Agnostic Tests of Stochastic Discount Factor Theory

By

Kuntara Pukthuanthong and Richard Roll

March 6, 2015

Abstract

We propose and implement tests for the existence of a common stochastic discount factor (SDF). Our tests are agnostic because they do not require macroeconomic data or preference assumptions; they depend only on observed asset returns. After examining test features and power with simulations, we apply the tests empirically to data on U.S. equities, bonds, currencies, commodities and real estate. The empirical evidence is consistent with a unique positive SDF that prices all assets and satisfies the Hansen/Jagannathan bounds.

Authors’ Coordinates

<table>
<thead>
<tr>
<th></th>
<th>Roll</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pukthuanthong¹</td>
<td>Department of Finance</td>
</tr>
<tr>
<td></td>
<td>University of Missouri</td>
</tr>
<tr>
<td></td>
<td>Columbia, MO 65211</td>
</tr>
<tr>
<td></td>
<td>California Institute of Technology</td>
</tr>
<tr>
<td></td>
<td>Pasadena, CA 91125</td>
</tr>
<tr>
<td>573-884-9785</td>
<td>626-395-3890</td>
</tr>
<tr>
<td><a href="mailto:pukthuanthongk@missouri.edu">pukthuanthongk@missouri.edu</a></td>
<td><a href="mailto:rroll@caltech.edu">rroll@caltech.edu</a></td>
</tr>
</tbody>
</table>

¹ Corresponding author
The Stochastic Discount Factor (SDF) has become a dominant paradigm in recent asset pricing research. For example, Cochrane (2001) begins with the basic SDF relation in chapter 1 and expands it into almost all other known models of assets. Exactly the same foundation is established in the first chapter of Singleton (2006) and exploited to study asset price dynamics. Campbell (2014) ordains the SDF as “The Framework of Contemporary Finance,” (p. 3.) in his essay explaining the 2013 Nobel Prizes awarded to Fama, Hansen, and Shiller.

The empirical success of SDF theory is less apparent. In many previous empirical applications, the SDF is proxied by a measured construct that depends usually on aggregate consumption, but occasionally on some other macroeconomic quantity, combined with a risk aversion parameter. Cochrane (1996) employs aggregate consumption changes along with power utility (and a particular level risk aversion) to measure the SDF. Despite giving this specification every empirical benefit of the doubt, Cochrane (2001, p. 45) admits that it still “…does not do well.” A similar imperfect fit between consumption changes, over various horizons, and both equities and bonds, is reported by Singleton (1990.)

Lettau and Ludvigson (2000) add in macro variables such as labor income and find the deviation in wealth from its shared trend with consumption and labor income has strong predictive power for excess stock returns at business cycle frequencies, thereby suggesting that risk premia vary countercyclically. Chapman (1997) adds technology shocks and a battery of conditioning variables, transforming them with orthogonal polynomials, which serve to eliminate the small firm effect but still produce “statistically and economically large pricing errors”, (p. 1406.)

In research published prior to the hegemony of the SDF paradigm, Long (1990) shows that a “Numeraire” portfolio has many similar properties. Long’s Numeraire portfolio \( \eta \) has strictly positive gross returns \( (1+R_\eta) \) and exists only if there is no arbitrage within a list of assets from which it is composed. In this case, the expected value of the ratio \( (1+R_j)/(1+R_\eta) \) is unity for all assets \( j \) on the list, which implies that \( 1/(1+R_\eta) \) is essentially the same as the modern SDF. Long notes that the Numeraire portfolio is also the growth optimum portfolio. The latter is examined by Roll (1973) who provides an empirical test of whether the expected ratio above is the same for all assets. (He does not find evidence against it.)
Excellent reviews are provided by Ferson (1995) and Cochrane and Culp (2003).

Recognizing that aggregate consumption changes are too “smooth” to be very correlated with asset prices, (Mehra and Prescott (1985)), and that consumption is likely measured with significant error, (Rosenberg and Engle (2002)), recent literature avoids aggregate consumption data. In addition to Rosenberg and Engle, such an approach is taken by Aït-Sahalia and Lo (1998, 2000), and Chen and Ludvigson (2009). However, as pointed out by Araujo, Issler, and Fernandes (2005) and Araujo and Issler (2011), the above scholars still find it necessary to impose what might be considered rather ad hoc restrictions on preferences.

Hansen and Jagannathan (1991) avoid the specification of preferences and are still able to develop their famous bound on the mean and volatility of the SDF, given that SDF theory is true. Campbell (1993) surmounts the annoyance with various approximations of nonlinear multiperiod consumption and portfolio-choices. He develops a formula for risk premia that can be tested without using consumption data and suggests a new way to use imperfect data about both market returns and consumption.

