and 90th percentiles of the distributions of the parameter estimates for Model 3, which includes only a spurious risk factor, S. Details of the Monte Carlo experiments are provided in the Appendix and main text.

APPENDIX TABLE 11: **Distributions of Direct and Indirect Parameter Estimates: Model 4**

	GMM Step 1			GMM Step 2			Iterated GMM		
	P10	Median	P90	P10	Median	P90	P10	Median	P90
Intercept Normalization									
$\delta_{\text{Mkt}-\text{Rf}}$	-0.12	0.00	0.12	-0.09	0.03	0.14	-0.09	0.03	0.14
δ_S	99.0	102	105	98.2	101	104	98.2	101	104
$\delta(\gamma)_{\text{Mkt}-\text{Rf}}$	-0.58	0.03	0.56	-4.01	0.89	4.70	-4.01	0.89	4.86
$\delta(\gamma)_S$	83.4	98.3	121	-82.3	67.3	261	-87.5	66.8	259
Mean Normalization									
$\gamma_{\text{Mkt}-\text{Rf}}$	1.74	3.18	4.34	2.50	2.75	2.99	2.50	2.74	2.98
γ_S	-2000	-39.7	1991	-292	-4.35	294	-292	-3.55	292
$\gamma(\delta)_{\text{Mkt}-\text{Rf}}$	-9.03	2.77	13.9	-7.93	2.70	12.5	-7.92	2.70	12.5
$\gamma(\delta)_S$	-20218	2166	20421	-16466	3847	20273	-16438	3849	20288

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The table reports 10th percentiles, medians, and 90th percentiles of the distributions of the parameter estimates for Model 4, which includes Mkt-Rf and a spurious risk factor, S. Details of the Monte Carlo experiments are provided in the Appendix and main text.

APPENDIX TABLE 12: **Distributions of Direct and Indirect Parameter Estimates: Model 5**

	GMM Step 1			GMM Step 2			Iterated GMM		
	P10	Median	P90	P10	Median	P90	P10	Median	P90
Intercept Normalization									
$\delta_{\text{Mkt-Rf}}$	0.36	0.89	1.42	0.47	0.89	1.32	0.47	0.89	1.32
δ_{SMB}	-0.18	-0.04	0.10	-0.18	-0.04	0.10	-0.18	-0.04	0.10
δ_{HML}	0.55	1.36	2.15	0.71	1.35	1.99	0.71	1.35	1.99
δ_S	57.9	74.1	90.2	61.3	74.2	86.9	61.3	74.2	86.9
$\delta(\gamma)_{\text{Mkt-Rf}}$	2.01	3.21	7.06	2.22	3.28	6.03	2.22	3.28	6.02
$\delta(\gamma)_{\text{SMB}}$	-0.55	-0.14	0.14	-0.50	-0.15	0.12	-0.50	-0.15	0.12
$\delta(\gamma)_{\text{HML}}$	3.04	4.85	10.8	3.37	4.97	9.13	3.37	4.97	9.12
$\delta(\gamma)_S$	-119	2.08	39.1	-85.0	-0.10	32.7	-85.0	0.03	32.7
Mean Normalization									
$\gamma_{\text{Mkt-Rf}}$	3.50	3.70	3.91	3.50	3.70	3.90	3.50	3.70	3.90
γ_{SMB}	-0.47	-0.17	0.14	-0.46	-0.17	0.13	-0.46	-0.17	0.13
γ_{HML}	5.34	5.61	5.90	5.34	5.62	5.90	5.34	5.62	5.90
γ_S	-67.4	0.21	66.0	-53.7	-0.59	53.2	-53.7	-0.55	53.1
$\gamma(\delta)_{\text{Mkt-Rf}}$	3.24	3.70	4.15	3.30	3.71	4.12	3.29	3.71	4.12
$\gamma(\delta)_{\text{SMB}}$	-0.87	-0.17	0.55	-0.81	-0.16	0.47	-0.81	-0.16	0.48
$\gamma(\delta)_{\text{HML}}$	4.99	5.62	6.24	5.05	5.62	6.20	5.05	5.62	6.20
$\gamma(\delta)_S$	143	299	797	171	307	650	172	307	650

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 10000$. The true risk factors are synthetic mimics of the Mkt-Rf, SMB and HML factors in U.S. data. The table reports 10th percentiles, medians, and 90th percentiles of the distributions of the parameter estimates for Model 5, which includes the true risk factors and a spurious risk factor, S. Details of the Monte Carlo experiments are provided in the Appendix and main text.