NBER WORKING PAPER SERIES

IDENTIFICATION AND INFERENCE IN LINEAR STOCHASTIC DISCOUNT FACTOR MODELS

Craig Burnside

Working Paper 16634
http://www.nber.org/papers/w16634

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2010

This paper is a substantial revision of, and replaces, NBER Working Paper (#13357) "Empirical Asset Pricing and Statistical Power in the Presence of Weak Risk Factors." I am grateful to the National Science Foundation for financial support (SES-0516697). The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Identification and Inference in Linear Stochastic Discount Factor Models
Craig Burnside
NBER Working Paper No. 16634
December 2010, Revised October 2012
JEL No. C3,G12

ABSTRACT

When linear asset pricing models are estimated using excess return data, a normalization of the model must be selected. Several normalizations are equivalent when the model is correctly specified, but the identification conditions differ across normalizations. In practice, some or all of these identification conditions fail statistically when conventional consumption-based models are estimated, and inference is not robust across normalizations. Using asymptotic theory and Monte Carlo simulations, I present evidence that the lack of robustness in qualitative inference across normalizations can be attributed to model misspecification and lack of identification. I propose the use of tests for failure of the rank conditions. Using a calibrated model, I show that these tests are effective in detecting non-identified models.

Craig Burnside
Department of Economics
Duke University
213 Social Sciences Building
Durham, NC 27708-0097
and NBER
craig.burnside@duke.edu

Standard approaches to the estimation and evaluation of linear stochastic discount factor (SDF) models rely on the identifying assumption that the covariance matrix (or the raw cross-moment matrix) of a vector of asset returns with a vector of risk factors has full column rank. Using asymptotic theory, as well as small sample simulation-based evidence, I show that standard Generalized Method of Moments (GMM) estimates have non-standard properties when the rank conditions fail and the underlying model is nonetheless valid. When the rank conditions fail, and the underlying model is false, GMM can have very low power to reject the model; in fact, in some cases, a false model can appear to be remarkably successful in terms of fit and in terms of the statistical significance of parameter estimates.

I propose a simple diagnostic to test for identification, based on Cragg and Donald (1997) and Wright (2003). If there are k risk factors, and $n > k$ returns, the covariance matrix (or raw cross-moment matrix) must have rank k for the model to be identified. A simple test of the null hypothesis that the rank of the matrix is some $r < k$ can be implemented after estimating the elements of the matrix by GMM. In Monte Carlo experiments using calibrated models, I find that the null hypothesis of reduced rank is always rejected when the matrix whose rank is being tested corresponds to the true model and has full rank. The diagnostic test is also quite successful in identifying cases of reduced rank, although the size of the test exceeds its asymptotic size in repeated samples. My findings can be viewed as a formalization of the notion that researchers should test for significant spread among the factor betas (the slope coefficients in time series regressions of the returns on the risk factors).[1] In fact, a similar rank test can be based on the matrix of factor betas. For calibrated examples, I find that tests based on the covariance matrix have better size properties than tests based on the betas in small samples.

The issue of identification is not simply a theoretical curiosity, it is a practical reality. To see this, consider panel (A) of Table 1, which shows results of the rank tests for several conventional models in the literature. In all but the last two examples, I use the quarterly returns of the Fama-French 25 portfolios sorted on the basis of size and value to estimate and test the model. In the last two cases, sets of currency portfolios are used. When traditional financial factors are used (the CAPM and Fama-French models), the null of reduced rank is strongly rejected. Therefore, these models appear to be identified. On the other hand, the null of reduced rank is not rejected for most (though not all) of the models based on

[1] Kan and Zhang (1999b) make the same suggestion and point out that while Chen, Roll, and Ross (1986) and Ferson and Harvey (1993) performed such tests, relatively few researchers do so.

macroeconomic factors. This suggests that these models are poorly identified.

I consider the conventional case where the model SDF is estimated using excess returns. In this case, the mean of the SDF is unidentified but a subset of the model parameters can still be identified by adopting a normalization. Two candidate normalizations are described by Cochrane (2005). One uses raw risk factors, the other uses demeaned risk factors. The identification condition for the first normalization is that the cross-moment matrix of the returns and factors has full column rank. The identification condition for the second normalization is that the covariance matrix of the returns and factors has full column rank.

The two normalizations are theoretically equivalent when the model is true in the sense that the two representations of the SDF are proportional to one another. At first glance, this suggests that it does not matter which normalization is adopted by an applied researcher. In contrast, I show that empirical results are dramatically different for the two normalizations when conventional consumption-based models are estimated. In particular, the normalization that expresses the SDF in terms of raw risk factors tends to cast the models in more favorable light. Measures of fit are close to 1, and the estimated parameters of the SDF are highly statistically significant. For the normalization that uses demeaned risk factors, the models appear to fit the data very poorly, and estimated parameters have a much lower degree of statistical significance.

A combination of misspecification and lack of identification of the demeaned normalization can explain this non-robustness across the estimates. Why misspecification? For the consumption based models I consider, the rank tests suggest that the cross moment matrix has greater rank than the covariance matrix. As I show below, this is inconsistent with the model being true. How does lack of identification play a role? Using asymptotic theory I demonstrate that the normalization that uses raw risk factors has the following properties when its rank condition is satisfied, but the rank condition for the demeaned normalization is not: (i) the estimated parameters of the SDF converge in probability to a non-zero limit, (ii) an R^2 measure of model fit converges in probability to 1, (iii) the t-statistic associated with any parameter of the SDF that converges to a non-zero limit (of which there is at least one) will diverge in probability to $\pm\infty$ leading to the conclusion that the associated risk factor helps to price the assets, (iv) under some additional regularity assumptions the test of the over-identifying restrictions rejects the model with the same probability as the size of the test, that is, as if the model were true, yet (v) the estimated SDF is uncorrelated with

the returns. My simulation-based evidence indicates that these properties are borne out in finite samples.

The normalization that uses demeaned risk factors does not share these properties. When the rank condition fails the parameters of this normalization of the SDF are asymptotically unidentified. Under additional assumptions about the data generating processes, I show that the parameter estimates and test statistics have asymptotic distributions, but these results do not lead to general statements about the statistical significance of the parameter estimates and the likelihood of rejecting the model in large samples. Simulation-based evidence suggests that in finite samples a researcher is more likely to conclude that the model has poor fit, reject it based on the test of the pricing errors, and conclude that the spurious factor does not price the assets, when the SDF is expressed in terms of demeaned factors. Nonetheless, even with the demeaned normalization the power to reject the model can be low.

Although my results are specifically relevant for empirical work that uses GMM and one of the two normalizations, they have broader relevance, because the first stage of GMM with the demeaned normalization is equivalent to the two-pass regression method pioneered by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973), and widely used in the consumption-based asset pricing literature.

There is an extensive literature relating to this paper which examines the properties of asset pricing tests in the presence of *spurious* risk factors, that is risk factors that are uncorrelated with the returns. Kan and Zhang (1999a) examine the behavior of GMM estimators in the presence of spurious risk factors. In their setup, the estimated SDF nests the true model but includes an additional factor that is uncorrelated with the returns. Their results relate to mine for the demeaned specification, although they study risk factors that are mean zero by construction, and I consider cases where the estimated SDF may or may not nest the true model. Kan and Zhang (1999b) study similar issues in the context of the two-pass approach to model evaluation. A more closely related paper to this one is Kan and Robotti (2008). They examine the behavior of the Hansen and Jagannathan (1997) distance measure under the two model normalizations discussed here. They show that results can depend dramatically on the normalization chosen. Another related literature does not focus on issues of identification, but examines the effects of factor model misspecification on estimation and inference [for example, Hou and Kimmel (2006), Shanken and Zhou (2007),

Kan and Robotti (2009), and Kan, Robotti, and Jay Shanken (2012)]. Some of this literature proposes using misspecification-robust standard errors. These standard errors, however, still assume that the model is properly identified.

I consider a further case, that a risk factor, or linear combination of risk factors, has the same covariance with every asset return being explained. This situation arises, in practice, when factor betas are statistically significant, but have very little spread, as is the case for the quarterly Fama-French portfolios and US consumption growth.[2] A risk factor of this type may be relevant, but it clearly cannot explain any spread in the cross-sectional distribution of the expected returns being studied. I show that in sufficiently large samples both normalizations will lead to rejection of the over-identifying restrictions. However, in finite samples, if the covariance between the returns and the factor is small, the performance of the raw-factor normalization mimics the case where the factor is entirely spurious. Additionally, when there is no spread in the covariances associated with a factor, the identification condition for the two-pass regression method with a constant common pricing error fails, and thereby affects the reliability of inference for that procedure.

One way of dealing with the issue of model normalization would be to work with gross returns rather than excess returns. Alternatively one could include a moment condition associating the mean of the SDF with the mean of the price of a risk free asset. In either case there would be no need to choose a normalization of the SDF. This would not, however, deal with the problem of identification. If the covariance matrix of excess returns with the risk factors lacks full column rank, the rank condition for identification of the factor loadings still fails.

What should an applied researcher conclude from this paper? First, given that it is straightforward to do so, test for spread among the factor betas for each factor. Second, given that the issue of identification goes beyond a factor-by-factor look at the betas, test the rank conditions, and do so using the covariances, not the betas, since the covariance-based test seems to perform better in small samples. Looking at the rank condition directly helps to identify situations where each factor, by itself, is relevant, but some linear combination of the

[2] In related work, Lewellen and Nagel (2006) and Lewellen, Nagel and Shanken (2010) criticize empirical estimates of the conditional CAPM (and CCAPM) arguing that the spread in the betas in these models is too small for inference to be reliable. Daniel and Titman (2012) discuss the properties of the cross-sectional regressions when the test assets lie in a low-dimensional subspace of the full payoff space. Their analysis undoubtedly has bearing on GMM-based procedures as well, and one of the sets of data I examine in this paper (the Fama-French 25 returns, described below) is central to their discussion.

factors is redundant. Third, in situations where the identification appears to be weak, avoid the raw factor normalization. Fourth, even when adopting the demeaned normalization, be aware that weak identification affects inference about parameters and weakens tests of the pricing errors.

The paper is organized as follows. Section 1 lays out a standard linear factor model. Section 2 discusses the two normalizations of the SDF. Section 3 discusses the approach to be used in estimating the two model normalizations and testing for identification. Section 4 discusses empirical findings for simple consumption based models, using the Fama and French (1993) returns on 25 portfolios sorted by market capitalization and book-to-market value as test assets. It lays the foundation for the rest of the paper by showing that inference regarding consumption-based models is sharply dependent on the chosen normalization. Section 5 discusses the asymptotic properties of the estimates and diagnostic tests under the maintained assumption that the model is valid, and under the alternative assumption that it is misspecified. Section 6 performs a series of small-sample 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. Section 7 discusses some alternative approaches to addressing the identification problem. Section 8 concludes.

1 Model Setup and Moment Conditions

I consider the estimation of a class of linear factor pricing models where the SDF takes the form

$$m_t = a - f_t' b. \tag{1}$$

Here f_t is a $k \times 1$ vector of risk factors, a is a scalar constant and b is a $k \times 1$ vector of parameters. If m_t is the true SDF, then standard arguments imply that the price of any asset at time $t-1$ whose payoff at time t is x_t is $p_{t-1} = E_{t-1}(m_t x_t)$. Therefore the expected price of the asset is $E(p_{t-1}) = E(m_t x_t)$. It follows that the return to the asset, $R_t = x_t/p_{t-1}$, satisfies $E_{t-1}(m_t R_t) = 1$ and $E(m_t R_t) = 1$. Consequently, the difference between the returns on two assets $R_{1t} - R_{2t}$ satisfies $E[m_t(R_{1t} - R_{2t})] = 0$.

Now consider a particular $n \times 1$ vector, R_t^e, 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. We will assume, throughout, that $n \geq k$. If m_t given by (1) is the true SDF, then

$$E\left(R_t^e m_t\right) = 0. \tag{2}$$

Restriction (2) implies that

$$E(R_t^e) = -\frac{\text{cov}(R_t^e, m_t)}{E(m_t)}. \tag{3}$$

Equation (3) means that variation in $E(R_{it}^e)$ across i implies variation in $\text{cov}(R_{it}^e, m_t)$ across i.

A beta representation of the model is obtained by substituting (1) into the numerator of (3):

$$E(R_t^e) = \frac{\text{cov}(R_t^e, f_t')b}{E(m_t)} = \underbrace{\text{cov}(R_t^e, f_t')\,\text{var}(f_t)^{-1}}_{\beta}\underbrace{\frac{\text{var}(f_t)b}{E(m_t)}}_{\lambda}. \tag{4}$$

The term β is an $n \times k$ matrix of *factor betas*, while λ is a $k \times 1$ vector of *factor risk premia.* Row i in the matrix β is the vector of slope coefficients in a time series regression of R_{it}^e on f_t.

2 Normalizations of the SDF

The moment restriction (2) does not separately identify the parameters a and b. This is because the GMM errors, $R_t^e m_t$, for the parameter pair (a, b) are proportional to the GMM errors for the parameter pair $(\kappa a, \kappa b)$, for any scalar κ. I primarily consider two approaches to achieving identification, both of which involve picking a particular normalization of the SDF.

The first normalization rewrites (1) as

$$m_t = a m_t^* = a(1 - f_t' b^*) \tag{5}$$

where $b^* = b/a$. Equation (2) implies:

$$E(R_t^e m_t^*) = 0. \tag{6}$$

I refer to this as the *A-normalization* since it factors the parameter a out of the SDF.

Alternatively (1) can be rewritten as

$$m_t = \xi m_t^\diamond = \xi[1 - (f_t - \mu)' b^\diamond], \tag{7}$$

where μ is the unconditional mean of f_t, ξ is a scalar, $\xi = a - \mu' b$ and $b^\diamond = b/(a - \mu' b)$. Equation (2) implies:

$$E\left(f_t\right) = \mu \qquad E(R_t^e m_t^\diamond) = 0. \tag{8}$$

I refer to this as the *M-normalization* since it factors the mean of m_t out of the SDF.

Although b^* and $b^\diamond$ are different parameters (except when $\mu = 0$) the two normalizations are equivalent in the sense that the GMM errors $1 - f_t' b^*$ and $1 - \left(f_t - \mu\right)' b^\diamond$ are proportional to one another when evaluated at the true parameter values. Also, $b^\diamond$ can be obtained from b^* and μ: $b^\diamond = b^*/(1 - \mu' b^*)$.

I also consider a variant of the M-normalization which introduces a common pricing error, denoted α, that appears in every pricing equation. Under the null that the model is valid, $\alpha = 0$. The parameters, μ, $b^\diamond$ and α are estimated by exploiting the moment restrictions:

$$E\left(f_t\right) = \mu \qquad E(R_t^e m_t^\diamond - \alpha) = 0. \tag{9}$$

I refer to this variant as the *TP-normalization*. This normalization has been used in a GMM context by Parker and Julliard (2005) and is closely related to the two-pass regression method for estimating the beta-representation of the model. The latter approach, which is related to the methods introduced by Fama and MacBeth (1973), usually includes a constant equivalent to α in the second-pass regression.

3 Estimation and Inference using GMM

3.1 The A-Normalization

Using the n moment restrictions given by (6), b^* is estimated using GMM. Define $u_t^*(b^*) = R_t^e m_t^* = R_t^e(1 - f_t' b^*)$ and let $g_T^*(b^*) = \frac{1}{T}\sum_{t=1}^T u_t^* = \bar{R}^e - D_T b^*$ 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 $a_T^* g_T^* = 0$, where a_T^* is a $k \times n$ matrix and takes the form $a_T^* = D_T' W_T^*$, where W_T^* is an $n \times n$ positive definite weighting matrix. It follows that the GMM estimator of b^* is

$$\hat{b}^* = \left(D_T' W_T^* D_T\right)^{-1} D_T' W_T^* \bar{R}^e. \tag{10}$$

I consider multi-stage GMM estimators. In the first stage $W_T^* = I_n$. In the jth subsequent stage, $W_T^* = (S_T^*)^{-1}$ where $S_T^* = \frac{1}{T}\sum_{t=1}^T \hat{u}_t^* \hat{u}_t^{*\prime}$, $\hat{u}_t^* = R_t^e(1 - f_t' \hat{b}_{j-1}^*)$ and $\hat{b}_j^*$ represents the jth-stage estimator of b^*.

Let $\delta_T^* = -D_T$. A test of the pricing errors is based on the statistic

$$J^* = T g_T^*(\hat{b}^*)'(\hat{V}_g^*)^+ g_T^*(\hat{b}^*), \tag{11}$$

where

$$\hat{V}_g^* = A_T^* S_T^* A_T^{*\prime} \text{ with } A_T^* = I_n - \delta_T^*(a_T^*\delta_T^*)^{-1}a_T^*, \tag{12}$$

and X^+ indicates the generalized inverse of the matrix X.

Equations (5) and (6) imply that

$$E(R_t^e) = E\left(R_t^e f_t'\right) b^*. \tag{13}$$

In a finite sample, corresponding to the left-hand side of (13) is the vector of *realized expected returns*, $\bar{R}^e$. Corresponding to the right-hand side of (13) is a vector of *predicted expected returns* given by $D_T\hat{b}^*$. The model's fit is evaluated using the cross-sectional R^2:

$$R_*^2 = 1 - \frac{(\bar{R}^e - D_T\hat{b}^*)'(\bar{R}^e - D_T\hat{b}^*)}{(\bar{R}^e - \ddot{R}^e)'(\bar{R}^e - \ddot{R}^e)}, \tag{14}$$

where $\ddot{R}^e = \frac{1}{n}\sum_{i=1}^n \bar{R}_i^e$ is the cross-sectional average of the realized expected returns.