Araujo, Issler and Fernandes (hereafter AIF), get around these difficulties by noting that the SDF should be the only serial correlation common feature of the data in the sense of Engle and Kozicki (1993). Then, by exploiting a log transform of returns, they derive a measure of the SDF that does not depend on a macroeconomic variable (notably including the problematic aggregate consumption) and also avoids the imposition of preferences.

Araujo and Issler (hereafter AI) take a similar tack, noting via a logarithmic series expansion that the natural logarithm of the SDF is the only common factor in the log of all returns. Thus, the log SDF can be eliminated by a simple difference in returns. Essentially, the log SDF represents the (single) common APT factor in the sense of Ross (1976).

In both AIF and AI, the SDF measure is a function of average arithmetic and geometric asset returns. AIF compute their measure empirically and report its temporal evolution along with various statistical properties. They also compare it to the time series of riskless returns. AI find that relative low risk aversion parameters are consistent with their estimated SDFs. They also are able to price some stocks successfully, but not stocks with low capitalization levels.
Both AIF and AI essentially assume that the SDF theory is true, rejecting it only indirectly in the case of AI with low cap stocks. Our primary goal is to develop tests that offer an opportunity to directly reject the SDF theory. Our SDF estimator, which we exploit to develop such tests, does not depend on a factor model or a logarithmic approximation, or any other structural assumption.

In the next section, we derive an estimator of the SDF that depends only on observed returns for a sample of assets and hence is “agnostic” with respect to both macroeconomic quantities and preferences. By collecting samples of different assets observed over the same time period and estimating SDFs for each collection, it becomes possible to test the theory’s main prediction: a unique SDF prices all assets in complete markets. Like many tests, this one involves a joint hypothesis, complete markets plus the SDF pricing equation. We can also examine whether the SDF is positive, which implies the absence of arbitrage. This too involves a joint hypothesis. If both hypotheses are rejected, then markets are incomplete and there are arbitrage opportunities, or else there’s something wrong with SDF theory itself.2

I. An Agnostic Test for the SDF

This section first shows (in sub-section I.A) how SDFs can be approximated by a transformation of returns, without any additional information about preferences, consumption or other macroeconomic data. The following sub-section (I.B) proves that the same SDF estimator arises naturally from minimizing a particular sum of average surprises. This development allows us to infer some useful properties of the SDF estimator. The next sub-section (I.C) proposes a battery of tests of SDF theory using the SDF estimator derived in I.A and I.B.

I.A. Estimating the SDF from Returns Alone

Let \( p_{i,t} \) denote the cash value of asset \( i \) at time \( t \). When markets are complete, SDF theory implies the existence of a unique \( m_i \), such that

---

2 Kan and Zhou (1999) argue that the SDF theory is more subject to estimation error than other approaches and that it has inherently weak power in empirical tests. Cochrane (2001) offers a rebuttal.
Denoting a gross return between \( t-1 \) and \( t \) by \( R_{i,t} \equiv \frac{p_{i,t}}{p_{i,t-1}} \), equation (1) is the same as

\[
E_{t-1}(\tilde{m}_t \tilde{R}_{i,t}) = 1 \quad \forall i,t.
\]

Corresponding to the expectation in (2), there is a realization; i.e.,

\[
m_t R_{i,t} = E_{t-1}(\tilde{m}_t \tilde{R}_{i,t}) + \epsilon_{i,t}
\]

where \( \epsilon_{i,t} \) denotes the (complete) surprise in the mR product for asset \( i \) in period \( t \). This suggests a sample analog to equation (2) since, for each asset \( i \), over a sample of size \( T \), we have, from (3) and (2),

\[
\frac{1}{T} \sum_{t=1}^{T} m_t R_{i,t} = \frac{1}{T} \sum_{t=1}^{T} \left[ E_{t-1}(\tilde{m}_t \tilde{R}_{i,t}) \right] + \frac{1}{T} \sum_{t=1}^{T} \epsilon_{i,t} = 1 + \frac{1}{T} \sum_{t=1}^{T} \epsilon_{i,t} \equiv 1.
\]

where the approximation indicates that the average surprise is not exactly zero in a finite sample, though is should vanish as \( T \to \infty \).

The approximation error in (4) equals the time series sample mean of the surprises in the SDF-gross return product, a mean for asset \( i \) which we hereafter denote

\[
\bar{\epsilon}_i = \frac{1}{T} \sum_{t=1}^{T} \epsilon_{i,t}.
\]

Rational expectations rules out any serial dependence in the surprises,

\[
\text{Cov}(\epsilon_{i,t}, \epsilon_{i,t-j}) = 0, \quad j \neq 0
\]

but the surprises could be heteroscedastic. Hence,

\[
\text{Var}(\bar{\epsilon}_i) = \frac{1}{T^2} \sum_{t=1}^{T} \text{Var}(\epsilon_{i,t}) = \frac{1}{T} \sigma_i^2
\]

\(^3\) For a representative agent, \( m \) is the discounted future marginal utility of consumption divided by the current marginal utility of consumption. The tilde denotes a random variable as of period \( t-1 \).
where $\tilde{\sigma}_i^2$ denotes the mean variance of surprises for asset $i$ over the particular sample period, $t=1,...,T$. Unless the mean variance is growing without bound, the approximation error should disappear as $T$ grows larger.

Now consider a sample of $N$ assets with simultaneous observations over $T$ periods, with $N > T$. The ensemble of gross returns for the $N$ assets can be expressed as a matrix $R$ (hereafter boldface denotes a matrix or vector). There are $N$ rows in $R$ and the $i^{th}$ row is $[R_{i,1:...:R_{i,T}}]$. We also need a column vector $m \equiv [m_1:...:m_T]'$ to hold the $T$ values of the SDF and a $N$-element column unit vector $1 \equiv [1:...:1]'$. The entire SDF sample analog approximation for all asset and periods can be written compactly as

$$Rm/T \equiv 1. \quad (5)$$

Pre-multiply (5) by the transpose of $R$, to obtain

$$(R'R)m/T \equiv R'1.$$ 

Since we have chosen $N > T$, the cross-sectional time-product matrix $R'R$ is non-singular unless there are two periods with linearly dependent cross-sectional vectors of returns.\(^4\) Hence, we can usually solve for the sample vector of time-varying stochastic discount factors as\(^5\)

$$m/T \equiv (R'R)^{-1}R'1. \quad (6)$$

Collecting the individual asset sample mean surprises in a column $N$ vector, $\bar{\epsilon} = (\bar{\epsilon}_1:...:\bar{\epsilon}_N)'$, the approximation error in (6) is equal to

$$(R'R)^{-1}R'\bar{\epsilon}.$$ 

This error is not exactly zero because, for each $t$, there are related components in $R$ and $\bar{\epsilon}$. For very large $N$ and $T$, these components should become immaterial, but they add sampling error to the estimated SDFs with smaller $N$ and $T$.

\(^4\) That is, unless the return of every individual asset in a given period is a linear function of the return on that asset in another period, (not that the returns are linearly dependent relative to each other in a given period.)

\(^5\) There is an expression similar to (6) in Hansen and Jagannathan (1991), but it involves an expectation; they do not write down a sample analog.
I.B. The Minimum Sum of Squared Average Surprises

The exact form of equation (5), (i.e., with no approximation), is

\[ \frac{Rm}{T} = (1 + \bar{\epsilon}) \]  

(8)

where \( \bar{\epsilon} \) is the column \( N \) vector that contains the average surprises for each asset. A least squares estimator for \( m \) is available by minimizing the sum of squared average surprises with respect to \( m \); i.e.,

\[ \min_m [ (\bar{\epsilon}'\bar{\epsilon}) = (Rm / T - 1)'(Rm / T - 1) ] . \]

The first-order condition is

\[ \frac{\partial}{\partial m} \left( m'R'Rm / T^2 - 2m'R1/T \right) = 2R'Rm / T^2 - 2R'1/T = 0 \]

and the extremum is achieved for the \( \hat{m} \) that satisfies

\[ \hat{m} / T = (R'R)^{-1}R'1 \]  

(9)

which shows that \( \hat{m} \) is the approximation (6) in section I.A. The second order condition is strictly positive because \( R'R \) is positive definite (by assumption); hence \( \hat{m} \) provides the minimum sum of squares for the average SDF surprises.

The least squares estimator in (9) differs from a standard regression estimator in one important respect; since the “dependent” variable here is the \( T \) element unit vector, (with every element a constant 1.0), there could be a connection between \( R \) and \( \bar{\epsilon} \), which would violate the customary spherical regression assumptions. Consequently, the estimator could be biased. There is indeed a linear connection between the \( mR \) product and \( \bar{\epsilon} \) but this is slightly different than the source of typical regression bias induced by linear dependence of the disturbances and explanatory variables.

To elucidate this issue, solve (8) for \( 1 \) and substitute the result in (9), which simplifies to,
\[ \hat{m} - m = - T (R' R)^{-1} R' \tilde{\epsilon}. \]

The expected value of this expression is the bias. Expanding \( R' \tilde{\epsilon} \), term by term, we observe that most elements are innocuous and close to zero because they involve products such as \((\epsilon_{i,t} R_{i,t,k})\) for \(i \neq j\) and \(k \neq 0\). However, there are a few elements that are unlikely to disappear. For period \(t\), there is

\[ R_{1,t} \epsilon_{1,t} + R_{2,t} \epsilon_{2,t} + \ldots + R_{N,t} \epsilon_{N,t} = m_t (R_{1,t}^2 + R_{2,t}^2 + \ldots + R_{N,t}^2) - \sum_{j=1}^{N} R_{j,t} \]

and there are similar terms for other periods. We will study the extent of the resulting bias in the next section using simulation but note already that the bias terms are atypical because the dependence between the explanatory variables (the \(R\)'s) and the disturbances (the \(\epsilon\)'s) is not linear.