3.2 The M-Normalization

Using the $n+k$ moment restrictions given by (8), $b^\diamond$ and μ are estimated using GMM. Define $u_{1t}^\diamond(b^\diamond,\mu) = R_t^e m_t^\diamond = R_t^e[1-(f_t-\mu)'b^\diamond]$ and let $g_{1T}^\diamond(b^\diamond,\mu) = \frac{1}{T}\sum_{t=1}^T u_{1t}^\diamond = \bar{R}^e - \left(D_T - \bar{R}^e\mu'\right)b^\diamond$. Define $u_{2t}^\diamond(\mu) = f_t - \mu$ and let $g_{2T}^\diamond(\mu) = \frac{1}{T}\sum_{t=1}^T u_{2t}^\diamond = \bar{f} - \mu$, where $\bar{f} = \frac{1}{T}\sum_{t=1}^T f_t$. Define $u_t^\diamond = \begin{pmatrix} u_{1t}^{\diamond\prime} & u_{2t}^{\diamond\prime}\end{pmatrix}'$ and $g_T^\diamond = \begin{pmatrix} g_{1T}^{\diamond\prime} & g_{2T}^{\diamond\prime}\end{pmatrix}'$. I consider GMM estimators that set $a_T^\diamond g_T^\diamond = 0$, where $a_T^\diamond$ is a $2k \times (n+k)$ matrix and takes the form

$$a_T^\diamond = \begin{pmatrix} d_T' W_T^\diamond & 0 \\ 0 & I_k \end{pmatrix}, \tag{15}$$

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

$$\hat{b}^\diamond = \left(d_T' W_T^\diamond d_T\right)^{-1} d_T' W_T^\diamond \bar{R}^e \tag{16}$$

$$\hat{\mu} = \bar{f}. \tag{17}$$

I consider multi-stage GMM estimators. In the first stage $W_T^\diamond = I_n$. In the jth subsequent stage, $W_T^\diamond = \left(P_T S_T^\diamond P_T'\right)^{-1}$ where $P_T = \begin{pmatrix} I_n & \bar{R}^e(\hat{b}_{j-1}^\diamond)'\end{pmatrix}$, $\hat{b}_j^*$ represents the jth-stage estimator

of $b^\diamond$ and $S_T^\diamond$ is a consistent estimator of $S_0^\diamond = \sum_{j=-\infty}^{+\infty} E(u_t^\diamond u_{t-j}^{\diamond\prime})$.[3] Because $u_{2t}^\diamond$ may be serially correlated I use a VARHAC estimator, described in more detail in Appendix A, to compute $S_T^\diamond$.

Let

$$\delta_T^\diamond = \begin{pmatrix} -d_T & \bar{R}^e \hat{b}^{\diamond\prime} \\ 0 & -I_k \end{pmatrix}. \tag{18}$$

A test of the pricing errors is based on

$$J^\diamond = T g_T(\hat{b}^\diamond, \hat{\mu})' (\hat{V}_g^\diamond)^+ g_T(\hat{b}^\diamond, \hat{\mu}), \tag{19}$$

where

$$\hat{V}_g^\diamond = A_T^\diamond S_T^\diamond A_T^{\diamond\prime} \text{ with } A_T^\diamond = I_{n+k} - \delta_T^\diamond \left(a_T^\diamond \delta_T^\diamond\right)^{-1} a_T^\diamond. \tag{20}$$

Equations (7) and (8) imply that

$$E(R_t^e) = E\left[R_t^e \left(f_t - \mu\right)'\right] b^\diamond. \tag{21}$$

Corresponding to the right-hand side of (21) is a vector of predicted expected returns, $d_T \hat{b}^\diamond$. The cross-sectional R^2 measure is:

$$R_\diamond^2 = 1 - \frac{(\bar{R}^e - d_T \hat{b}^\diamond)'(\bar{R}^e - d_T \hat{b}^\diamond)}{(\bar{R}^e - \ddot{R}^e)'(\bar{R}^e - \ddot{R}^e)}. \tag{22}$$

Yogo (2006) uses a different, optimal, GMM procedure in conjunction with the M-normalization. Noting that the derivative of $g_{1T}^\diamond$ with respect to μ is non-zero, he uses a variant of $a_T^\diamond$ that is not block diagonal because this improves asymptotic efficiency. In his case $\hat{\mu}$ 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. I discuss his procedure in more detail in Appendix C.

3.3 The TP-Normalization

When estimating the TP-normalization, I set up the first stage of GMM so that the point estimates are consistent with the two-pass regression method in the following sense: if the factor risk premia are evaluated as $\hat{\lambda}_f^\diamond = \mathrm{S}_f \hat{b}^\diamond$, then $\hat{\alpha}$ and $\hat{\lambda}_f^\diamond$ are numerically identical to

[3] The first stage 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 stages, Cochrane (2005) suggests using the matrix $\begin{pmatrix} I_n & 0_{n\times k} \end{pmatrix}$ in place of P_T in the expression for $W_T^\diamond$. This is less efficient in terms of the covariance matrix of $\hat{b}^\diamond$, but is asymptotically equivalent in terms of the test of the overidentifying restrictions.

the coefficients obtained in the second-pass regression. To avoid repetitive descriptions in the main text, I relegate the formulaic details of the estimation of the TP-normalization to Appendix A.

3.4 Testing Identification

3.4.1 Rank Conditions for Identification

Each normalization is associated with a rank condition that must hold for the model to be identified. I defer asymptotic theory to Section 5, but it is useful, at this stage, to outline these conditions. They are:

- A-normalization: For the parameter b^* to be identified it is necessary that the $n \times k$ matrix $D_0 \equiv E(R_t^e f_t')$ have rank k.
- M-normalization: For the parameter $b^\diamond$ to be identified it is necessary that the $n \times k$ matrix $d_0 \equiv \text{cov}(R_t^e, f_t)$ have rank k.
- TP-normalization: For the parameters α and $b^\diamond$ to be identified it is necessary that the $n \times (k+1)$ matrix $d_0^+ = (\ \iota \quad d_0\ )$, where ι is an $n \times 1$ vector of ones, have rank $k+1$.

3.4.2 The Distinction Between Model Validity and Identification

Failure of one or more of the rank conditions does not necessarily imply that the model of the SDF is false. Consider the following examples.

Imagine a single factor model. Suppose that $\text{rank}(d_0) = 0$, so that the rank condition for the M-normalization fails. This implies that $d_0 = 0$. The asset pricing model may still be valid, but only if $E(R_t^e) = d_0 b^\diamond = 0$ for the particular assets being studied. This, in turn, implies that $D_0 = 0$, and that $d_0^+ = (\ \iota \quad 0\ )$, so that the rank conditions for the A and TP-normalizations also fail.

To take another single factor example, suppose that $d_0 = \iota c$ for some scalar $c \neq 0$. This implies that $\text{rank}(d_0) = 1$, so that the rank condition of the M-normalization is satisfied. The TP-normalization, however, is unidentified because the matrix $d_0^+ = (\ \iota \quad \iota c\ )$ has rank 1. The asset pricing model is still valid as long as every element in the vector $E(R_t^e)$ is equal to $cb^\diamond$.

Now imagine a model with $k > 1$ factors, for which $\text{rank}(d_0) = r < k$. So the rank condition for the M-normalization fails. This implies that a lower dimensional model can

correctly price the returns. To see this let $X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}$ where X_1 is a $k \times r$ matrix whose columns span the rowspace of d_0, denoted $\mathcal{R}(d_0)$, and X_2 is a $k \times (k-r)$ matrix whose columns span the nullspace of d_0, denoted $\mathcal{N}(d_0)$. The columns of X span R^k, by construction. Let the $r \times 1$ vector $\tilde{b}^\diamond$ be the vector containing the first r elements of $X^{-1}b^\diamond$, and define the $n \times r$ matrix $\tilde{d}_0 = d_0 X_1$. Because $d_0 X_2 = 0$ we can write $E(R_t^e) = d_0 b^\diamond = d_0 X X^{-1} b^\diamond = \tilde{d}_0 \tilde{b}^\diamond$. The original model may well be valid, but only the lower dimensional model is identified for the particular assets whose excess returns are included in the vector R_t^e.

The model being true does, however, imply an important restriction across the rank conditions for the A and M-normalizations. When the model is true, $\mu_R = d_0 b_0^\diamond$ and, therefore, $D_0 = d_0(I_k + b_0^\diamond \mu_0')$. Hence $\text{rank}(D_0) \leq \text{rank}(d_0)$ when the model is true. A natural specification test for any model is to check whether this inequality holds in the data.

3.4.3 Testing the Rank Conditions

To test whether the rank conditions hold I borrow directly from Cragg and Donald (1997) and Wright (2003). Let B be an $n \times k$ matrix. Let $\hat{B}$ be a consistent estimator for B and assume that $\sqrt{T}\,\text{vec}(\hat{B} - B_0) \xrightarrow{d} N(0, V_B)$, 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 $\text{rank}(B) = r < k$ I form the statistic

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

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 A and M-normalizations by letting B be d_0 or D_0 and estimating the elements of these matrices by GMM. When $r = 0$ the test associated with d_0 is analogous to an F-test for $\text{cov}(R_{it}^e, f_{jt}) = 0$ for all i, j.

Analogous tests can be constructed for the rank condition of the TP-normalization. To test the null hypothesis that $\text{rank}\begin{pmatrix} \iota & B \end{pmatrix} = r$ I form the statistic

$$L^+(r) = \min_{P \in \Omega_r} T\,\text{vec}[\begin{pmatrix} \iota & \hat{B} \end{pmatrix} - P]' \begin{pmatrix} 0 & 0 \\ 0 & V(\hat{B}) \end{pmatrix}^+ \text{vec}[\begin{pmatrix} \iota & \hat{B} \end{pmatrix} - P].$$

If the true rank of $\begin{pmatrix} \iota & B_0 \end{pmatrix}$ is r, $L^+(r) \xrightarrow{d} \chi^2_{(n-r)(k+1-r)}$. When $r = 1$ the test associated with d_0^+ is analogous to an F-test for $\text{cov}(R_{it}^e, f_{jt}) = \text{cov}(R_{\ell t}^e, f_{jt})$ for all i, ℓ and every j.

One aspect of these tests that may seem disadvantageous is that the asymptotic distribution of $L(r)$ is derived under the null that B has reduced rank ($r < k$). This may trouble researchers who would prefer to have full rank as the null hypothesis. Two considerations

are important. First, the null hypothesis of reduced rank is not equivalent to the asset pricing model being false. Second, the rank tests are very powerful, asymptotically, against the alternative hypothesis that the matrix B has full rank; the test rejects the null with probability one in this case. I investigate the finite sample size and power of the $L(r)$ and $L^+(r)$ tests in Section 6.

4 Empirical Findings

I consider a set of data widely used in the asset pricing literature: the 25 portfolios of US stocks sorted on size and the book-to-market value ratio introduced by Fama and French (1993) and henceforth referred to as the FF25.[4] Using the real excess returns to the these portfolios at the quarterly frequency over the period 1949Q1–2008Q4, I estimate two consumption-based factor models.

- The CCAPM, which uses a single factor: the log-growth rate of real per capita consumption of nondurables and services (referred to from now on as "consumption growth").
- A Durables CCAPM, which uses two factors: consumption growth and durables growth (defined as the log-growth rate of the real stock of consumer durables).[5]

I present empirical results from the first, second and fifth stages of GMM. In practice, I found that for most of the GMM procedures and models I considered further iterations of the weighting matrix produced only minor changes in the results.

4.1 The CCAPM

The first row of Table 2 presents results for the CCAPM using the A-normalization. The GMM estimates of b^* are statistically significant at all GMM stages. The fit of the model is quite good at the first stage of GMM as measured by the cross-sectional R^2, but deteriorates with further GMM iterations. The model is not rejected at conventional significance levels according to the J-statistic.

[4] The data are described in more detail in Appendix B.

[5] Yogo (2006) estimates an extension of the durables C-CAPM that adopts recursive preferences. He estimates a linear approximation of the model that adds a third factor: the market return, defined as the return of a value-weighted portfolio of all US stocks.

These findings are not robust when we turn to the M-normalization in the second row of Table 2. While the GMM estimates of $b^\diamond$ are statistically significant at all GMM stages, the fit of the model is very poor as measured by the cross-sectional R^2. The model is also rejected on the basis of the J-statistic.

Results for the TP-normalization are shown in the last row of Table 2. The estimates of $\hat{b}^\diamond$ remain positive but are no longer statistically significant. Compared to the M-normalization the fit of the model improves a little due to the inclusion of the constant. But the model is still rejected on the basis of the J-statistic.

Is the CCAPM identified? The L statistic for $\text{rank}(D_0) = 0$ has a tiny p-value (see Table 1, panel B) suggesting that D_0 has full rank and that the A-normalization is identified. The results are slightly less clear when we turn to the M-normalization: the L statistic for $\text{rank}(d_0) = 0$ has a p-value of 0.036. For the TP-normalization, the L^+ statistic for $\text{rank}(d_0^+) = 1$ has a p-value of 0.24, suggesting that even if d_0 has full rank (1), one cannot reject that d_0^+ has less than full rank. This circumstance would arise if d_0 was a constant vector. The evidence about identification is mixed. Below I ask whether identification problems explain the differing performance of the CCAPM across the A, M and TP-normalizations.

4.2 The Durables CCAPM

Next, consider the model that uses consumption growth and durables growth as risk factors. The first row of Table 3 presents results using the A-normalization. The estimates of b^* for both factors are positive and significant (except for consumption growth in the first stage of GMM). The R^2 measures suggest that the model can explain about 98 percent of the cross-sectional variation in expected returns. The model also easily passes the test of the pricing errors.

With the M-normalization (second row of Table 3) the results are dramatically different. The estimates of $b^\diamond$ for durables growth remain positive but are no longer statistically significant. Strikingly, the fit of the model is very poor, the R^2 being uniformly negative. The model is not rejected at the second stage of GMM, but is after a few iterations on the weighting matrix.

Finally, the TP-normalization (third row of Table 3) also casts doubt on the model. None of the estimates of $b^\diamond$ are statistically significant, the estimate of α is always large and at least marginally statistically significant, the R^2 of the model hovers around zero, and with enough

iterations on the weighting matrix the model is rejected on the basis of the J-statistic.

Tests of the rank conditions are provided in Table 1. The L statistic for $\text{rank}(D_0) = 1$ has a p-value of 0.014, suggesting that D_0 has full rank (2) and that the A-normalization is identified. The L statistic for $\text{rank}(d_0) = 1$, on the other hand, has a p-value of 0.977. This indicates that we cannot reject that d_0 has reduced rank, and that the M-normalization is not identified. The L^+ statistic for $\text{rank}(d_0^+) = 2$ has a p-value of 0.968, so we cannot reject that d_0^+ has reduced rank, and that the TP-normalization is not identified.

In summary, the Durables CCAPM seems to be poorly identified. This poor identification stems from the weak correlation between the consumption factors and the asset returns. Let f_{1t} denote consumption growth and f_{2t} denote durables growth. Recall that the rank test on d_0 for the CCAPM is equivalent to an F-test for $\text{cov}(R_t^e, f_{1t}) = 0$. This test had a p-value of 0.036. Similarly, the rank test on d_0^+ for the CCAPM is equivalent to an F-test for $\text{cov}(R_t^e, f_{1t}) = c\iota$. This test had a p-value of 0.24. This suggests that, at best, consumption growth has a common covariance with all of the returns. A test for $\text{cov}(R_t^e, f_{2t}) = 0$ has a p-value of 0.974. This suggest that durables growth is uncorrelated with the returns. Indeed, if it were the case that $\text{cov}(R_t^e, f_{1t}) = c\iota$ and $\text{cov}(R_t^e, f_{2t}) = 0$ then d_0 would have rank 1 and d_0^+ would have rank 1, consistent with the findings of the rank tests.

The fact that D_0 appears to have rank 2, while d_0 has rank 1, also suggests that the model is misspecified. As mentioned above, the model cannot be true when $\text{rank}\,(D_0) > \text{rank}(d_0)$.

Poor identification in combination with model misspecification is a plausible explanation of the lack of robustness we observe across normalizations for consumption-based models. As we will see in the next section, when the model is false, D_0 has full column rank and d_0 has less than full rank, asymptotic theory predicts exactly the lack of robustness exhibited by estimates of these models for the different normalizations.

In Burnside (2007) I present results for a wider set of the factor models, and an even larger set of results, for additional models, is available upon request. These results confirm that lack of robustness across normalizations is a common occurrence when the rank conditions fail for the M and TP-normalizations, especially if $\text{rank}\,(D_0) > \text{rank}(d_0)$.

5 Large Sample Properties of the GMM Estimators

In this section I explore the asymptotic properties of the GMM procedures described in Section 3.

Assumption 1 Let the true SDF be given by (1), the true values of the parameters a and b be denoted a_0 and b_0, and the true values of b^* and $b^\diamond$ be denoted $b_0^* = b_0/a_0$, and $b_0^\diamond = b_0/(a_0 - \mu_0' b_0)$.