Despite its possible bias, the estimator in (9) shares some attractive features with OLS regression estimates. In particular, it can be used to define residuals, estimates of the true disturbances, as

\[ \hat{\epsilon} = R (R' R)^{-1} R' I = - [I - R (R' R)^{-1} R'] I. \]

The matrix in brackets on the right side of (10) is idempotent, so the sum of squared residuals divided by the degrees-of-freedom, \(N-T\), is

\[ \frac{\hat{\epsilon}' \hat{\epsilon}}{N-T} = \frac{1 [I - R (R' R)^{-1} R'] I}{N-T} = \frac{T}{N-T} \left( \frac{N}{N-T} \right), \]

(11)

For a large enough \(N\), (specifically, \(N>2T\)), the mean squared residual in (11) declines with \(N\), holding \(T\) constant.\(^6\) Consequently, the quality of our SDF estimator should be better when \(N\) is

\(^6\) Proof: The second term on the right side of (11) can be written as \( \frac{N^2}{N-T} \bar{R} (R' R)^{-1} \bar{R} = \frac{N^2}{N-T} \Psi \) where \( \bar{R} \) is the \(T\) element column vector whose \(t\)th element is the cross-sectional mean gross return in period \(t\). The positive quadratic form \( \Psi \) does not depend directly on \(N\), so \( \frac{\partial}{\partial N} \left[ \frac{N^2}{N-T} \Psi \right] = \Psi \left[ 1 - \left( \frac{T}{N-T} \right)^2 \right] \), which is positive for \(N>2T\), at which point both terms in (11) decline with \(N\); QED.
large relative to $T$; i.e., when there are at least twice as many assets as time periods. The square root of (11) gives the standard error of the estimate,

$$s = \sqrt{\hat{\varepsilon}'\hat{\varepsilon}/(N - T)}.$$

The covariance matrix of the estimated SDFs is given by

$$E[(\hat{\mathbf{m}} - \mathbf{m})(\hat{\mathbf{m}} - \mathbf{m})'|R, T] = (R'R)^{-1}R'E(\mathbf{V}_{\varepsilon}^2)R(R'R)^{-1}$$  \hspace{1cm} (12)

where the $(N \times N)$ symmetric matrix $\mathbf{V}_{\varepsilon}^2$ has the following element in the $j^{th}$ row and $k^{th}$ column:

$$(\varepsilon_{j,1} + \varepsilon_{j,2} + \ldots + \varepsilon_{j,T})(\varepsilon_{k,1} + \varepsilon_{k,2} + \ldots + \varepsilon_{k,T}).$$

Unlike the analogous covariance matrix of disturbances in standard OLS regressions, the diagonal elements of $\mathbf{V}_{\varepsilon}^2$ are not necessarily equal to each other and the off-diagonal elements need not have zero expectation. However, we can safely assume that cross-products separated in time, such as $\varepsilon_{j,t}\varepsilon_{k,\tau}$ for $t \neq \tau$, are zero; otherwise, the $\varepsilon$'s would not be surprises. This implies that the element in the $j^{th}$ row and $k^{th}$ column of $\mathbf{V}_{\varepsilon}^2$ reduces to $\sum_{t=1}^{T} \varepsilon_{j,t}\varepsilon_{k,t}$. Moreover, if the $\varepsilon$'s are not correlated across assets, an arguably dubious condition, this sum has an expected value of zero for $j \neq k$ and then $E(\mathbf{V}_{\varepsilon}^2)$ becomes diagonal and equal to $I\sigma^2_{\varepsilon}$, where $I$ is the identity matrix and $\mathbf{\sigma}^2_{\varepsilon}$ is the $N$ element column vector whose $j^{th}$ element is $\text{Var}(\sum_{t=1}^{T} \varepsilon_{j,t})$. If the variance of the surprises were the same scalar $\sigma^2$ for all assets and time periods, perhaps an even more dubious condition, then (12) simplifies further to

$$E[(\hat{\mathbf{m}} - \mathbf{m})(\hat{\mathbf{m}} - \mathbf{m})'|R, T] = T\sigma^2(R'R)^{-1}.$$  \hspace{1cm} (13)

Except for the presence of $T$, this is the standard regression covariance matrix of the coefficients given IID disturbances.
The square roots of the T diagonal elements of (12) or (13) provide the standard errors of the SDFs period-by-period. We will examine their properties using simulation in the next section. One pertinent property is obvious already, however. For a fixed number of assets, N, the standard errors of estimated SDFs increase with the time series sample size, T. Thus, we anticipate that our estimator will perform better when N-T is large.

I. C. Testing the SDF Theory

The vector on the right side of (9) is an estimate based on N assets and a sample period t = 1,...,T. But the SDF theory stipulates that any other set of assets should produce the same \( \hat{m} \) from the same time series observations. Hence, if we denote by \( \hat{m}(k) \) a sample \( \hat{m} \) computed according to (9) (where k indicates a set of K assets) and then, from the same calendar observations, choose a complement set \( j \subset k \) with J assets (and \( J > T \)), the SDF null hypothesis can be expressed as

\[
H_0: E[\hat{m}(k) - \hat{m}(j)] = 0. \tag{14}
\]

Any standard test of equality could be employed for equation (14). For example, the Hotelling (1931) \( T^2 \) test could check whether the means of \( \hat{m}(k) \) and \( \hat{m}(j) \) are statistically indistinguishable. This is a necessary condition for the SDF theory. A sufficient condition is that the entire vectors \( \hat{m}(k) \) and \( \hat{m}(j) \) are congruent. The non-parametric Kruskal-Wallis (1952) test is designed for this purpose and will reject the null hypothesis if \( \hat{m}(k) \) stochastically dominates \( \hat{m}(j) \) or vice versa.

It might be sensible to conduct tests with assets that seem unlikely, \textit{a priori}, to share the same SDF, such as equities in one group and bonds in another (over the same sample period, of course) or perhaps equities in two different countries. This would represent a tougher hurdle for the SDF theory but any viable theory should be able to surmount the most severe test possible.

There is no reason to restrict our attention to just two sets of assets. Every vector computed according to (9), with the same time series of observations but with different assets, should be congruent. The Welch (1951) test would serve nicely to check whether the means of all such
vectors are the same and the Kruskal-Wallis test can handle multiple comparisons of entire distributions. The Welch test is robust against heterogeneity in the variances of the distributions being compared. On the other hand, the non-parametric Brown/Forsythe (1974) test is designed specifically to check for unequal volatilities using absolute deviations. By implementing all three tests, we should be able to ascertain whether two or more estimated SDFs have equal means, volatilities or display stochastic dominance. Violation of any one of the three tests would be evidence against the SDF theory.

Test power is a more difficult issue. As indicated in section I.B, power undoubtedly depends on the relative sizes of the time period, T, and the cross-sections, N, K, J etc. Unless the data are extremely high frequency, one usually has more assets than time periods. But in the present case, unlike with most asset pricing tests, this is an advantage. On the other hand, a large T, but not nearly as large as N, might sometimes confer an advantage because the time series sums of the expectation surprises, (the ε₁,i’s in (3)) will compromise the accuracy of the SDF estimates for short time series. We investigate this issue in section II using simulated data.

Nothing above requires specification of a proxy for the SDF. Even a riskless rate, if there is one, whose gross return \( R_f \) satisfies the useful property, \( E(m_t) = 1/R_f \), is not necessary. Moreover, tests can be conducted with relatively short time series samples, but still with the caveat that longer samples may be less prone to estimation error.

II. Qualities of Our SDF Estimator

II. A. Comparing the Estimated SDF and the True SDF

To provide some insight about the performance of our SDF estimator, this sub-section offers a series of simulations to compare true SDFs with estimated SDFs. We assume temporarily that the SDF theory is true. We generate “true” SDFs with a mean equal to the reciprocal of the gross riskless interest rate, as the SDF theory stipulates, and with different levels of time series variation about the mean. We then simulate gross returns so that their product with the true SDF averages to unity over a specified sample period. Then, we perturb each gross return with a
random disturbance to induce sampling error. Finally, using the resulting sample returns, we calculate our SDF estimator and compare it with the known “true” SDF.

Step 1 is to generate a time series sample of “true” SDFs of length T. Specifically, we select a gross riskless rate, $R_F$, $(1+$ the riskless return), and generate the SDF at time $t$ as

$$m_t = \frac{1}{R_F} \exp(\xi_t - \sigma_\xi^2 / 2) , \ (t=1,\ldots,T) \quad (15)$$

where $\xi$ is a IID random variable with mean zero and standard deviation $\sigma_\xi$. The exponential in (15) has a mean of 1.0 if $\xi$ is normally distributed, which we assume to be the case initially$^7$ and, in accordance with SDF theory and the absence of arbitrage, (15) provides a strictly positive $m_t$.