Assumption 2 Let D_0 and d_0 have full column rank.

Assumption 3 Assume that $D_T \xrightarrow{a.s.} D_0$, $d_T \xrightarrow{a.s.} d_0$, $W_T^* \xrightarrow{a.s.} W_0^*$ and $W_T^\diamond \xrightarrow{a.s.} W_0^\diamond$, with W_0^* and $W_0^\diamond$ positive definite.

For compactness of notation let $\theta^\diamond = (b^\diamond, \mu)$ and $\theta_0^\diamond = (b_0^\diamond, \mu_0)$. Define

$$\delta_0^* = E\left[\frac{\partial u_t^*(b_0^*)}{\partial b^*}\right] = -D_0$$

and

$$\delta_0^\diamond = E\left[\frac{\partial u_t^\diamond(\theta_0^\diamond)}{\partial \theta^\diamond}\right] = \begin{pmatrix} -d_0 & E(R_t^e) b_0^{\diamond\prime} \\ 0 & -I_k \end{pmatrix}$$

Let $S_0^* = E\left[\sum_{j=-\infty}^{+\infty} u_t^*(b_0^*) u_{t-j}^*(b_0^*)'\right]$ and $S_0^\diamond = E\left[\sum_{j=-\infty}^{+\infty} u_t^\diamond(\theta_0^\diamond) u_{t-j}^\diamond(\theta_0^\diamond)'\right]$. Define $a_0^* = D_0' W_0^*$ and

$$a_0^\diamond = \begin{pmatrix} d_0' W_0^\diamond & 0 \\ 0 & I_k \end{pmatrix}.$$

Theorem 1 *Suppose assumptions 1—3 are satisfied. Under additional regularity conditions provided in Hansen (1982)* $\hat{b}^* \xrightarrow{a.s.} b_0^*$, $\hat{\theta}_0^\diamond \xrightarrow{a.s.} \hat{\theta}_0^\diamond$, $\sqrt{T}(\hat{b}^* - b_0^*) \xrightarrow{d} N(0, V_b^*)$ *and* $\sqrt{T}(\hat{\theta}^\diamond - \theta_0^\diamond) \xrightarrow{d} N(0, V_\theta^\diamond)$ *with* $V_b^* = (a_0^* \delta_0^*)^{-1} a_0^* S_0^* a_0^{*\prime} (\delta_0^{*\prime} a_0^*)^{-1}$ *and* $V_\theta^\diamond = (a_0^\diamond \delta_0^\diamond)^{-1} a_0^\diamond S_0^\diamond a_0^{\diamond\prime} (\delta_0^{\diamond\prime} a_0^\diamond)^{-1}$. *Also,* $R_*^2 \xrightarrow{a.s.} 1$ *and* $R_\diamond^2 \xrightarrow{a.s.} 1$. *The statistics* J^* and $J^\diamond$ *both converge in distribution to* χ^2 *random variables with* $n - k$ *degrees of freedom.*

The proof of Theorem 1 is provided in Appendix A. The interpretation of Theorem 1 is that when the model is true (Assumption 1, model validity), and when the moment conditions are informative about the risk factors (Assumption 2, identification), both approaches to estimation work well.

Next I turn to a situation where the model remains true, but the returns in the data being studied do not fully shed light on the relevance of the risk factors.

Assumption 2a Let $\text{rank}(d_0) = r < k$.

If we maintain Assumption 1, that the model is true, then, as shown above, under Assumption 2a D_0 has rank less than or equal to r.

Theorem 2 *Suppose assumptions 1, 2a and 3 are satisfied. It follows that neither b^* nor $b^\diamond$ is asymptotically identified. Nonetheless, $R^2_* \xrightarrow{a.s.} 1$ and $R^2_\diamond \xrightarrow{a.s.} 1$.*

The proof of Theorem 2 is provided in Appendix A. The interpretation of Theorem 2 is that even though the model is true (Assumption 1), the moment conditions are insufficiently informative about the risk factors (Assumption 2a), and neither approach to estimation will work well in large samples. The lack of asymptotic identification might be manifested in large standard errors for $\hat{b}^*$ and $\hat{b}^\diamond$ in finite samples. While the parameters of the SDF are not identified, this is because there are multiple values of b^* and $b^\diamond$ that satisfy the moment conditions asymptotically. As a result the measures of fit limit to 1. This suggests that the models being true, but lack of identification being pervasive, is not responsible for the lack of robustness we saw in Section 4. There, we found that for some normalizations R^2 was very close to 1, whereas, for other R^2 was very far from 1.

Next I turn to a situation where the model is false.

Assumption 1b Assume that the true SDF is not given by (1) so that, in general, $\mu_R \neq d_0 b^\diamond$ for the particular vector f_t being studied.

Assumption 2b Let $\text{rank}(D_0) = k$ but $\text{rank}(d_0) = k - 1$.

We can make Assumption 2b because we are no longer maintaining Assumption 1.

Theorem 3 *Suppose assumptions 1b, 2b and 3 are satisfied. Under additional regularity conditions provided in Hansen (1982) $\hat{b}^* \xrightarrow{a.s.} b^*_s = x/(x'\mu_0)$, where x is the unique element of the nullspace of d_0 whose elements sum to 1, and $R^2_* \xrightarrow{a.s.} 1$. At least one element of b^*_s is non-zero. In contrast $b^\diamond$ is not asymptotically identified.*

The proof of Theorem 3 is provided in Appendix A. Since the model is not true, the parameter vector b^*_s has no interpretation as the "true" value of b^*. Rather b^*_s is a degenerate value of b^* for which the moment condition (5) holds, even though (1) is not the true SDF. The sum of the elements of b^*_s is $1/(x'\mu_0)$ which is the inverse of a weighted average of the means of the risk factors.

Before turning to the asymptotic distributions of $\hat{b}^*$ and J^* it is helpful to define the following notation. Let $S^*_s = E\sum_{j=-\infty}^{+\infty} u^*_t(b^*_s)u^*_{t-j}(b^*_s)'$ and $V^*_s = E[u^*_t(b^*_s)u^*_t(b^*_s)']$. In general

$S_s^* \neq V_s^*$ under Assumption 1b. Let $A_0^* = I_n - \delta_0^*(a_0^*\delta_0^*)^{-1}a_0^*$ and define $V_g^* = A_0^* V_s^* A_0^{*\prime}$ and $V_0^* = A_0^* S_s^* A_0^{*\prime}$. Finally, diagonalize V_0^* as $V_0^* = P_0\Lambda_0 P_0'$ where the columns of P_0 are the orthonormal eigenvectors of V_0^* and Λ_0 is a diagonal matrix with the eigenvalues of V_0^* on the diagonal. Let $\tilde{V}_0^* = P_0\Lambda_0^{1/2}$ so that $\tilde{V}_0^*\tilde{V}_0^{*\prime} = V_0^*$.

Theorem 4 *Under the assumptions of Theorem 3,* $\sqrt{T}(\hat{b}^* - b_s^*) \xrightarrow{d} N(0, V_b^*)$ *with* $V_b^* = (a_0^*\delta_0^*)^{-1}a_0^* S_s^* a_0^{*\prime}(\delta_0^{*\prime}a_0^*)^{-1}$. *If* $S_s^* = V_s^*$, $J^* \xrightarrow{d} \chi^2_{n-k}$. *When* $S_s^* \neq V_s^*$, $J^* \xrightarrow{d} \sum_{i=1}^{n-k}\lambda_i z_i^2$ *where* z_1, z_2, ..., z_{n-k} *are mutually independent standard normal random variables and* λ_1, λ_2, ..., λ_{n-k} *are the non-zero eigenvalues of the matrix* $\tilde{V}_0^{*\prime}(V_g^*)^{+}\tilde{V}_0^*$.

The proof of Theorem 4 is provided in Appendix A. Some examples are helpful in interpreting Theorems 3 and 4.

5.1 Single Factor Model With a Spurious Factor

For a single factor model, the assumptions of Theorem 3 imply that $d_0 = 0$ and, since D_0 has rank 1, that $D_0 = \mu_R\mu_0 \neq 0$. It follows that $b_s^* = 1/\mu_0$. A researcher testing the model under the null would compute an inconsistent estimate of V_b^* that would converge, instead, to $(a_0^*\delta_0^*)^{-1}a_0^* V_s^* a_0^{*\prime}(\delta_0^{*\prime}a_0^*)^{-1}$. Nonetheless, because this matrix is finite, the t-statistic for $\hat{b}^*$ would diverge to $+\infty$ if $\mu_0 > 0$ or $-\infty$ if $\mu_0 < 0$. The predicted expected returns, $D_T\hat{b}^*$, would converge almost surely to $D_0 b_s^* = \mu_R$. Therefore $R_*^2 \xrightarrow{a.s.} 1$.

Thus, in large samples, a researcher testing the model using the A-normalization would conclude that the factor f_t prices the returns (due to the statistical significance of $\hat{b}^*$) and that the model's fit is perfect. A researcher testing the over-identifying restrictions using, say, a 5 percent critical value from the χ^2_{n-1} distribution would only reject the model about 5 percent of the time in repeated large samples if $S_s^* = V_s^*$. It is unclear what would happen in the more general case when $S_s^* \neq V_s^*$.

For the M-normalization, Hansen's identification condition requires that there should be a unique $b^\diamond$ that solves $(d_0' W_0^\diamond d_0)\, b^\diamond = d_0' W_0^\diamond \mu_R$. The reason $b^\diamond$ is unidentified is that when $d_0 = 0$ any value of $b^\diamond$ is a solution.

5.2 Multi-factor Model with a Single Spurious Factor

Without loss of generality, let the kth factor be spurious: $\text{cov}(R_t^e, f_{kt}) = 0$ and let the rank of d_0 be $k-1$ as in the assumptions of Theorem 3. In this case, the vector x referred to in

the statement of Theorem 3 has a 1 as its kth element and zeros elsewhere. So $b_s^* = x/(x'\mu_0)$ has $1/E(f_{kt})$ as its kth element and zeros elsewhere. This means, oddly enough, that a researcher testing the model using the A-normalization and a large sample of data would conclude that the factor f_{kt} prices the returns and that the other factors are irrelevant, even though they are the only ones correlated with R_t^e. As in the previous example, $R_*^2 \xrightarrow{a.s.} 1$. A researcher testing the over-identifying restrictions using a 5 percent critical value from the χ^2_{n-k} distribution would only reject the model about 5 percent of the time in repeated large samples if $S_s^* = V_s^*$.

As in the previous example, for the M-normalization, the identification condition is not satisfied because there is no unique $b^\diamond$ that solves $(d_0' W_0^\diamond d_0)\, b^\diamond = d_0' W_0^\diamond \mu_R$. When the last column is d_0 is zero, but the rest of the matrix has full column rank, it is $b_k^\diamond$ that is unidentified asymptotically. The rest of the parameter vector $b^\diamond$ is identified.

To some extent, this case resembles the Durables CCAPM model in that durables growth appears to be spurious, while consumption does not. A test of $\text{cov}(R_t^e, f_t) = 0$ for consumption growth alone has a p-value of 0.036, whereas a test of $\text{cov}(R_t^e, f_t) = 0$ for durables growth alone has a p-value of 0.974. Consistent with what the theoretical results predict, as Table 3 shows, the A-normalization delivers a highly significant estimate of $\hat{b}^*$ for durables growth (the t-statistic is about 8 in the second and later stages of GMM). The R^2 for the Durables CCAPM is 0.98 at every GMM stage, and the model is far from being rejected on the basis of the over-identifying restrictions.

In contrast, for the M-normalization, at most stages of GMM, the estimate of $b^\diamond$ is significant for consumption growth, and not for durables growth. It is important to note, however, that lack of identification of the M-normalization means that we cannot trust conventional inference. This is discussed more below.

5.3 Multi-factor Model with Colinear Covariances

Now consider the more general case where $k > 1$, where d_0 has no zero columns, but d_0 has rank $k-1$. In this case, no one factor is spurious, but there is an identification problem for the M-normalization due to multicollinearity. Using the A-normalization, however, $\hat{b}^* \xrightarrow{a.s.} b_s^* = x/(x'\mu_0)$ and $R_*^2 \xrightarrow{a.s.} 1$. Remarkably, the estimated model puts all its weight on an irrelevant linear combination of the risk factors because

$$\text{cov}(R_t^e, m_t^*) = -\text{cov}(R_t^e, f_t' b_s^*) = -\frac{d_0 x}{x'\mu_0} = 0.$$

5.4 Greater Numbers of Spurious Factors

Theorems 3 and 4 rely on the assumption that only one linear combination of the risk factors is spurious, in the sense that $\text{rank}(d_0) = k - 1$. If $\text{rank}(d_0) = r < k - 1$, then b^* is asymptotically unidentified because there will be many b^* such that $a_0^*(\mu_R - D_0 b^*) = 0$. All these b^*, however, share the property that $\mu_0' b^* = 1$ and $\mu_R = D_0 b^*$. So, although the individual elements of $\hat{b}^*$ are unidentified, the particular linear combination $\mu_0' b^*$ is identified. Presumably, then $\hat{\mu}'\hat{b}^*$ would be centered around 1 and the measure of fit, R_*^2, would be roughly 1.

5.5 Asymptotic Properties of the M-Normalization

To this point we have only been able to show lack of asymptotic identification of the M-normalization when the rank condition on d_0 fails. Although $b^\diamond$ is unidentified asymptotically, and does not have a well defined probability limit, it is still possible, under additional regularity conditions, to derive its asymptotic distribution, the distribution of the associated t-statistic, the model's R^2 and the J-statistic used to test the over-identifying restrictions. Here I consider only the case of a single factor model with the returns and factors being iid and independent of each other. The more general case of a multifactor model with persistent returns or factors can be worked out, but at the cost of algebraic complexity.

Theorem 5 *Make the assumptions of Theorem 3. Let R_t^e and f_t be iid over time and mutually independent, and let $k = 1$. Let $\Sigma_R = \text{var}(R_t^e)$, and $\sigma_f^2 = \text{var}(f_t)$. Define the random variables $X \sim N(0, \sigma_f^2 \Sigma_R)$, $Z = (X'\mu_R)/(X'X)$ and $\tilde{Z} = (X'\Sigma_R^{-1}\mu_R)/(X'\Sigma_R^{-1}X)$. Then $T^{1/2} d_T \xrightarrow{d} X$. It follows that at the first stage of GMM $T^{-1/2}\hat{b}^\diamond \xrightarrow{d} Z$, $t(\hat{b}^\diamond) \xrightarrow{d} (X'\mu_R)/[\sigma_f^2 Z^2 (X'\Sigma_R X)]^{1/2}$ and $R_\diamond^2 \xrightarrow{d} 1 - (\mu_R' M \mu_R)/(\mu_R' M_\iota \mu_R)$, where $M = I_n - X(X'X)^{-1}X'$, $M_\iota = I_n - \iota\iota'/n$ and ι is an $n \times 1$ vector of ones. At the second stage of GMM $T^{-1/2}\hat{b}^\diamond \xrightarrow{d} \tilde{Z}$, $t(\hat{b}^\diamond) \xrightarrow{d} (X'\Sigma_R^{-1}\mu_R)/[\sigma_f^2 Z^2 (X'\Sigma_R^{-1}X)]^{1/2}$ and $R_\diamond^2 \xrightarrow{d} 1 - (\mu_R' \tilde{M}' \tilde{M} \mu_R)/(\mu_R' M_\iota \mu_R)$, where $\tilde{M} = I_n - X(X'\Sigma_R^{-1}X)^{-1}X'\Sigma_R^{-1}$, and $J \xrightarrow{d} (\mu_R' \tilde{M}' \Sigma_R^{-1} \tilde{M} \mu_R')/(\sigma_f^2 Z^2)$.*

The proof of Theorem 5 is provided in Appendix A.[6] Interpreting Theorem 5 in general terms is difficult because of the dependence of the asymptotic distributions on μ_R and Σ_R. The distribution of $\hat{b}^\diamond$ will spread out as the sample size increases at a rate of $T^{1/2}$. The t-statistic

[6] Similar asymptotic properties are derived in Kan and Zhang (1999a, 1999b) for the case where f_t is zero mean or, equivalently, has a known mean.

for $\hat{b}^{\diamond}$ has a well-defined asymptotic distribution at both stages of GMM, so the probability of finding $\hat{b}^{\diamond}$ to be significantly different from zero will converge to a number generally not equal to zero. Similarly, the J statistic used to test the over-identifying restrictions has a well-defined asymptotic distribution, so the probability of rejecting the model will converge to a number generally not equal to zero nor one.

5.6 Approximate Failure of the Rank Condition

The assumption that the rank condition literally does not hold may seem extreme. In single factor models it requires that d_0 is exactly zero. A standard device in the theory of weak instruments and unit root testing is also useful when it is preferable to assume that d_0 is small and asymptotically vanishing but not literally zero. Mimicking Hall's (2005) discussion of Staiger and Stock (1997), in the single factor case we might suppose that $R_t^e = \mu_R + c_T(f_t - \mu) + u_t$ where u_t is an $n \times 1$ vector that is uncorrelated with f_t and $c_T = T^{-1/2}c$ where c is an $n \times 1$ vector of constants. Working with this alternative assumption, however, does not change the result stated in Theorem 5 that $T^{1/2}d_T \xrightarrow{d} N(0, \sigma_f^2 \Sigma_R)$ nor does it change the fact that $\bar{R}^e \xrightarrow{p} \mu_R$. As a consequence, the results in Theorem 5 go through unchanged.

5.7 Insufficient Spread in the Covariances

The rank condition for the TP-normalization is that d_0^+ should have full column rank. We have already studied the case where this rank condition fails if d_0 has less than full column rank. Another possibility is that there is a unique $k \times 1$ vector x, whose elements sum to 1 such that $\text{cov}(R_t^e, f_t'x) = \iota c$, where ι is an $n \times 1$ vector of ones, and $c \neq 0$ is some scalar constant.

Theorem 6 *Suppose assumptions 1 (or 1b), 2 and 3 are satisfied. Under additional regularity conditions provided in Hansen (1982)* $\hat{b}^* \xrightarrow{a.s.} b_s^* = (x - \beta_{\iota D}c)/(\mu'x)$, *where* $\beta_{\iota D} = (D_0'W_0^*D_0)^{-1}(D_0'W_0^*\iota)$ *and the predicted expected returns of the A-normalization converge almost surely to*

$$D_0 b_s^* = D_0 b_s^* = \mu_R + (\iota - D_0\beta_{\iota D})\, c/(\mu'x).$$

If ι *is in the space spanned by the columns of* D_0 *the model is is true since this would imply that* $\iota = D_0\beta_{\iota D}$. *Otherwise the model is false. Given assumption 2,* $\hat{b}^{\diamond}$ *also has a well defined*

probability limit, whether or not the model is true.

The proof of Theorem 6 is provided in Appendix A. It is clear that in a sufficiently large sample the model will be rejected if it is false.

6 Small Sample Properties of the GMM Estimators

To further demonstrate the sensitivity of empirical results to the choice of normalization in the presence of weak identification, I conduct Monte Carlo experiments. In each experiment I generate data from an artificial asset pricing model in which, by construction, three factors price 25 asset returns. I calibrate the model to resemble the Fama and French (1993) three factor model and the asset returns to resemble the FF25 portfolios.

I first study the properties of the A, M, and TP-normalizations when they are used to estimate the true model. I then study the properties of the three normalizations when the data are confronted with misspecified and, in some cases, under-identified models. The first of these models, which is calibrated to resembled the CAPM, is misspecified, in that it uses only the first of the three factors, but it is well identified. The second model uses a purely spurious factor, and is therefore misspecified. The third model, which is calibrated to resemble the consumption-based CCAPM, is misspecified. By construction, the factor has a common covariance with all of the returns, so the A and M-normalizations are identified, but the TP-normalization is not. The fourth model, which is calibrated to resemble the durables-based CCAPM, is misspecified. This model uses the consumption factor from the second model, as well as a purely spurious factor, which is calibrated to resemble durables growth in US data. The A-normalization is identified, but the M and TP-normalizations are not.

6.1 The Data Generating Process

I generate artificial data from a three factor model. The true SDF is given by $m_t = a - f_t'b$ where a is a scalar, f_t and b are $k \times 1$ vectors, and f_t follows the law of motion $f_t \sim Niid(\mu, \Sigma_f)$. The model of the true SDF is calibrated to mimic first stage GMM estimates for the Fama-French 3-factor model over the sample period 1949:Q1–2008:Q4.[7] I set $k = 3$, $a = 1.1404$ and $b = (\ 3.85 \quad 0.06 \quad 6.46\ )'$. I set μ and Σ_f equal to the sample mean and

[7] Details of the data and the Fama and French (1993) model are provided in Appendix B.

covariance matrix of the *Rm-Rf*, *SMB* and *HML* factors from the Fama-French database. It follows that $b^* = (\ 3.38 \quad 0.06 \quad 5.67\ )'$, $b^\diamond = (\ 3.86 \quad 0.06 \quad 6.48\ )'$ and $\xi = 0.9975$.

I generate an $n \times 1$ (with $n = 25$) vector of artificial excess returns $R_t^e = \mu_R + \beta(f_t - \mu) + \Psi\xi_t$ 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, and $\xi_t \sim Niid(0, I_n)$ and is independent of f_t. Given this definition 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 with actual data, I set Σ_R equal to its sample equivalent in the FF25 data. I set β equal to the matrix of factor betas for the FF25 returns regressed on *Rm-Rf*, *SMB* and *HML*, the three Fama-French factors. I set Ψ equal to the Cholesky decomposition of the covariance matrix of the residuals from those regressions.

Given the assumptions above we have

$$\begin{aligned} E(R_t^e m_t) &= E\left\{[\mu_R + \beta(f_t - \mu) + \Psi\xi_t]\,(a - f_t'b)\right\} \\ &= (a - \mu'b)\,[\mu_R - \beta\Sigma_f b/(a - \mu'b)]\,. \end{aligned} \tag{23}$$

To ensure that the right hand side of (23) is zero, I set $\mu_R = \beta\Sigma_f b/(a - \mu'b)$. This means that the model expected returns correspond to the model-predicted expected returns for the first stage GMM estimates for the Fama-French 3-factor model over the sample period 1949:Q1–2008:Q4.

Before proceeding to the results, it is important to note that in US data the b coefficient corresponding to the *SMB* factor is statistically insignificant, so it does not have an important role to play in pricing the FF25 portfolios. This does not mean, however, that it is a spurious factor. The *SMB* factor covaries with the portfolio returns, but not in a way that helps to explain the cross-sectional distribution of the expected returns. So there is no identification problem. In the simulated model, the simulated *SMB* factor does help to price the assets, but only marginally so, because the b coefficient corresponding to the *SMB* factor is numerically small.

6.2 Estimating the True Model

In these experiments the true SDF, m_t, prices the returns. In large samples GMM estimators based on the two normalizations deliver consistent parameter estimates and lead to correct inference about the model. To check small sample performance I simulate 10000 samples of 240 observations (the size of quarterly US data sample) each from the model.

Table 4 shows the results. For all three 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 *Rm-Rf* and *HML* factors) the parameters are statistically significant in almost all samples. They b parameter associated with the pseudo-SMB factor is usually not significant, consistent with it playing a very small role in pricing the returns. The R^2 associated with the model is usually large, and the test of the overidentifying restrictions usually does not reject the model.

Some differences across normalizations emerge. For example, after the initial GMM stage, the distribution of $\hat{b}^*$ (the A-normalization), drifts slightly away from the true values in that the median estimates of the coefficients on *Rm-Rf* and *HML* are closer to the true values of $b^\diamond$ than they are to the true values of b^*. Also, the R^2 associated with the A-normalization tends to be lower than for the other cases. Also, the slope coefficients associated with *Rm-Rf* are less often statistically significant for the TP-normalization than for the other cases. The test of the overidentifying restrictions has slightly excessive size with the M and TP-normalizations, with the opposite being true for the A-normalization. I find that these features of the simulations are much harder to discern if the sample size is increased to 1000.

6.3 Estimating a Pseudo-CAPM

Table 5 shows results for a second set of experiments in which I use the same data generating process, but the model being estimated uses only the first factor, the pseudo-*Rm-Rf* or CAPM factor. Since the model is misspecified, and since the pseudo-*HML* factor plays an important role in pricing the assets, we expect the estimated model fit to be less that perfect, and that the model should be rejected in large samples. As Table 5 indicates, even in samples as small as 240, the model is usually rejected at conventional significance levels, and the R^2 measure of fit is usually negative, regardless of the normalization.

One difference across normalizations emerges. For the A and M-normalizations the slope coefficient associated with the CAPM factor is usually positive and statistically significant, whereas for the TP-normalization this is not the case. This difference, while narrowing (in terms of percentages), persists in larger samples of 1000 observations. What explains this finding? GMM with the A and M-normalizations is akin to running a regression with factor betas on the right-hand side and average returns on the left-hand side, but no constant included in the regression. Given that the betas are all positive and the average returns are

all positive, the estimated regression line will tend to have a positive slope given that it has to go through the origin. However, when the regression line is fitted with a constant (as is the case with the TP-normalization) the significance of the slope coefficient now relies on whether the betas actually line up with the average returns. There is only a weak tendency of the betas with respect to the pseudo-CAPM factor to line up in the right way, so including a constant pricing error in the model tends to weaken the significance of the slope coefficient in the TP-normalization.

6.4 Estimating a Model with a Single, Spurious Factor

Table 6 shows results for a third set of experiments in which I use the same data generating process, but the model being estimated uses a single, completely spurious factor. I generate the proposed risk factor $x_{2t} = \mu_{x2} + u_{2t}$, where μ_{x2} is equal to the sample mean of quarterly US consumption growth in the period 1949Q1–2008Q4, and $u_{2t} \sim Niid(0, \sigma_{u2}^2)$ with σ_{u2}^2 equal to the sample variance of US consumption growth over the same period. Unlike actual consumption growth and the FF25 portfolio returns, this series is uncorrelated, by construction, with the simulated returns.

In this case, the three normalizations behave very differently. For the A-normalization the distribution of the estimates of b^* lies completely to the right of zero, $\hat{b}^*$ is almost always statistically significant, the model's R^2 is often very high (especially at the first GMM stage), and the model is rarely rejected (especially at the second GMM stage).

In contrast, for the M and TP-normalizations, the distribution of the estimates of $b^\diamond$ are centered roughly at zero, and are statistically significant much less often than for the A-normalization. The model's R^2 is usually negative. Interestingly, the test of the over-identifying restrictions is quite weak at the second GMM stage, but is quite successful in rejecting the model at later GMM stages. The contrast between the raw and demeaned normalizations becomes sharper if the sample size is increased to 1000.

6.5 Estimating a Pseudo–CCAPM

Table 7 shows results for a fourth set of experiments in which I use the same data generating process, but I estimated a model based on a pseudo-CCAPM factor. I generate the proposed risk factor, denoted x_{3t}, as follows:

$$x_{3t} = \mu_{x3} + c\iota'\Sigma_R^{-1}\left(R_t^e - \mu_R\right) + u_{3t},$$

where c is a scalar, ι is an $n \times 1$ vector of ones, and $u_{3t} \sim Niid(0, \sigma^2_{u3})$ is independent of R^e_t and u_{1t}. I set μ_{x3} equal to the sample mean of quarterly US consumption growth in the period 1949Q1–2008Q4. I set c equal to the cross-sectional average of the sample covariance between US consumption growth and the FF25 returns over the same period. I set σ^2_{u3} so that the variance of x_{3t} equals the sample variance of US consumption growth over the same period.

As Table 7 indicates, in samples of 240 observations, the A-normalization provides misleading inference about the model. The estimates of b^* are positive and statistically significant in nearly every sample. This is not unexpected. After all, x_{2t} is a relevant factor in that it is correlated with R^e_t. But the model also has good fit in many samples, with the median R^2 being 0.74 and 0.55 and 0.51 at the first, second and fifth stages of GMM. Additionally, the test of the pricing errors leads to rejection at the 5 percent level in only about 20 percent of the samples.

A very different picture emerges when the model is estimated using the M-normalization. Here $\hat{b}^\diamond$ is usually positive and statistically significant, reflecting the fact that the pseudo-CAPM factor is correlated with the returns. The fit of the model, on the other hand is quite poor with the R^2 being very low. Interestingly, the model is not rejected that often at the second GMM stage, but the test of the pricing errors becomes more powerful with further iterations on the GMM weighting matrix.

The TP-normalization presents yet another view of the model. Here the distribution of $\hat{b}^\diamond$ is centered roughly at zero, and $\hat{b}^\diamond$ is rarely statistically significant. This reflects the inclusion of the constant pricing error in the TP-normalization. With this constant included in the model, the fact that there is no spread among the factor covariances means that there is nothing left for the factor to price. The fit of the model is generally poor, but as was the case with the M-normalization, the model is not rejected that often at the second GMM stage, but the test of the pricing errors becomes more powerful with further iterations on the GMM weighting matrix. I find that the performance of the pricing error test improves dramatically in large samples. When the sample size in the simulations increases to 1000, the model is rejected in nearly every case, regardless of which normalization is adopted.

6.6 Estimating a Pseudo–Durables-CCAPM

Table 8 shows results for a fifth set of experiments in which I use the same data generating process, but estimate a two-factor pseudo-Durables-CCAPM model. The first factor is x_{3t}, defined above. The second factor is $x_{4t} = \mu_{x4} + u_{4t}$ where $u_{4t} \sim Niid(0, \sigma^2_{u4})$ is independent of R^e_t, u_{1t} and u_{3t}. I set μ_{x4} and σ^2_{u4}, respectively, equal to the sample mean and variance of US durables growth over the period 1949Q1–2008Q4.

The results for the A-normalization follow the predictions of Theorem 4. The estimates of b^*_2, the coefficient on the spurious factor, are nearly always found to be statistically significant. In contrast, the estimates of b^*_1, the coefficient on the more relevant factor, is less often found to be significant. The cross-sectional R^2 measures tend to be very high and the test of the over-identifying restrictions rarely leads to the model being rejected. These features of the A-normalization sharpen in larger samples.

The M and TP-normalizations paint a very different picture. For these normalizations, estimates of the slope coefficient on the durables factor are centered around zero, and are rarely statistically significant. The model's R^2 is usually quite low. But at the second stage of GMM, for both of these normalizations, the model is usually not rejected. It is only with further iterations of the weighting matrix that the model is rejected with reasonable frequency. These features of the M and TP-normalizations persist in larger samples because $b^\diamond_2$ is not identified asymptotically.

6.7 Tests for Identification

The results presented above show that the three normalizations lead to similar conclusions when the model being tested is the true model and all three of the normalizations are identified. But very different results emerge across normalizations when the model is misspecified, and one or more of the normalizations is not identified. Some normalizations tend to shed positive light on misspecified model when they are under-identified. This suggests that tests for lack of identification might be useful in guiding inference. A natural question is whether the tests I proposed above perform well in repeated samples.

The $L(r)$ statistic is used to test the null hypothesis that the rank of a matrix is r, with r being less than the number of columns in the matrix. If the $L(r)$ statistic does not exceed its 5 or 10 percent critical value, I state that "lack of identification was detected". If the opposite is true, I state that "lack of identification was not detected". The strength of the

test obviously depends on which critical value is used. Given the data generating processes used in the simulations, Table 9 summarizes the identification of each normalization in each of the 5 cases that were analyzed above. It also indicates the frequency with which, across simulations, the rank tests correctly assess the identification of the model.

In samples of 240 observations the tests perform quite well, with one exception. When the true model is estimated, all three of the normalizations are asymptotically identified, and the tests always conclude that the model is identified. The next model is the pseudo-CAPM, where only the first factor from the true model is included in the estimated SDF. In this case, even though the model is misspecified, it is identified for all normalizations because the covariances between the pseudo-CAPM factor and the returns are non-zero, and vary across assets. Once again, the rank tests always lead to the conclusion that the model is identified.

When the estimated model consists of a single spurious factor, only the A-normalization is identified. When 5 percent critical values are used, the testing procedure I described above correctly assesses the identification of the A-normalization in 72.2 percent of the samples, and the non-identification of the M and TP-normalizations in roughly 88 percent of the samples.

When the estimated model is the pseudo-CCAPM, the A and M-normalizations are identified, but the TP-normalization is not. The testing procedure correctly assesses the identification of the A and M-normalizations in 99.4 and 81.3 percent of the samples, and the non-identification of the TP-normalization in 88.5 percent of the samples.

When the estimated model is the pseudo-Durables CCAPM, the A-normalization is identified, but the M and TP-normalizations are not. This is the one case where the rank tests does not work as well. It only concludes that the A-normalization is identified in 26.2 percent of the samples. It correctly assesses the non-identification of the M and TP-normalizations in 94.1 and 99.3 percent of the samples.

The performance of the tests is as expected if the 10 percent critical value is adopted. This makes it more likely that the null hypothesis of non-identification will be rejected. As Table 9 also indicates, the tests become more accurate in larger samples. The size of the tests approaches asymptotic size.

Tests based on covariances appear, in general, to be more powerful than tests based on betas, in that across simulations beta-based tests are more likely to conclude in favor of identification.

While tests of the rank conditions are not entirely reliable (given the one example of the A-normalization in samples of 240 observations), tests of the rank condition for the TP-normalization seem to work very well, and are conservative, in that size tends to exceed asymptotic size in small samples. So reduced rank is rejected less often than it should be in finite samples. Failure of the rank condition for the TP-normalization is indicative of a problem with the proposed SDF, because it suggests the presence of a factor, or linear combination of factors, for which there is no spread in the covariances. At a minimum, in this case, a lower-dimensional model should be adopted.

7 Addressing Lack of Identification

7.1 Working with Gross Returns

At first glance, it might seem that the solution to the identification problems highlighted here would be to work with gross returns. While excess returns are often used in practice, working with gross returns is equally feasible, and certainly dispenses with the need to adopt a normalization. The moment conditions used to estimate the model become

$$E\left[R_t(a - f_t'b)\right] = \iota, \tag{24}$$

where $R_t = R_t^e + R_t^f \iota$, where R_t^f is the gross risk free rate between periods $t-1$ and t.