In Step 2, initial gross unscaled returns are generated to be strictly positive (thus assuming limited liability) with a pre-specified mean and volatility (which are assumed to be the same for all individual assets); i.e., for asset $i$,

$$\hat{R}_{i,t} = \mu \exp(\xi_{i,t} - \sigma_\xi^2 / 2) , \ (t=1,\ldots,T; \ i=1,\ldots,N) \quad (16)$$

where $\mu$ is the expected gross return ($1 +$ the net return) and $\sigma_\xi$ is the standard deviation of the unscaled gross return $\hat{R}$. We find in simulations (in the robustness section below) that imposition of equal means and variances at this stage has an immaterial effect because the final scaled returns used in all subsequent calculations are computed as

$$R_{i,t} = \frac{\hat{R}_{i,t}}{\sum_{i=1}^{N} m_t \hat{R}_{i,t} / T} \exp(\vartheta_{i,t} - \sigma_\vartheta^2 / 2) \quad (17)$$

where $\vartheta$ is an IID return perturbation with mean zero and standard deviation $\sigma_\vartheta$. As required by SDF theory, (17) implies that

$$E\left[\frac{1}{T} \sum_{i=1}^{T} m_i R_{i,t}\right] = 1 .$$

$^7$ Later, we consider non-normally distributed variation.
Final gross returns on N assets are generated independently for T time periods according to (17). Consequently, except for their common dependence on the average SDF, the returns are uncorrelated with each other. We consider the consequences of this assumption in the robustness section below.

The final simulation step uses the estimator (9) with the final returns from (17) to obtain \( \hat{m}_t \) (t=1,…,T), for comparison with the true values from (15), \( m_t \) (t=1,…,T). We make this comparison using two criteria, the simple correlation between \( m \) and \( \hat{m} \) and the Theil (1966) \( U_2 \) statistic. The latter is closely related to the mean square prediction error, (MSE). Specifically,

\[
MSE = \frac{\sum_{t=1}^{T} (m_t - \hat{m}_t)^2}{T}, \quad \text{and}
\]

\[
U_2 = \frac{MSE}{\left( \sum_{t=1}^{T} m_t^2 / T \right)}.
\]

The correlation is easy to understand but it can be a bit misleading because it fails to measure whether \( m \) and \( \hat{m} \) are congruent. For example, if \( \hat{m} = 2m \), the correlation would be perfect. An advantage of the MSE is that it can be decomposed into three components, one due to a difference in means, another to a difference in volatilities, and third due to a lack of correlation; i.e.,

\[
MSE \equiv (\bar{m} - \bar{\hat{m}})^2 + (s_m - s_{\hat{m}})^2 + 2(1 - \rho)s_m s_{\hat{m}}
\]

(18)

where the superior bars indicate means, the s’s are standard deviations and \( \rho \) is the correlation between \( m \) and \( \hat{m} \). This decomposition is particularly relevant in our application because we would expect \( \hat{m} \) to have more volatility than \( m \) due to sampling error and to be imperfectly correlated. However, if the SDF theory is true, the two means should be close to one another.

Before beginning full-scale simulations, we present a proof of concept by showing a comparison between the true and estimated SDF when the IID perturbations in (17) are relatively small. This illustration uses 120 assets and 60 time periods, (a modest degrees-of-freedom according to section I.B), a riskless rate of .4% per period, and a true SDF standard deviation of 4% per period. Initial returns have means of .8% per period (mean gross returns of 1.008) and standard
deviations of 8% per period, a material level of return volatility. However, the standard deviation of the perturbations in (17) is intentionally small, .01% per period.

The final returns, (after making sure the means of the SDF-Return product is 1.0 on average), still have substantial volatility. Their average standard deviation is 8.1% over the 120 simulated assets with a minimum (maximum) individual asset standard deviation of 5.17% (11.2%).\(^8\)

Figure 1 plots the resulting estimated SDFs against the true SDFs for the 60 time periods. Their difference is trifling. Their correlation is 0.99946; the MSE is nevertheless almost entirely due to imperfect correlation, whose decomposition fraction is 0.959, because the SDF means and volatilities are virtually identical. This shows that the theoretical bias discussed in section 1.B is empirically trivial when the return perturbations are minor.

Turning now to simulations with larger return perturbations, we examine the relative influences of the time series and cross-sectional sample sizes, \(T\) and \(N\), respectively, and also the impact of return perturbations, \(\sigma_\theta\), the volatility of the true SDF, \(\sigma_\xi\), and the risk-free rate \(R_F\). With this many parameters, it is hard to summarize results compactly over a continuum of parameter values, so we resort to a hopefully more illuminating expedient. We simply generate the simulated \(\hat{m}\) and \(\tilde{m}\) with several different choices of the parameters and then present summary linear regressions of the correlations and Theil’s \(U_2\) on all the parameters jointly.

Our estimator of the SDF requires \(N>T\), so we let \(T=30, 60, 90,\) and 120 and for each \(T\), we set \(N=240, 360, 480,\) and 960. These choices are made to roughly match sample sizes and numbers of assets in our later empirical work below. For each \(N\) and \(T\), we let the true SDF volatility take the values \(\sigma_\xi=.5\%, 1\%, 1.5\%\) and 2% per month. For each \(N, T,\) and \(\sigma_\xi\), the perturbation volatility \(\sigma_\theta\) takes on nine values beginning with \(\sigma_\xi/5\) and increasing by this increment to terminate at \(1.8\sigma_\xi\). Finally, for each choice of the previous parameters, we let the risk-free rate vary as follows: \(R_F=.1\%, .2\%, .3\%, .4\%\) and .5% per month. This results in 2,880 different parameter combinations. For each parameter combination, we generate completely different true SDFs and returns and hence have independent sets of sample SDFs.

\(^8\) The minimum (maximum) individual return is -25.3% (35.6%).
Table 1 gives the results, panel A for the correlation between \( m \) and \( \hat{m} \), and Panel B for Theil’s \( U_2 \). In Panel A, we see that the correlation falls with \( T \), rises with \( N \), rises with \( \sigma_\xi \), the volatility of the true SDF, and falls with \( \sigma_\varphi \), the perturbation volatility, all with a very high levels of significance. Each regression coefficient, of course, indicates the marginal influence holding constant other parameters. For the two volatilities, the directions are intuitively obvious because a greater spread of the true values and a smaller perturbation variance should improve the fit. For \( N \) and \( T \), the fit seems related to the degrees-of-freedom, \( N-T \), (remember, \( N>T \)). Fewer degrees-of-freedom result in less precise estimation. The riskless rate has no significance whatsoever; this too is hardly surprising because a simple translation of the mean SDF should essentially be immaterial.\(^9\)

The results for Theil’s \( U_2 \) in Panel B essentially agree with the results for the correlations in Panel A, with opposite signs as expected (since \( U_2 \) is larger when the fit is worse), except for the volatility of the true SDF, which has the same sign but less statistical significance. This exception might be explained by the fact that \( U_2 \) is scaled by a denominator that relates to the variance of the true SDF. The other three significant variables in panel A are even more significant in Panel B and the overall explanatory power is larger.

We find that the decomposition of the MSE into its three components, equation (18), reveals virtually no effect at all from the first component, a difference in means between the true and estimated SDFs. On average over the 2,880 combinations of parameters, the mean difference component’s fraction of the total MSE has a value of 0.0000 and the largest value is only 0.0012. In contrast, the averages of the standard deviation difference component and the correlation component are, respectively, 0.2426 and 0.7574 as fractions of the total MSE. (For each parameter set, the three fractional components sum to 1.0 by construction.) The largest and smallest values are, respectively .8346 and 0.000 (1.0000 and 0.1654) for the standard deviation difference component (correlation component.)

Each of the 2,880 parameter combinations uses a different simulated set of “true” SDFs, which results in a corresponding and different set of estimated SDFs. Consequently, we can compare the 2,880 means of true and estimated SDFs. They are very close. The averages over 2,880 sets

\(^9\) In unreported results, we verify that this is also true of the mean and variance of the initial returns in (16).
are 0.9956 and 0.9960 for, respectively, the estimated and true SDF means. The standard deviations of the means across the 2,880 sets are, respectively, 0.2438 and 0.2439. Their correlation is 0.9977. Hence the mean of our estimator is close to the true mean SDF regardless of the parameters.

However, although the means are close, the period-by-period estimated and true SDFs display substantial divergence for some parameter combinations. The average correlation is .189 and the maximum and minimum correlations over the 2,880 parameter combinations are, respectively, 0.951 and -0.547. This makes it very clear that ill-considered parameters degrade the performance of our SDF estimator when there is a large amount of sampling variation.\(^\text{10}\)

Panel C of Table 1 reports determinants of the time series standard deviation of the estimated SDFs. The impact of degrees-of-freedom (essentially N-T) is apparent; Larger N and smaller T reduce sampling error and result in a better-behaved estimated SDF. Holding N and T constant, more volatility in the return perturbation brings, not surprisingly, in a more volatile estimated SDF. The time series volatility of the true SDF, however, has no significant impact and neither does the riskless rate.