Working with gross returns does not make the problem of identification go away. To see this, consider the gradient of (24) with respect to a and b, which is the $n \times (k+1)$ matrix

$$G = \begin{pmatrix} E(R_t) & -E(R_t f_t') \end{pmatrix}.$$

This matrix must have rank $k+1$ for a and b to both be identified. This can only be true if $\text{cov}(R_t, f_t)$ has full column rank. To see this, notice that G is

$$G = \begin{pmatrix} E(R_t) & -E(R_t)E(f_t)' - \text{cov}(R_t, f_t) \end{pmatrix}.$$

Let $x^+ = \begin{pmatrix} x_0 & x' \end{pmatrix}'$, where x_0 is a scalar and x is a $k \times 1$ vector. Then

$$Gx^+ = E(R_t)\left[x_0 - E(f_t)'x\right] - \text{cov}(R_t, f_t)x.$$

If $\text{cov}(R_t, f_t)$ has less than full column rank, then there exists a non-zero x such that $\text{cov}(R_t, f_t)x = 0$. It follows that by setting $x_0 = E(f_t)'x$ there is a non-zero x^+ such that

$Gx^+ = 0$, in which case a and b are not identified. A researcher working with gross returns cannot be unconcerned with identification. The only difference is that the rank condition associated with the gross returns case is different from the one associated with excess returns, since the former requires that $\text{cov}(R_t, f_t)$ has full column rank, while the latter requires that $\text{cov}(R_t^e, f_t)$ has full column rank.[8]

Alternatively, a researcher might define $m_t = a - f_t'b$, and use the moment conditions $E(R_t^e m_t) = 0$ and $E(q_{t-1}) = E(m_t)$ to estimate a and b. Here q_t is the price of a risk free asset at time t. For this case, the gradient of the moment conditions with respect to a and b is

$$G = \begin{pmatrix} E(R_t^e) & -E(R_t^e f_t') \\ 1 & -E(f_t)' \end{pmatrix}.$$

Let $x^+ = \begin{pmatrix} x_0 & x' \end{pmatrix}'$, where x_0 is a scalar and x is a $k \times 1$ vector. Then

$$Gx^+ = \begin{pmatrix} E(R_t^e)\left[x_0 - E(f_t)'x\right] - \text{cov}(R_t^e, f_t)x \\ x_0 - E(f_t)'x \end{pmatrix}.$$

If $\text{cov}(R_t^e, f_t)$ has less than full column rank, then there exists a non-zero x such that $\text{cov}(R_t^e, f_t)x = 0$. It follows that by setting $x_0 = E(f_t)'x$ there is a non-zero x^+ such that $Gx^+ = 0$, in which case a and b are not identified. Here, identification fails under the same circumstances that it fails for the M-normalization.

7.2 Examining Betas over Subsamples

A less formal procedure than a rank test is sometimes used by researchers concerned about spurious factors. This involves examining factor betas over subsamples, looking for changes of sign. There are at least three problems with this procedure. First, it cannot address the possibility that identification fails due to colinearity among the betas across factors. Second, it does not address the issue of lack of spread among the betas. Third, while it is a reasonable procedure for detecting a purely spurious factor, which leads to non-identification, it is not obvious that it will work under weak identification [see Section 5.6]. In the latter case, estimated betas may converge to zero asymptotically with few or no sign switches.

[8] Of course, $\text{cov}(R_t, f_t) = \text{cov}(R_t^e, f_t) + \iota\,\text{cov}(R_t^f, f_t)$. This means that if $\text{cov}(R_t^e, f_t)$ has rank $k-1$, $\text{cov}(R_t, f_t)$ has full column rank as long as $\text{cov}(R_t^f, f_t)x \neq 0$ for the x such that $\text{cov}(R_t^e, f_t)x = 0$. If $\text{cov}(R_t^e, f_t)$ has rank $k-2$ or less, then $\text{cov}(R_t, f_t)$ is also of reduced rank.

7.3 Continuously Updated GMM

Hansen, Heaton and Yaron (1996) propose a variant of optimal GMM in which the weighting matrix is continuously updated as a function of the parameter vector (CU-GMM). CU-GMM estimates of the model under the two normalizations are equivalent in finite samples, in the sense that the errors $1 - f_t'\hat{b}^*$ and $1 - (f_t - \hat{\mu})'\hat{b}^\diamond$ are proportional to one another, $\hat{b}^\diamond = \hat{b}^*/(1 - \hat{\mu}'\hat{b}^*)$, and the test statistic for the overidentifying restrictions is numerically identical [Penaranda and Santana (2010)]. CU-GMM does not, however, resolve the identification problem. I demonstrate this in Appendix C for the single factor case. In the case where the risk factor being used is spurious, but has a non-zero mean, the A-normalization combined with the CU-GMM estimator leads to positive inference about the model. The M-normalization, while numerically equivalent to the A-normalization in finite samples, suffers from an identification problem in the limit.

8 Conclusion

When excess returns are used to estimate linear SDFs, GMM estimation requires that a normalization of the SDF be adopted. Standard normalizations of the SDF using raw or demeaned factors are asymptotically equivalent when the model is true and identified. The conditions under which these normalizations are identified, however, are 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 wildly in terms of statistical significance, measures of fit can differ dramatically, and tests of overidentifying restrictions can differ sharply in outcome. I have demonstrated this, here, for consumption-based models fit to US data. My Monte Carlo simulations demonstrate that model misspecification, combined with identification problems, is a plausible explanation of these empirical findings.

The concrete message of this paper is that researchers can easily check their models for identification using tests of rank conditions. Monte Carlo evidence suggests that these tests are powerful in detecting failure of the rank conditions. When identification problems seem to be present, researchers should be cautious in conducting inference.

REFERENCES

Black, Fischer, Michael Jensen and Myron Scholes (1972) "The Capital Asset Pricing Model: Some Empirical Tests," in Michael Jensen, ed. *Studies in the Theory of Capital Markets*. Praeger: New York.

Burnside, Craig (2007) "Empirical Asset Pricing and Statistical Power in the Presence of Weak Risk Factors," NBER Working Paper No. 13357.

Campbell, John Y. (2003) "Consumption-Based Asset Pricing," in George M. Constantinides, Milton Harris, and René Stulz, ed.: *Handbook of the Economics of Finance*, Vol. 1B, pp. 801–885. Elsevier: Amsterdam.

Chen, Naifu, Richard R. Roll, and Stephen A. Ross (1986) "Economic Forces and the Stock Market," *Journal of Business* 59, 383–403.

Cochrane, John H. (2005) *Asset Pricing*, Revised edition. Princeton: Princeton University Press.

Cragg, John G. and Stephen G. Donald (1997) "Inferring the Rank of a Matrix," *Journal of Econometrics* 76, 223–50.

Daniel, Kent and Sheridan Titman (2012) "Testing Factor-Model Explanations of Market Anomalies," *Critical Finance Review* 1, 103–39.

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* 71, 607–36.

Ferson, Wayne E. and Campbell Harvey (1993) "The Risk and Predictability of International Equity Returns," *Review of Financial Studies* 6, 527–566.

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

Hansen, Lars P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029–54.

Hansen, Lars P., John C. Heaton and Amir Yaron (1996) "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics* 14, 262–80.

Hansen, Lars P. and Ravi Jagannathan (1997) "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance* 52, 557-90.

Hou, Kewei, and Robert Kimmel (2006) "On the 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," *Journal of Finance* 62, 1623–61.

Kan, Raymond and Cesare Robotti (2008) "Specification Tests of Asset Pricing Models Using Excess Returns," *Journal of Empirical Finance* 15, 816–38.

Kan, Raymond and Cesare Robotti (2009) "Model Comparison Using the Hansen-Jagannathan Distance," *Review of Financial Studies* 22, 3449–90.

Kan, Raymond, Cesare Robotti and Jay Shanken (2012) "Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology," *Journal of Finance*, forthcoming.

Kan, Raymond and Chu Zhang (1999a) "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics* 54, 103–27.

Kan, Raymond and Chu Zhang (1999b) "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance* 54, 203–35.

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, 1238–87.

Lewellen, Jonathan and Stefan Nagel (2006) "The Conditional CAPM Does Not Explain Asset-Pricing Anomalies," *Journal of Financial Economics* 82, 289–314.

Lewellen, Jonathan, Stefan Nagel and Jay Shanken (2010) "A Skeptical Appraisal of Asset-Pricing Tests," *Journal of Financial Economics* 96, 175–94.

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

Parker, Jonathan A. and Christian Julliard (2005) "Consumption Risk and the Cross Section of Expected Returns," *Journal of Political Economy* 113, 185–222.

Penaranda, Francisco and Enrique Santana (2010) "A Unifying Approach to the Empirical Evaluation of Asset Pricing Models," CEPR Discussion Paper No. DP7943.

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.

Staiger, Douglas and James H. Stock and (1997) "Instrumental Variables Regression with Weak Instruments," *Econometrica* 65: 557–86.

Wright, Jonathan H. (2003) "Detecting Lack of Identification in GMM," *Econometric Theory* 19, 322–30.

Yogo, Motohiro (2006) "A Consumption-Based Explanation of Expected Stock Returns," *Journal of Finance* 61, 539–80.

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

		Test of ...				
	Number of	A)		B)	C)	
	factors	rank[cov(R^e, f)] $= k-1$		rank$[E(R^e f')]$	rank$[\iota$ cov$(R^e, f)] = k$	
Model	(k)	Covariance test	Beta test	$= k-1$	Covariance test	Beta tests
CAPM	1	0.000	0.000	0.000	0.000	0.000
Fama-French 3 factors	3	0.000	0.000	0.000	0.001	0.000
CCAPM	1	0.036	0.002	0.000	0.241	0.018
Durables-CCAPM	2	0.977	0.518	0.014	0.968	0.509
Yogo (2006)	3	0.852	0.995	0.558	0.819	0.993
Lettau and Ludvigson (2001)	3	0.675	0.000	0.376	0.686	0.001
Jagannathan and Wang (2007)	1	0.000	0.000	0.000	0.000	0.000
Lustig and Verdelhan (2007)	3	0.732	0.657	0.790	0.783	0.699
Currency portfolios, CCAPM	1	0.869	0.876	0.000	0.871	0.773

Note: The table presents results from testing the rank conditions associated with each of the normalizations. The p-value associated with the null hypothesis of reduced rank (non-identification) is shown. The tests in panel A are relevant for the M-normalization, for which factors are demeaned. Tests based on the covariance matrix, cov(R^e, f), and the matrix of factor betas are shown. The tests in panel B are relevant for the A-normalization, for which raw risk factor are used. The tests in panel C are relevant for the TP-normalization, which uses demeaned risk factors and a common pricing error across moment conditions. Results are presented for nine models. For the first seven cases, the asset returns studied are the real quarterly excess returns to the Fama and French (1993) 25 portfolios sorted on the basis of size and book-to-market value (FF25 portfolios). The CAPM case uses these returns and the market excess return (CAPM) as a risk factor. The Fama-French 3 factor case uses the FF25 returns, and the three Fama and French (1993) risk factors: the market excess return, the SMB factor and the HML factor. The CCAPM case uses the FF25 returns and consumption growth as a risk factor. The Durables-CCAPM case uses the FF25 returns and consumption and durables growth as risk factors. The Yogo (2006) case uses the FF25 returns, and consumption growth, diurables growth, and the market returns as risk factors. The Lettau and Ludvigson (2001) case uses the FF25 returns, consumption growth, *cay*, and the product of consumption growth and *cay* as risk factors. The Jagannathan and Wang (2007) case uses the FF25 returns on an annual basis and Q4-Q4 consumption growth as the risk factor. The Lustig and Verdelhan (2007) case uses eight currency portfolios sorted by interest rate differential versus the US, and the risk factors from Yogo (2006), all at the annual frequency. The currency portfolios CCAPM case uses five currency portfolios sorted by interest rate differential versus the US, at the quarterly frequency, and uses consumption growth as the risk factor. Details of all data sets are provided in Appendix C.

TABLE 2: **GMM Estimates of the CCAPM Model**

	First Stage		Second Stage			Fifth Stage		
Factor	b	R^2	b	R^2	J	b	R^2	J
	A-Normalization (b^*)							
Consumption growth	114.0 (23.2)	0.75	137.1 (10.1)	0.14	31.0 (0.153)	140.7 (11.1)	−0.07	28.1 (0.254)
	M-Normalization ($b^\diamond$)							
Consumption growth	245.9 (111.3)	−0.25	131.7 (54..6)	−3.24	44.2 (0.007)	100.1 (37.5)	−5.12	67.7 (0.000)
	TP-Normalization (α and $b^\diamond$)							
Pricing error (α)	1.67 (0.63)	0.06	2.16 (0.42)	−0.23	52.8 (0.000)	2.25 (0.41)	−0.29	54.2 (0.000)
Consumption growth	76.1 (79.0)		59.0 (36.6)			53.1 (35.5)		

Note: Quarterly data, 1949–2008. The table reports first, second and fifth stage GMM estimates, obtained using the moment restriction $E(R_t^e m_t) = 0$, where R_t^e is a 25×1 vector of excess returns of the Fama-French 25 portfolios of US stocks sorted on size and the book-to-market value ratio, and m_t is the SDF. For the A-normalization the SDF is $m_t = 1 - f_t b^*$, where f_t is real per capita consumption (nondurables & services) growth. For the M-normalization the SDF is $m_t = 1 - (f_t - \mu)b^\diamond$. For the TP-normalization the SDF is $m_t = 1 - (f_t - \mu)b^\diamond$ and the moment condition used is $E(R_t^e m_t - \alpha) = 0$. GMM-VARHAC standard errors are reported in parentheses for $\hat{b}^*$, $\hat{b}^\diamond$ and $\hat{\alpha}$. The table reports the R^2 measure of cross-sectional fit between the sample mean of R_t^e and the model predicted mean returns defined in the text. Tests of the overidentifying restrictions are also reported. The test statistic, J, is presented along with the associated p-value in parentheses.

TABLE 3: **GMM Estimates of the Durables-CCAPM Model**

	First Stage		Second Stage			Fifth Stage		
Factor	b	R^2	b	R^2	J	b	R^2	J
	A-Normalization (b^*)							
Consumption growth	42.2	0.98	46.7	0.98	17.5	50.5	0.98	18.4
	(30.3)		(13.7)		(0.782)	(13.3)		(0.735)
Durables growth	71.1		66.2			63.7		
	(19.7)		(8.6)			(8.2)		
	M-Normalization ($b^\diamond$)							
Consumption growth	299.2	−0.16	73.1	−7.56	16.9	90.4	−6.89	57.1
	(149.2)		(78.1)		(0.815)	(39.8)		(0.000)
Durables growth	149.7		16.7			43.1		
	(164.3)		(68.5)			(37.1)		
	TP-Normalization (α and $b^\diamond$)							
Pricing error (α)	1.70	0.16	1.17	−3.16	15.8	1.95	0.07	45.5
	(0.98)		(0.67)		(0.826)	(0.43)		(0.002)
Consumption growth	129.7		24.0			51.3		
	(103.3)		(59.2)			(38.3)		
Durables growth	160.3		46.3			41.9		
	(129.9)		(48.6)			(34.9)		

Note: Quarterly data, 1949–2008. The table reports first, second and fifth stage GMM estimates, obtained using the moment restriction $E(R_t^e m_t) = 0$, where R_t^e is a 25×1 vector of excess returns of the Fama-French 25 portfolios of US stocks sorted on size and the book-to-market value ratio, and m_t is the SDF. For the A-normalization the SDF is $m_t = 1 - f_t' b^*$, where f_t is a 2×1 vector containing real per capita consumption (nondurables & services) growth, and durable consumption growth. For the M-normalization the SDF is $m_t = 1 - (f_t - \mu)' b^\diamond$. For the TP-normalization the SDF is $m_t = 1 - (f_t - \mu)' b^\diamond$ and the moment condition used is $E(R_t^e m_t - \alpha) = 0$. GMM-VARHAC standard errors are reported in parentheses for $\hat{b}^*$, $\hat{b}^\diamond$ and $\hat{\alpha}$. The table reports the R^2 measure of cross-sectional fit between the sample mean of R_t^e and the model predicted mean returns defined in the text. Tests of the overidentifying restrictions are also reported. The test statistic, J, is presented along with the associated p-value in parentheses.

TABLE 4: **Monte Carlo Experiments with Artificial Quarterly Data; Estimation of the True Model**

	GMM Stage 1					*GMM Stage 2*					*GMM Stage 5*				
	Percentiles			Percent Significant at		Percentiles			Percent Significant at		Percentiles			Percent Significant at	
	5	50	95	10% level	5% level	5	50	95	10% level	5% level	5	50	95	10% level	5% level
							A-Normalization								
b_1^*	2.02	3.39	4.72	99.0	97.9	2.27	3.80	5.25	99.6	99.2	2.31	3.87	5.38	99.7	99.3
b_2^*	-2.14	0.07	2.24	11.2	5.8	-2.33	0.11	2.51	17.6	10.7	-2.40	0.13	2.61	19.2	12.0
b_3^*	3.86	5.68	7.42	99.9	99.8	4.36	6.35	8.26	100	99.9	4.45	6.47	8.44	100	99.9
R^2	0.84	0.94	0.98			-0.35	0.69	0.94			-1.07	0.60	0.93		
J						13.1	21.8	33.2	9.3	3.9	13.1	21.7	33.0	8.7	3.7
							M-Normalization								
$b_1^\diamond$	2.26	3.90	5.75	98.9	97.6	2.24	3.90	5.78	99.3	98.0	2.25	3.91	5.83	99.3	98.1
$b_2^\diamond$	-2.45	0.09	2.61	10.1	5.0	-2.47	0.08	2.59	12.4	6.4	-2.49	0.09	2.63	12.9	6.7
$b_3^\diamond$	4.22	6.53	9.11	99.9	99.8	4.21	6.53	9.17	99.9	99.8	4.22	6.56	9.24	99.9	99.8
R^2	0.81	0.92	0.97			0.53	0.86	0.95			0.42	0.85	0.95		
J						13.4	22.8	35.6	13.8	7.4	13.4	22.7	35.6	13.8	7.3
							TP-Normalization								
$b_1^\diamond$	0.23	3.34	6.93	50.0	35.8	0.55	3.20	6.23	62.9	50.7	0.50	3.20	6.27	63.4	51.4
$b_2^\diamond$	-2.65	0.37	3.27	10.3	5.2	-2.37	0.48	3.25	13.1	7.2	-2.38	0.50	3.27	13.6	7.5
$b_3^\diamond$	3.81	6.27	9.13	99.4	98.7	3.81	6.22	9.04	99.7	99.2	3.82	6.25	9.11	99.7	99.2
R^2	0.82	0.93	0.97			0.56	0.87	0.95			0.50	0.86	0.95		
J						12.4	21.5	34.2	13.0	7.0	12.4	21.6	34.3	13.5	7.5

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 240$. The true risk factors, $f_t \sim Niid(\mu, \Sigma_f)$ with μ and Σ_f equal to the sample mean and covariance matrix of the Fama-French factors (1949Q1–2008Q4). The SDF is given by $m_t = a - f_t'b$, with $a = 1.1404$ $b = (\ 3.85 \quad 0.06 \quad 6.46\ )'$. I generate $R_t^e = \mu_R + \beta(f_t - \mu) + \Psi\xi_t$, where μ_R is a 25×1 vector, β is an 25×3 matrix, Ψ is a 25×25 lower triangular matrix, and $\xi_t \sim Niid(0, I_n)$ and is independent of f_t. The elements of β and Ψ are set so that the model-implied var(R_t^e) and cov(R_t^e, f_t) are equal to their sample equivalents for the FF25 portfolios and Fama-French factors (1949Q1–2008Q4). The vector μ_R is set to ensure that $E(R_t^e m_t) = 0$. The model is estimated by GMM using the A, M and TP-normalizations.

TABLE 5: **Monte Carlo Experiments with Artificial Quarterly Data; Estimation of the Pseudo-CAPM Model**

	GMM Stage 1					*GMM Stage 2*					*GMM Stage 5*				
	Percentiles			Percent Significant at		Percentiles			Percent Significant at		Percentiles			Percent Significant at	
	5	50	95	10% level	5% level	5	50	95	10% level	5% level	5	50	95	10% level	5% level
							A-Normalization								
b^*	1.76	3.02	4.28	98.3	96.7	1.90	3.42	4.90	98.8	97.9	1.93	3.61	5.20	98.8	97.8
R^2	-0.94	-0.35	0.13			-2.55	-0.61	0.02			-4.63	-0.85	-0.02		
J						28.9	42.8	58.8	86.8	76.5	28.8	42.5	58.0	86.6	75.9
							M-Normalization								
$b^\diamond$	1.78	3.17	4.76	98.2	96.2	1.36	2.66	4.21	95.6	90.9	1.26	2.57	4.13	94.3	89.2
R^2	-1.21	-0.49	0.00			-1.89	-0.90	-0.35			-2.15	-1.06	-0.44		
J						29.3	43.9	60.9	88.1	79.1	29.6	44.4	61.5	88.9	80.4
							TP-Normalization								
$b^\diamond$	-3.32	-0.66	1.94	13.5	7.1	-1.34	0.64	2.63	17.2	10.2	-1.04	0.92	2.96	22.7	14.6
R^2	0.00	0.05	0.34			-1.20	-0.67	-0.07			-1.78	-0.98	-0.25		
J						27.1	40.8	57.5	84.1	73.7	27.0	40.9	57.7	84.2	73.8

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 240$. The returns and true risk factors are the same as in Table 4. The estimated model uses the pseudo-CAPM factor as a single factor. It is defined as the first of the three true risk factors. The model is estimated by GMM using the A, M and TP-normalizations.

TABLE 6: **Monte Carlo Experiments with Artificial Quarterly Data; Estimation of a Spurious Factor Model**

	GMM Stage 1					*GMM Stage 2*					*GMM Stage 5*				
	Percentiles			Percent Significant at		Percentiles			Percent Significant at		Percentiles			Percent Significant at	
	5	50	95	10% level	5% level	5	50	95	10% level	5% level	5	50	95	10% level	5% level
							A-Normalization								
b^*	135	200	326	98.2	96.3	114	142	175	99.6	99.3	109	137	169	100	100
R^2	-0.67	0.78	0.94			–6.99	-0.45	0.83			-7.61	-0.67	0.81		
J						6.9	16.4	29.7	2.3	0.9	16.6	27.6	41.0	23.2	12.8
							M-Normalization								
$b^\diamond$	-1135	4.6	1170	11.5	4.1	-129	0.6	131	4.5	2.5	-115	-0.1	117	32.5	23.1
R^2	-21.4	-3.65	0.04			-35.0	-13.2	-3.26			-35.7	-13.6	-3.6		
J						1.7	6.7	32.5	4.7	3.8	31.4	48.1	66.4	92.5	87.1
							TP-Normalization								
$b^\diamond$	-346	5.0	351	31.7	14.9	-114	0.0	116	17.0	9.2	-104	0.0	107	26.3	17.4
R^2	0.00	0.13	0.61			-6.79	-1.71	0.02			-8.38	-2.25	-0.02		
J						10.3	27.1	49.0	36.6	28.5	24.3	38.9	56.2	77.5	65.2

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 240$. The returns and true risk factors are the same as in Table 4. The estimated model uses a single spurious factor. I generate the factor as $x_{2t} = \mu_{x2} + u_{2t}$, where μ_{x2} is equal to the sample mean of quarterly US consumption growth in the period 1949Q1–2008Q4, and $u_{2t} \sim Niid(0, \sigma^2_{u2})$ with σ^2_{u2} equal to the sample variance of US consumption growth over the same period. The model is estimated by GMM using the A, M and TP-normalizations.

TABLE 7: **Monte Carlo Experiments with Artificial Quarterly Data; Estimation of the Pseudo-CCAPM Model**

	GMM Stage 1					*GMM Stage 2*					*GMM Stage 5*				
	Percentiles			Percent Significant at		Percentiles			Percent Significant at		Percentiles			Percent Significant at	
	5	50	95	10% level	5% level	5	50	95	10% level	5% level	5	50	95	10% level	5% level
							A-Normalization								
b^*	80.6	115	156	99.8	99.6	91.0	112	135	100	100	90.4	112	135	100	100
R^2	0.42	0.74	0.91			-0.91	0.55	0.85			-1.24	0.51	0.85		
J						18.9	29.9	42.6	32.7	18.9	19.9	31.0	43.6	37.6	23.0
							M-Normalization								
$b^\diamond$	123	249	526	87.6	70.9	23.2	89.3	162	58.6	45.0	9.04	74.4	149	73.6	64.9
R^2	-3.22	-0.28	0.43			-20.8	-6.10	-0.49			-22.3	–7.3	-0.70		
J						8.9	22.8	41.4	16.7	10.5	26.7	43.5	62.4	84.2	75.4
							TP-Normalization								
$b^\diamond$	-349	3.7	352	31.8	14.2	-112	-0.50	116	16.0	8.9	-102	-0.78	105	25.4	16.8
R^2	0.00	0.13	0.61			-6.94	-1.70	0.02			-8.48	-2.23	-0.01		
J						10.3	27.0	49.5	36.7	28.8	24.2	39.1	56.1	77.4	65.9

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 240$. The returns and true risk factors are the same as in Table 4. The estimated model uses the pseudo-CCAPM factor as a single factor. It is defined as $x_{3t} = \mu_{x3} + c\iota'\Sigma_R^{-1}(R_t^e - \mu_R) + u_{3t}$, where c is a scalar, ι is an $n \times 1$ vector of ones, and $u_{3t} \sim Niid(0, \sigma_{u3}^2)$ is independent of R_t^e and u_{1t}, where μ_{x3} is set equal to the sample mean of quarterly US consumption growth in the period 1949Q1–2008Q4, c is set equal to the cross-sectional average of the sample covariance between US consumption growth and the FF25 returns over the same period, and σ_{u3}^2 is set so that the variance of x_{3t} equals the sample variance of US consumption growth over the same period. The model is estimated by GMM using the A, M and TP-normalizations.

TABLE 8: **Monte Carlo Experiments with Artificial Quarterly Data; Estimation of the Pseudo-Durables-CCAPM Model**

	GMM Stage 1					*GMM Stage 2*					*GMM Stage 5*				
	Percentiles			Percent Significant at		Percentiles			Percent Significant at		Percentiles			Percent Significant at	
	5	50	95	10% level	5% level	5	50	95	10% level	5% level	5	50	95	10% level	5% level
							A-Normalization								
b_1^*	-28.1	20.7	68.3	21.9	14.3	5.5	29.5	53.1	69.0	60.0	6.7	30.6	54.3	77.7	70.0
b_2^*	47.4	80.8	114	96.7	95.0	55.1	71.4	87.9	100	100	53.8	70.3	86.6	100	100
R^2	0.87	0.97	0.99			0.41	0.90	0.98			0.25	0.88	0.98		
J						11.0	19.7	30.9	3.6	1.5	14.3	23.3	34.8	10.0	4.6
							M-Normalization								
$b_1^\diamond$	64.7	222	457	64.9	49.5	10.1	83.9	160	42.6	30.0	4.7	74.5	152	70.3	60.8
$b_2^\diamond$	-423	-7.2	409	17.4	7.1	-114	-0.6	111	8.4	4.2	-107	-1.1	102	27.1	18.6
R^2	-1.76	0.02	0.63			-20.0	-5.77	-0.33			-21.3	-6.6	-0.5		
J						5.8	15.9	32.7	5.7	3.3	23.0	39.6	58.8	76.3	66.0
							TP-Normalization								
$b_1^\diamond$	-344	2.8	345	25.1	10.8	-113	0.9	117	9.7	4.7	-107	-0.3	109	24.3	15.8
$b_2^\diamond$	-323	-4.9	314	25.5	11.5	-108	-0.9	106	9.8	4.8	-101	-1.0	98.5	24.4	16.2
R^2	0.04	0.36	0.74			-7.18	-1.61	0.15			-8.23	-1.97	0.09		
J						7.4	16.8	37.2	11.7	8.0	20.8	35.5	52.8	68.8	56.5

Note: The table reports results from 10000 Monte Carlo experiments with sample size $T = 240$. The returns and true risk factors are the same as in Table 4. The estimated model uses x_{3t} and x_{4t} as risk factors, where x_{3t} is the pseudo-CCAPM factor defined in Table 7 and $x_{4t} = \mu_{x4} + u_{4t}$, where $u_{4t} \sim Niid(0, \sigma_{u4}^2)$ is independent of R_t^e and u_{1t} and u_{3t}, μ_{x4} is set equal to the sample mean of quarterly US durable consumption growth in the period 1949Q1–2008Q4, and σ_{u4}^2 is set equal to the sample variance of US durables growth over the same period. The model is estimated by GMM using the A, M and TP-normalizations.

TABLE 9: **Performance of the Rank Tests in the Monte Carlo Experiments (percentage of samples in which the rank test correctly determines matrix rank)**

	A-normalization		M-normalization			TP-normalization		
	Identified?	D_0-test	Identified?	d_0-test	β-test	Identified?	$[\iota\ d_0]$-test	$[\iota\ \beta]$-test
				Using 5% Critical Values in Samples of 240 Observations				
True DGP	Yes	100	Yes	100	100	Yes	100	100
Pseudo CAPM	Yes	100	Yes	100	100	Yes	100	100
Spurious factor model	Yes	72.2	No	87.7	64.6	No	88.6	64.6
Pseudo C-CAPM model	Yes	99.4	Yes	81.3	95.3	No	88.5	95.3
Pseudo Durables C-CAPM model	Yes	26.2	No	94.1	76.0	No	99.3	76.0
				Using 10% Critical Values in Samples of 240 Observations				
True DGP	Yes	100	Yes	100	100	Yes	100	100
Pseudo CAPM	Yes	100	Yes	100	100	Yes	100	100
Spurious factor model	Yes	81.6	No	79.5	54.8	No	80.4	54.8
Pseudo C-CAPM model	Yes	99.8	Yes	88.5	97.2	No	80.6	97.2
Pseudo Durables C-CAPM model	Yes	38.6	No	88.0	65.2	No	98.4	65.2
				Using 5% Critical Values in Samples of 1000 Observations				
True DGP	Yes	100	Yes	100	100	Yes	100	100
Pseudo CAPM	Yes	100	Yes	100	100	Yes	100	100
Spurious factor model	Yes	100	No	93.0	89.0	No	93.0	89.0
Pseudo C-CAPM model	Yes	100	Yes	100	100	No	93.0	100.0
Pseudo Durables C-CAPM model	Yes	90.0	No	96.0	96.0	No	100	96.0
				Using 10% Critical Values in Samples of 1000 Observations				
True DGP	Yes	100	Yes	100	100	Yes	100	100.0
Pseudo CAPM	Yes	100	Yes	100	100	Yes	100	100.0
Spurious factor model	Yes	100	No	84.0	81.0	No	85.0	81.0
Pseudo C-CAPM model	Yes	100	Yes	100	100	No	91.0	100.0
Pseudo Durables C-CAPM model	Yes	95.0	No	95.0	90.0	No	100	90.0

Note: The table summarizes the performance of the rank testing procedure across repeated samples. The model is declared "not-identified" k if the null ypothsis of reduced rank is not rejected at the 5 or 10 percent critical values of the chi-squared distribution associated with the rank test statistics. Otherwise the model is declared "identified". The table reports whether each normalization is identified in population for the various factor models studied in the Monte Carlo experiments. The table reports the percentage of samples in which the "identified" or "not-identified" declaration based on the test statistic matches the limiting identification of the normalization.

A Proofs and Other Details

A.1 Estimation of the TP Normalization

Using the $n+k$ moment restrictions given by (9), $b^\diamond$, α and μ are estimated using GMM. Define $u_{1t}^\diamond(\alpha, b^\diamond, \mu) = R_t^e m_t^\diamond - \alpha = R_t^e[1-(f_t-\mu)'b^\diamond] - \alpha$ and let $g_{1T}^\diamond(\alpha, b^\diamond, \mu) = \frac{1}{T}\sum_{t=1}^T u_{1t}^\diamond = \bar{R}^e - \alpha - \left(D_T - \bar{R}^e\mu'\right)b^\diamond$. Define $u_{2t}^\diamond(\mu) = f_t - \mu$ and let $g_{2T}^\diamond(\mu) = \frac{1}{T}\sum_{t=1}^T u_{2t}^\diamond = \bar{f} - \mu$. Define $u_t^\diamond = \left(\begin{array}{cc} u_{1t}^{\diamond\prime} & u_{2t}^{\diamond\prime}\end{array}\right)'$ and $g_T^\diamond = \left(\begin{array}{cc} g_{1T}^{\diamond\prime} & g_{2T}^{\diamond\prime}\end{array}\right)'$. Let $\tilde{d}_T = \left(\begin{array}{cc} \iota & d_T\end{array}\right)$, where ι is an $n\times 1$ vector of ones.

I consider GMM estimators that set $\tilde{a}_T^\diamond g_T^\diamond = 0$, where $\tilde{a}_T^\diamond$ is a $(2k+1)\times(n+k)$ matrix and takes the form

$$\tilde{a}_T^\diamond = \left(\begin{array}{cc} \tilde{d}_T' W_T^\diamond & 0 \\ 0 & I_k \end{array}\right), \tag{A1}$$

and $W_T^\diamond$ is an $n\times n$ positive definite weighting matrix. It follows that the GMM estimators are

$$\left(\begin{array}{c}\hat{\alpha}\\ \hat{b}^\diamond\end{array}\right) = \left(\tilde{d}_T' W_T^\diamond \tilde{d}_T\right)^{-1}\tilde{d}_T' W_T^\diamond \bar{R}^e \tag{A2}$$

$$\hat{\mu} = \bar{f}. \tag{A3}$$

I consider multi-stage GMM estimators. In the first stage $W_T^\diamond = I_n$. In the jth subsequent stage, $W_T^\diamond = (P_T S_T^\diamond P_T')^{-1}$ where $P_T = \left(\begin{array}{cc} I_n & \bar{R}^e(\hat{b}_{j-1}^\diamond)'\end{array}\right)$, $\hat{b}_j^*$ represents the jth-stage estimator of $b^\diamond$ and $S_T^\diamond$ is a consistent estimator of $S_0^\diamond = \sum_{j=-\infty}^{+\infty} E(u_t^\diamond u_{t-j}^{\diamond\prime})$.

Let

$$\tilde{\delta}_T^\diamond = \left(\begin{array}{cc} -\tilde{d}_T & \bar{R}^e\hat{b}^{\diamond\prime} \\ 0 & -I_k \end{array}\right). \tag{A4}$$

A test of the pricing errors is based on

$$\tilde{J}^\diamond = T g_T(\hat{\alpha}, \hat{b}^\diamond, \hat{\mu})'(\tilde{V}_g^\diamond)^+ g_T(\hat{\alpha}, \hat{b}^\diamond, \hat{\mu}), \tag{A5}$$

where

$$\tilde{V}_g^\diamond = \tilde{A}_T^\diamond S_T^\diamond \tilde{A}_T^{\diamond\prime} \text{ with } \tilde{A}_T^\diamond = I_{n+k} - \tilde{\delta}_T^\diamond\left(\tilde{a}_T^\diamond\tilde{\delta}_T^\diamond\right)^{-1}\tilde{a}_T^\diamond. \tag{A6}$$

The cross-sectional R^2 measure is

$$\tilde{R}_\diamond^2 = 1 - \frac{(\bar{R}^e - \hat{\alpha} - d_T\hat{b}^\diamond)'(\bar{R}^e - \hat{\alpha} - d_T\hat{b}^\diamond)}{(\bar{R}^e - \ddot{R}^e)'(\bar{R}^e - \ddot{R}^e)}. \tag{A7}$$

A.2 Proof of Theorem 1

The additional regularity conditions required for consistency are stated in Hansen's Theorem 2.1. It follows from assumption 3 that $a_T^* \xrightarrow{a.s.} a_0^* = D_0' W_0^*$. Define $h_0^*(b^*) = a_0^*[E(R^e) - D_0 b^*]$. Given that D_0 has full column rank and W_0^* is positive definite, the function $h_0^*(b^*)$ has a unique zero, $b_z^* = (D_0' W_0^* D_0)^{-1} D_0' W_0^* E(R^e)$. Since the model is true $E(R^e) = D_0 b_0^*$. Substituting this into the expression for b_z^* we have $b_z^* = b_0^*$. From Hansen's (1982) Theorem 2.1, $\hat{b}^* \xrightarrow{a.s.} b_0^*$.

Similarly, it follows from assumption 3 that $a_T^\diamond \xrightarrow{a.s.} a_0^\diamond$ with

$$a_0^\diamond = \begin{pmatrix} d_0' W_0^\diamond & 0 \\ 0 & I_k \end{pmatrix}.$$

Define

$$h_0^\diamond(b^\diamond, \mu) = a_0^\diamond \begin{pmatrix} E(R^e) - d_0 b^\diamond \\ \mu_0 - \mu \end{pmatrix}.$$

Given that d_0 has full column rank and $W_0^\diamond$ is positive definite, the function $h_0^\diamond(b^\diamond, \mu)$ has a unique zero, $b_z^\diamond = (d_0' W_0^\diamond d_0)^{-1} d_0' W_0^\diamond E(R^e)$, $\mu_z = \mu_0$. Since the model is true $E(R^e) = d_0 b_0^\diamond$. Substituting this into the expression for $b_z^\diamond$ we have $b_z^\diamond = b_0^\diamond$. From Hansen's (1982) Theorem 2.1, $\hat{b}^\diamond \xrightarrow{a.s.} b_0^\diamond$.

The matrices δ_0^* and $\delta_0^\diamond$ have full column rank due to the properties of D_0 and d_0. It follows from Hansen's Theorem 3.1 that $\hat{b}^*$ and $\hat{\theta}^\diamond$ have the asymptotic distributions stated in the theorem.

The model-predicted expected returns are $D_T \hat{b}^*$ and $d_T \hat{b}^\diamond$, respectively, for the two normalizations. Given that results above these both converge almost surely to μ_R and, therefore, we get the result that $R_*^2 \xrightarrow{a.s.} 1$ and $R_\diamond^2 \xrightarrow{a.s.} 1$.

From the results above it follows that $\delta_T^* \xrightarrow{a.s.} \delta_0^*$, $\delta_T^\diamond \xrightarrow{a.s.} \delta_0^\diamond$, $S_T^* \xrightarrow{a.s.} S_0^*$ and $S_T^\diamond \xrightarrow{a.s.} S_0^\diamond$. Also $A_T^* \xrightarrow{a.s.} A_0^* = I_n - \delta_0^*(a_0^* \delta_0^*)^{-1} a_0^*$ and $A_T^\diamond \xrightarrow{a.s.} A_0^\diamond = I_{n+k} - \delta_0^\diamond(a_0^\diamond \delta_0^\diamond)^{-1} a_0^\diamond$. The results concerning the asymptotic distributions of J^* and $J^\diamond$ follow from Hansen's Lemma 4.1.■

A.3 Proof of Theorem 2

Since D_0 has rank less than k, the function $h_0^*(b^*) = a_0^*[E(R^e) - D_0 b^*]$, defined in the proof of Theorem 1, does not have a unique zero. Instead any b^* such that $(D_0' W_0^* D_0)\, b^* = D_0' W_0^* D_0 b_0^*$ is a zero of $h_0^*(b^*)$. This means that $b_0^* + x$ is a zero for any x in the nullspace of $D_0' W_0^* D_0$– denoted $\mathcal{N}(D_0' W_0^* D_0)$—which is a non-empty set when $\text{rank}(D_0) < k$. So b^* is asymptotically unidentified.

As in the proof of Theorem 1, the last k rows of the function $h_0^\diamond(b^\diamond, \mu)$ have a unique zero, $\mu_z = \mu_0$. However, because d_0 has rank less than k, the first n rows of the function $h_0^\diamond(b^\diamond, \mu)$, which are $d_0' W_0^\diamond \left[E(R^e) - d_0 b^\diamond\right]$, do not have a unique zero. Instead any $b^\diamond$ such that $\left(d_0' W_0^\diamond d_0\right) b^\diamond = d_0' W_0^\diamond d_0 b_0^\diamond$ is a zero. This means that $b_0^\diamond + x$ is a zero for any $x \in \mathcal{N}(d_0' W_0^\diamond d_0)$, which is a non-empty set when $\text{rank}(d_0) < k$. So $b^\diamond$ is asymptotically unidentified.

The predicted expected returns from the A-normalization are $D_T \hat{b}^*$. Although $\hat{b}^*$ is not uniquely identified asymptotically, it lies almost surely in the set $B_0^* = \{b | b - b_0^* = x, x \in \mathcal{N}(D_0' W_0^* D_0)\}$. Since W_0^* is positive definite, any $x \in \mathcal{N}(D_0' W_0^* D_0)$ is in $\mathcal{N}(D_0)$. Therefore $D_T \hat{b}^* \xrightarrow{a.s.} D_0 b_0^* = \mu_R$. Therefore $R_*^2 \xrightarrow{a.s.} 1$. A similar result holds for $R_\diamond^2$.■

A.4 Proof of Theorem 3

As in the proof to Theorem 1, $a_T^* \xrightarrow{a.s.} a_0^* = D_0' W_0^*$. Because D_0 has full column rank and W_0^* is positive definite, the function $h_0^*(b^*)$ has a unique zero, $b_s^* = \left(D_0' W_0^* D_0\right)^{-1} D_0' W_0^* E(R^e)$. From Hansen's (1982) Theorem 2.1, $\hat{b}^* \xrightarrow{a.s.} b_s^*$. Of course, since the model is false, b_s^* does not have an interpretation as a "true" parameter value.

To get the expression for b_s^* in the statement of the theorem proceed as follows. Let x be the unique element of $\mathcal{N}(d_0)$ whose elements sum to 1 (all other elements of $\mathcal{N}(d_0)$ are proportional to x because d_0 has rank $k-1$). Let $X = \begin{pmatrix} X_1 & x \end{pmatrix}$ where X_1 is a $k \times (k-1)$ matrix whose columns span the rowspace of d_0, denoted $\mathcal{R}(d_0) = \mathcal{N}(d_0)^\perp$. The columns of X span R^k, by construction. Define $\tilde{b}_s^* = X^{-1} b_s^*$ and let $\tilde{b}_{s1}^*$ denote the first $k-1$ elements of $\tilde{b}_s^*$ and $\tilde{b}_{sk}^*$ denote the kth element of $\tilde{b}_s^*$. It follows that

$$\begin{aligned} E(R^e) - D_0 b_s^* &= E(R^e) - D_0 X X^{-1} b_s^* \\ &= E(R^e) - D_0 X \tilde{b}_s^* \\ &= E(R^e) - D_0 X_1 \tilde{b}_{s1}^* - D_0 x \tilde{b}_{sk}^* \\ &= E(R^e) - D_0 X_1 \tilde{b}_{s1}^* - \left[d_0 + E(R^e) E(f')\right] x \tilde{b}_{sk}^*. \end{aligned}$$

Since $x \in \mathcal{N}(d_0)$, $d_0 x = 0$, so we can write

$$E(R^e) - D_0 b_s^* = E(R^e) \left[1 - E(f') x \tilde{b}_{sk}^*\right] - D_0 X_1 \tilde{b}_{s1}^*.$$

This means we can set $E(R^e) - D_0 b_s^* = 0$ by choosing $\tilde{b}_{s1}^* = 0$ and $\tilde{b}_{sk}^* = 1/[E(f)' x]$. Since $b_s^* = X \tilde{b}_s^*$ it follows that $b_s^* = x/[x' \mu_0]$. By assumption $x' \mu_0$ cannot be zero, otherwise $\text{rank}[D_0] < k$ and we also know that at least one element of x is non-zero, so this means at least one element of b_s^* is non-zero. Since $E(R^e) = D_0 b_s^*$ we also have $R_*^2 \xrightarrow{a.s.} 1$.

As in the proof of Theorem 1, the last k rows of the function $h_0^\diamond(b^\diamond, \mu)$ have a unique zero, $\mu_z = \mu_0$. However, because d_0 has rank less than k, the first n rows of the function $h_0^\diamond(b^\diamond, \mu)$, which are $d_0' W_0^\diamond [E(R^e) - d_0 b^\diamond]$, do not have a unique zero. Instead any $b^\diamond$ such that $(d_0' W_0^\diamond d_0)\, b^\diamond = d_0' W_0^\diamond E(R^e)$ is a zero. Let $b_z^\diamond$ be a zero. This means that $b_z^\diamond + x$ is a zero for any x in the nullspace of $d_0' W_0^\diamond d_0$, which is a non-empty set because $\text{rank}(d_0) < k$. So $b^\diamond$ is asymptotically unidentified. Although there are arbitrarily many solutions to $d_0' W_0^\diamond [E(R^e) - d_0 b^\diamond] = 0$, in general, there is no solution to $E(R^e) - d_0 b^\diamond = 0$.■

A.5 Proof of Theorem 4

The matrix $\delta_0^* = D_0$ has full column rank. It follows from Hansen's Theorem 3.1 that $\hat{b}^*$ has the asymptotic distribution stated in the theorem.

From the results above it follows that $\delta_T^* \xrightarrow{a.s.} \delta_0^*$ and $A_T^* \xrightarrow{a.s.} A_0^* = I_n - \delta_0^*(a_0^* \delta_0^*)^{-1} a_0^*$, however, the matrix S_T^* will not generally be a consistent estimator for S_s^* because it imposes the restriction that $E(u_t^* u_{t-j}^{*\prime}) = 0$ for $j \neq 0$. This restriction only holds when the model is true. Instead $S_T^* \xrightarrow{a.s.} V_s^* = E[u_t^*(b_s^*) u_t^*(b_s^*)']$.

This means that $\hat{V}_g^* \xrightarrow{a.s.} V_g^* = A_0^* V_s^* A_0^{*\prime}$. We also know $\sqrt{T} g_T^*(\hat{b}^*) \xrightarrow{d} N(0, V_{g0}^*)$ where $V_{g0}^* = A_0^* S_s^* A_0^{*\prime}$. Diagonalize V_g^* as $V_g^* = P_g \Lambda_g P_g'$ where the columns of P_g are the orthonormal eigenvectors of V_g^* and Λ_g is a diagonal matrix with the eigenvalues of V_g^* on the diagonal. Diagonalize V_{g0}^* as $V_{g0}^* = P_0 \Lambda_0 P_0'$. Let $\tilde{\Lambda}_g = \Lambda_g^+$ and $\tilde{\Lambda}_0 = \Lambda_0^+$. These are diagonal matrices with zeros where Λ_g and Λ_0 have zeros, and whose non-zero elements are the inverses of the non-zero elements of Λ_g and Λ_0.

From these results it follows that $J^* \xrightarrow{d} Z' \Omega Z$, with $Z = \sqrt{T} \tilde{\Lambda}_0^{1/2} P_0' g_T^*(\hat{b}^*)$ and

$$\Omega = \Lambda_0^{1/2} P_0' P_g \tilde{\Lambda}_g P_g' P_0 \Lambda_0^{1/2}.$$

The vector Z converges in distribution to a vector of independent normal random variables, the first $n-k$ of which have unit variance and the last k of which have zero variance. The matrix Ω can be diagonalized as $\Omega = P_\Omega \Lambda_\Omega P_\Omega'$. When $V_s^* = S_s^*$ the first $n-k$ eigenvalues on the diagonal of Λ_Ω are ones while the rest are zeros. In this case $J \xrightarrow{d} \chi^2_{n-k}$. In general, however, $V_s^* \neq S_s^*$ and these eigenvalues will not be 1, so that $J^* \xrightarrow{d} \sum_{i=1}^{n-k} \lambda_{\Omega i} z_i^2$ where $\lambda_{\Omega 1}$, $\lambda_{\Omega 2}$, ..., $\lambda_{\Omega n-k}$ are the nonzero eigenvalues of Ω and z_1, z_2, ..., z_{n-k} are mutually independent standard normal random variables. Given the form of Ω, $\prod_{i=1}^{n-k} \lambda_{\Omega i} = \prod_{i=1}^{n-k} \lambda_{0i}/\lambda_{gi}$, however, in general, $\lambda_{\Omega i} \neq \lambda_{0i}/\lambda_{gi}$.■

A.6 Proof of Theorem 5

Let $R_t^e = \mu_R + u_t$ with $E(u_t u_t') = \Sigma_R$, and $f_t = \mu + \epsilon_t$ with $E(\epsilon_t^2) = \sigma_f^2$. The asymptotic distribution of $\hat{b}^\diamond$ depends on the asymptotic distribution of $d_T = \frac{1}{T}\sum_{t=1}^T R_t^e \left(f_t - \bar{f}\right)$. Scaling d_T by a factor of $T^{\frac{1}{2}}$ we have

$$T^{\frac{1}{2}} d_T = T^{-\frac{1}{2}} \sum_{t=1}^T u_t \epsilon_t - \sum_{t=1}^T \epsilon_t T^{-\frac{1}{2}} \sum_{t=1}^T u_t. \tag{A8}$$

The first expression on the right hand side of (A8) converges in distribution to $X \sim N(0, \sigma_f^2 \Sigma_u)$. The second expression converges in probability to 0. So $T^{\frac{1}{2}} d_T \xrightarrow{d} X$. Also, $\bar{R}^e \xrightarrow{p} \mu_R$.

At the first stage of GMM the weighting matrix is $W_T^\diamond = I_n$ so we have $T^{-\frac{1}{2}}\hat{b}^\diamond = T^{\frac{1}{2}} d_T' \bar{R}^e / (T^{\frac{1}{2}} d_T' T^{\frac{1}{2}} d_T)$. It follows that $T^{-\frac{1}{2}}\hat{b}^\diamond \xrightarrow{d} Z = (X'\mu_R)/(X'X)$. The t-statistic for $\hat{b}^\diamond$ is $t = \hat{b}^\diamond / \sqrt{V_b^\diamond}$ where $V_b^\diamond$ is the first element on the diagonal of

$$V_\theta^\diamond = (a_T^\diamond \delta_T^\diamond)^{-1} a_T^\diamond S_T^\diamond a_T^{\diamond\prime} (\delta_T^{\diamond\prime} a_T^\diamond)^{-1} / T, \tag{A9}$$

where $a_T^\diamond$ and $\delta_T^\diamond$ are defined in section 3.2 and $S_T^\diamond$ is a conventional estimate of the long-run covariance of the GMM errors in the first stage, which are

$$\begin{aligned} \hat{u}_{1t} &= R_t^e [1 - (f_t - \bar{f})\hat{b}^\diamond] \\ \hat{u}_{2t} &= f_t - \bar{f}. \end{aligned}$$

Considerable algebra shows that at the first stage of GMM $T^{-1} V_b^\diamond \xrightarrow{d} \sigma_f^2 Z^2 (X'\Sigma_R X)/(X'X)^2$. Hence $t \xrightarrow{d} Z / \sqrt{\sigma_f^2 Z^2 (X'\Sigma_R X)/(X'X)^2}$ or $(X'\mu_R)/[\sigma_f^2 Z^2 (X'\Sigma_R X)]^{\frac{1}{2}}$. We also have

$$R_\diamond^2 = 1 - \frac{(\bar{R}^e - d_T\hat{b}^\diamond)'(\bar{R}^e - d_T\hat{b}^\diamond)}{\left(\bar{R}^e - \iota\iota'\bar{R}^e/n\right)'\left(\bar{R}^e - \iota\iota'\bar{R}^e/n\right)} = 1 - \frac{\bar{R}^{e\prime} M_d \bar{R}^e}{\bar{R}^{e\prime} M_\iota \bar{R}^e}$$

where $M_d = I - d_T(d_T'd_T)^{-1}d_T'$ and $M_\iota = I_n - \iota\iota'/n$. So the R^2 is

$$R_\diamond^2 \xrightarrow{d} 1 - \frac{\mu_R' M \mu_R}{\mu_R' M_\iota \mu_R}$$

where $M = I_n - X(X'X)^{-1}X'$.

At the second stage of GMM the weighting matrix is $W_T^\diamond = (P_T S_T^\diamond P_T')^{-1}$ where $P_T = \begin{pmatrix} I_n & \bar{R}^e(\hat{b}^\diamond)' \end{pmatrix}$. Considerable algebra shows that $TW_T^\diamond \xrightarrow{d} W = \Sigma_R^{-1}/(\sigma_f^2 Z^2)$. We have $T^{-\frac{1}{2}}\hat{b}^\diamond = T^{\frac{1}{2}} d_T'(TW_T)\bar{R}^e / (T^{\frac{1}{2}} d_T'(TW_T) T^{\frac{1}{2}} d_T)$. It follows that $T^{-\frac{1}{2}}\hat{b}^\diamond \xrightarrow{d} \tilde{Z} = (X'W\mu_R)/(X'WX) =$

$(X'\Sigma_R^{-1}\mu_R)/(X'\Sigma_R^{-1}X)$. The t-statistic for $\hat{b}^\diamond$ is $t = \hat{b}^\diamond/\sqrt{V_b^\diamond}$ where $V_b^\diamond$ is again the first element on the diagonal of $V_\theta^\diamond$, and $V_\theta^\diamond$ is given by (A9). In this case, however, the matrix $a_T^\diamond$ depends on the weighting matrix and takes a form such that $T^{-1}V_b^\diamond \xrightarrow{d} \sigma_f^2 Z^2(X'W\Sigma_R WX)/(X'WX)^2$ or $\sigma_f^2 Z^2/(X'\Sigma_R^{-1}X)$. Hence $t \xrightarrow{d} \tilde{Z}/[\sigma_f^2 Z^2/(X'\Sigma_R^{-1}X)]^{\frac{1}{2}}$ or $(X'\Sigma_R^{-1}\mu_R)/[\sigma_f^2 Z^2(X'\Sigma_R^{-1}X)]^{\frac{1}{2}}$. We also have

$$R_\diamond^2 = 1 - \frac{\bar{R}^{e\prime}\tilde{M}_d'\tilde{M}_d\bar{R}^e}{\bar{R}^{e\prime}M_\iota\bar{R}^e}$$

where $\tilde{M}_d = I - d_T(d_T'W_T d_T)^{-1}d_T'W_T$. So the R^2 is

$$R_\diamond^2 \xrightarrow{d} 1 - \frac{\mu_R'\tilde{M}'\tilde{M}\mu_R}{\mu_R'M_\iota\mu_R}$$

where $\tilde{M} = I_n - X(X'WX)^{-1}X'W = I_n - X(X'\Sigma_R^{-1}X)^{-1}X'\Sigma_R^{-1}$. The test statistic for the over-identifying restrictions is $J = T(\bar{R}^e - d_T\hat{b})'W_T(\bar{R}^e - d_T\hat{b}) = T\bar{R}^{e\prime}\tilde{M}_d'W_T\tilde{M}_d\bar{R}^e$. Hence $J \xrightarrow{d} \mu_R'\tilde{M}'W\tilde{M}\mu_R = \mu_R'\tilde{M}'\Sigma_u^{-1}\tilde{M}\mu_R/(\sigma_f^2 Z^2)$.■

A.7 Proof of Theorem 6

As in the proof to Theorem 3, $\hat{b}^* \xrightarrow{a.s.} b_s^* = (D_0'W_0^*D_0)^{-1}D_0'W_0^*E(R^e)$. Given the assumption that there exists a unique $k\times 1$ vector x whose elements sum to 1, such that $\text{cov}(R_t^e, f_t'x) = \iota c$ it follows that $D_0 x = \iota c + E(R^e)\mu' x$. Hence

$$E(R^e) = \frac{D_0 x - \iota c}{\mu' x}.$$

and

$$b_s^* = \frac{1}{\mu' x}(x - \beta_{\iota D}c).$$

and

$$\beta_{\iota D} = (D_0'W_0^*D_0)^{-1}(D_0'W_0^*\iota).$$

The predicted expected returns are $D_T\hat{b}^* \xrightarrow{a.s.} D_0 b_s^*$ and

$$D_0 b_s^* = E(R^e) + (\iota - D_0\beta_{\iota D})\frac{c}{\mu' x}$$

Notice that $\lim_{c\to 0} b_s^* = x/(\mu' x)$ and $\lim_{c\to 0} D_0 b_s^* = E(R^e)$.■

A.8 Estimating Long-Run Covariance Matrices

A.8.1 The A-Normalization

As stated in section 4, I define $S_T^* = \frac{1}{T}\sum_{t=1}^T \hat{u}_t^*\hat{u}_t^{*\prime}$ when estimating the standard errors of $\hat{b}^*$ and testing the over-identifying restrictions of the model. This is a consistent estimate of S_0^* when the model is true because $E[u_t^*(b_0^*)u_{t-j}^*(b_0^*)'] = 0$ for $j \neq 0$.

When conducting inference about the price of risk we need an estimate of $\tilde{S}_0^*$. Since $\hat{u}_t^*(\mu_0, \varsigma_{f0})$ is not necessarily orthogonal to lagged information the simple covariance matrix $\frac{1}{T}\sum_{t=1}^{T} \tilde{u}_t^*(\tilde{\theta}_T^\diamond)\tilde{u}_t^*(\tilde{\theta}_T^\diamond)'$ will, in general, be an inconsistent estimator of $\tilde{S}_0^*$. For this reason I use den Haan and Levin's (2000) VARHAC procedure for estimating $\tilde{S}_0^*$. In doing so I impose the restriction that lagged variables do not appear in the equations for u_t^* but allow for lags in the equations for $\hat{u}_t^*$.

A.8.2 The M-Normalization

As stated in section 3, to compute $S_T^\diamond$ I use the same VARHAC procedure described above. In doing so I impose the restriction that lagged variables do not appear in the equations for $u_{1t}^\diamond$ (the errors corresponding to the asset pricing conditions) but allow for lags in the equations for $u_{2t}^{\diamond *}$ (the errors corresponding to $f_t - \mu$).

B Data Construction

B.1 FF25 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.

B.2 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.[9] 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

$$k_{t+1} = c_{t+1}^D + (1 - \delta)k_t \tag{A10}$$

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

$$k_{t+4} = c_{t+4}^D + (1 - \delta)c_{t+3}^D + (1 - \delta)^2 c_{t+2}^D + (1 - \delta)^3 c_{t+1}^D + (1 - \delta)^4 k_t. \tag{A11}$$

Once I identify the value of δ that applies within a year using (A11), I use (A10) 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.

B.3 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 (*RM-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 real quarterly versions of the Fama-French factors (*RM-Rf*, *SMB* and *HML*) I proceed as follows. To get quarterly excess returns I compound the monthly series geomet-

[9] I pass the NIPA population series through the Census X11 seasonal adjustment procedure because the NIPA series displays noticeable seasonal fluctuations.

rically within each quarter. I convert them to real excess returns by dividing the resulting series by $1 + \pi_t$.

B.4 Yogo Factors

Yogo (2006) proposes a three factor model that uses the two factors from the Durables CCAPM as well as the market return, *RM*, as factors. I use the data for the series as constructed above, to study the rank condition for Yogo's model using his sample period (1951:Q1–2001Q4), while also using Campbell's (2003) timing for consumption growth (that is, assuming that quarter q returns and quarter $q + 1$ consumption are determined simultaneously).

B.5 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 take the factor data directly from the authors over the sample period 1963Q3–1998Q3.

B.6 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. I examine the rank condition over the period 1954–2003 as in Jagannathan and Wang (2007).

B.7 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's (2006) model]. I take the data for the returns and factors directly from their paper.

B.8 Sorted Currency Portfolios

I construct a set of currency portfolios over the period 1976–2008. I compute the monthly payoff to taking a long position in foreign currency as

$$x_{t+1} = \frac{S_{t+1} - F_t}{F_t}$$

where S_t is the spot exchange rate measured as USD per foreign currency unit (FCU) and F_t is the one month forward exchange rate in the same units. I compute these payoffs for up to 19 currencies on a monthly basis: Australia, Austria, Belgium, Canada, Denmark, France, Germany, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the UK, and the U.S. The data source is Datastream, and I used Reuters/WMR quotes when they are available. I also use BBI quotes for Australia and New Zealand in the 1984–1996 period.

In each month I sort the available currencies into five portfolios based on the size of the forward discount $(S_t - F_t)/F_t$. Countries with large values of the forward discount are countries with high interest rates. I study the rank conditions for quarterly averages of the payoffs.

C Optimal Iterated GMM and CU-GMM

C.1 Optimal Iterated GMM and the M-normalization

Each of the variants of the M-normalization that I have described above sets the GMM estimator up in such a way that $\hat{\mu} = \bar{f}$. Consider the benchmark case where the GMM estimator sets $a_T^\diamond g_T^\diamond = 0$ where $g_T^\diamond = (\ g_{1T}^{\diamond\prime} \quad g_{2T}^{\diamond\prime}\ )'$, $g_{1T}^\diamond = \bar{R}^e - \left(D_T - \bar{R}^e\mu'\right) b^\diamond$, $g_{2T}^\diamond = \bar{f} - \mu$ and $a_T^\diamond$ is given by (15). A more traditional approach to GMM might, instead, define

$$a_T^\diamond = \begin{pmatrix} (D_T - \bar{R}\mu') & -\bar{R}^e b^{\diamond\prime} \\ 0 & I_k \end{pmatrix} W_T^\diamond \tag{A12}$$

where $W_T^\diamond$ would now be an $(n+k)\times(n+k)$ weighting matrix. With the $a_T^\diamond$ given by (A12), the equation $a_T^\diamond g_T^\diamond = 0$ is the first order condition corresponding to $\min_{b^\diamond,\mu} g_T^{\diamond\prime} W_T^\diamond g_T^\diamond$. It is clear that in this setup, μ is free to help match not only $E(f_t-\mu)=0$ but also the asset pricing equations. Under the null, this is asymptotically more efficient than the other approaches because it uses information about μ that lies in the asset pricing restrictions.

Yogo (2006) uses this approach. In the first stage of GMM he sets

$$W_T^\diamond = \begin{pmatrix} \kappa I_n & 0 \\ 0 & \mathrm{S}_f^{-1} \end{pmatrix}$$

and $\kappa = \det(\mathrm{S}_{R^e})^{-1/n}$. Here $\mathrm{S}_{R^e} = \frac{1}{T}\sum_{t=1}^T (R_t^e - \bar{R}^e)(R_t^e-\bar{R}^e)'$. As $\kappa\to 0$, $\hat{\mu}\to\bar{f}$, whereas, as $\kappa\to\infty$, $\hat{\mu}$ is determined solely by the asset pricing conditions. In the second stage of GMM, the inverse of a consistent estimate of $S_0^\diamond$ is used as the weighting matrix.

As the sample size grows, $a_T^\diamond$ converges to

$$a_0^\diamond = \begin{pmatrix} d_0 + \mu_R(\mu_0-\mu)' & -\mu_R b^{\diamond\prime} \\ 0 & I_k \end{pmatrix} \begin{pmatrix} \kappa_0 I_n & 0 \\ 0 & \Sigma_f^{-1} \end{pmatrix} \tag{A13}$$

where $\kappa_0 = \det(\Sigma_{R^e})^{-1/n}$. When the rank condition for the M-normalization fails, $(\mathrm{rank}[d_0] < k)$, this means identification fails here at $\mu=\mu_0$.

The objective function, $g_T^{\diamond\prime} W_T^\diamond g_T^\diamond$, is ill-conditioned asymptotically. This can most easily be demonstrated for the first stage of GMM and a single factor model. In this case $d_0=0$ and the objective function limits to the function

$$Q_0(b^\diamond,\mu) = \theta\left[1+(\mu-\mu_0)b^\diamond\right]^2 + (\mu-\mu_0)^2/\sigma_f^2,$$

where $\theta = \kappa\mu_R'\mu_R$. This function has no well-defined minimum. However, along the locus $b^\diamond = 1/(\mu_0-\mu_0)$, which is illustrated in Figure A1, the limit of $Q_0(b^\diamond,\mu)$ as $\mu\to\mu_0$ is 0. However, the function does not achieve the infimum for any value of $(b^\diamond,\mu)$. At $\mu=\mu_0$, for example, the function equals θ.

C.2 Continuously Updated GMM and the A-normalization

Here I consider the case of a single factor model, where the proposed factor is spurious. In this case $d_0=0$, but $D_0=\mu_R\mu_0\neq 0$. What happens to the CU-GMM estimator in this circumstance? The estimator solves

$$\min_{b^*} g_T(b^*)'\Omega_T(b^*)^{-1} g_T(b^*)$$

where $g_T(b^*) = \bar{R}^e - D_T b^*$ and $\Omega_T(b^*)$ is a HAC covariance matrix associated with the GMM errors $u_t = R_t^e(1 - f_t b^*)$. Asymtotically, $g_T(b^*) \to g_0(b^*) = E(R_t^e)(1 - \mu_0 b^*)$. If we assume that R_t^e and f_t are not only uncorrelated, but are independent of each other, and if we assume that $\Omega_T(b^*)$ is computed as the simple covariance matrix of u_t (so that the researcher is implicitly assuming the GMM errors are unpredictable) then,

$$\Omega_T(b^*) \xrightarrow{p} \Omega_0(b^*) = M_0(1 - \mu_0 b^*)^2 + M_0 \sigma_0^2 (b^*)^2$$

where $M_0 = E(R_t^e R_t^{e\prime})$ and $\sigma_0^2 = \text{var}(f_t)$. Notice that $g_0(1/\mu_0) = 0$ and $\Omega_0(1/\mu_0) = M_0\sigma_0^2/\mu_0^2$. Clearly, then, the CU-GMM estimator $\hat{b}^* \xrightarrow{p} 1/\mu_0$. This indicates that the uncentered CU-GMM estimator has exactly the same issue as the regular GMM estimator. A naive researcher will, asymptotically, think that his model holds because the moment conditions will hold exactly, the R^2 will be 1 and the estimate of the factor loading in the SDF will be non-zero and statistically significant. However, the model is false.

C.3 Continuously Updated GMM and the M-normalization

The CU-GMM objective function for the M-normalization is $Q(b^\diamond, \mu) = h_T(b^\diamond, \mu)' \Gamma_T(b^\diamond, \mu)^{-1} h_T(b^\diamond, \mu)$ where

$$h_T(b^\diamond, \mu) = \begin{pmatrix} \bar{R}^e[1 + (\bar{f} - \mu)b^\diamond] - d_T b^\diamond \\ \bar{f} - \mu \end{pmatrix}$$

and $\Gamma_T(b^\diamond, \mu)$ is a HAC covariance matrix associated with the GMM errors

$$v_t = \begin{pmatrix} R_t^e \left[1 - (f_t - \mu)\, b^\diamond\right] \\ f_t - \mu \end{pmatrix}.$$

I assume that R_t^e and f_t are not only uncorrelated, but are independent of each other, and I assume that $\Gamma_T(b^\diamond, \mu)$ is computed as the simple covariance matrix of v_t.

As in the case of optimal GMM, consider the locus $b^\diamond = 1/(\mu_0 - \mu)$, shown in Figure A1. Along this locus

$$h_T \xrightarrow{p} h_0 = \begin{pmatrix} 0 \\ \mu_0 - \mu \end{pmatrix}$$

and

$$\Gamma_T \xrightarrow{p} \Gamma_0 = \sigma_0^2 \begin{pmatrix} \frac{M_0}{(\mu_0 - \mu)^2} & \frac{\mu_R}{\mu_0 - \mu} \\ \frac{\mu_R'}{\mu_0 - \mu} & 1 \end{pmatrix}.$$

Consequently, along this locus the GMM objective function limits to

$$\frac{(\mu_0 - \mu)^2}{(1 - \mu_R' M_0^{-1} \mu_R)\sigma_0^2}.$$

So the CU-GMM objective function limits to zero as μ approaches μ_0 from above or below along the locus $b^\diamond = 1/(\mu_0 - \mu)$.

If we consider the case where $\mu = \mu_0$, then

$$h_T \xrightarrow{p} h_0 = \begin{pmatrix} \mu_R \\ 0 \end{pmatrix}$$

and

$$\Gamma_T \xrightarrow{p} \Gamma_0 = \begin{pmatrix} M_0[1 + \sigma_0^2 (b^\diamond)^2] & -\mu_R \sigma_0^2 b^\diamond \\ -\mu_R' \sigma_0^2 b^\diamond & \sigma_0^2 \end{pmatrix}.$$

Consequently, along this locus the GMM objective function limits to

$$\mu_R' [M_0 + \Sigma_R \sigma_0^2 (b^\diamond)^2]^{-1} \mu_R.$$

Clearly there is no finite value of $b^\diamond$ for which this is zero, so the GMM objective function does not obtain its infimum (zero) at any point inside the parameter space.

Figure A1, and the above discussion, illustrates that in the limit, the objective function for the M-normalization is near zero when $\mu \approx \mu_0$ and $b^\diamond \approx \pm\infty$. In finite samples, this manifests itself in an identification problem in that the CU-GMM objective function is low in the neighborhood of $\mu \approx \mu_0$ for very large positive and negative values of $b^\diamond$. Thus, confidence regions for $b^\diamond$ are typically disjoint subsets of the parameter space.

FIGURE A1

The M-Normalization and Optimal GMM