The variance of our estimated SDF should increase with T due to the approximation error. This is because the elements in the estimated SDF vector are equal to the right side of (6) multiplied by T. This multiplication converts the average approximation error to the sum of approximation errors, (summed over T periods.) The standard deviation of this sum increases with \(\sqrt{T}\). In an unreported alternative regression to Panel C in Table 1, using \(\sqrt{T}\) instead of T as a regressor, we find that virtually nothing is altered except the coefficient.\(^\text{11}\)

In Panel D, of Table 1, we finally see something that is influenced by the true riskless rate; viz., the implied riskless rate from the reciprocal of the estimated SDF. The t-statistic is 2.42, but the overall explanatory power is meager. Also, both the perturbation volatility and the volatility of

\(^{10}\) For Theil’s U\(_2\), the mean, maximum and minimum are, respectively, 0.339, 0.787, and 0.0359. Larger values indicate more disagreement between the estimated SDF and the true SDF.

\(^{11}\) The t-statistic for \(\sqrt{T}\) is 53.8 as opposed to the 53.9 reported for T in Table 1. Everything else is similarly close; e.g., the adjusted R-square is 0.730 as opposed to 0.731.
the true SDF are marginally significant, which may be explained by Jensen’s inequality (since the implied riskless rate is obtained from a reciprocal of an estimated SDF.)

II. B. Test Power

This sub-section provides evidence about the power of our proposed tests of SDF theory by tabulating type II errors under a variety of different simulated conditions. The type II error, often called the “power” of the test, is the probability of rejecting a false null hypothesis. To estimate power, we must set up a simulation so that the true SDFs for different groups of assets are not the same. For two or more sets of assets, we then estimate separate SDFs and tabulate the rejection frequency of the null hypothesis that all SDF estimates are the same except for sampling error. In a simulation, the rejection frequency is the fraction of replications with test p-values less than the type I error.

The three tests used here and in the empirical section are the Kruskal/Wallis (1952) (KW) non-parametric one-way analysis of variance based on ranks, which rejects a false null hypothesis if one or more sample SDFs is stochastically dominant or has an abnormal median, the Welch (1951) (WE) test for equal means, which allows for unequal variances, and the Brown/Forsythe (1974) (BF) test for unequal variances.

The relevant test depends on the nature of the difference among SDFs. For example, if the medians differ but the means and variances are about the same, the KW test should reject the null but the WE and BF test might not. Similarly, if the SDF distributions have similar location on the real line but have disparate volatilities, the BF test should reject but the other tests would not. This implies that simulations should examine various type of SDF heterogeneity; i.e., different locations or volatilities or both and perhaps differences in higher moments. Obviously, we cannot hope to examine every possible type and size of differences across SDFs, so this section is unavoidably limited. However, we will gladly supply the simulation Fortran code to anyone interested in examining power for other parameter choices.

To be most relevant for the empirical tests to follow, we perform power calculations for several choices of the most important parameters, which are the number of sample periods, $T$, the number of assets in each group, $N$, the means and variances of the true SDFs (which can differ
across groups), the number of asset groups, and the volatility of return perturbations. For each choice of parameters, the simulations are replicated 1,000 times and the power is tabulated as the null hypothesis rejection frequency.

Our first set of simulations has just two asset groups. Parameter combinations include \( N = 240, 480, 720 \) and 960. For each \( N \), \( T = 30, 60, 90 \), and 120. To illustrate differences in the tests, we conduct one simulation with SDFs that differ only in location; i.e., two values for the riskless rate, .001 and .050 per period, but with the same SDF volatility, 1% per period. The second simulation has two SDFs with the same mean, \( R_f = .001 \), but different standard deviations, 2% and 1%. In both simulations here (and in all that follow), SDFs are generated according to the true SDF model in Section II.A, equation (15), and returns are generated by (17).

For the first simulation, the one with differing SDF means, the test powers for KW and WE are plotted in Figure 2. The third test, BF, has virtually no power in this case because the SDF volatilities are the same; the average power of BF across all parameter combinations is 7.025%, not far from the type I error (as it should be.) For KW and WE, however, the power is respectable. It is better, of course, for small perturbation volatility, (which is given in parentheses after the test acronym.) The figure is sorted by the test power of KW(1%), but power does not increase monotonically with either \( T \) or \( N \) nor with \( N-T \).

For the smaller perturbation volatility (1%), the highest power is for \( N = 960 \) and \( T = 120 \), both being the largest values simulated. The lowest power is for \( N = 240 \) and \( T = 30 \), which are the lowest values considered. Overall, power increases with both \( N \) and \( T \) for these 1% perturbation tests, but it is not strictly monotonic. For instance, given \( N = 240, T = 90 \) has slightly more power than \( T = 120 \).

For the larger return perturbation volatility (2%), the power is very weak (around 10%) with the worst parameter combination, which is \( N = 240 \) and \( T = 120 \). However, power exceeds 80% for the best combination, \( N = 960 \) and \( T = 120 \). Some of the 2% return perturbation power patterns are inconsistent with each other. For example, holding \( N = 240 \), we see that power decreases with \( T \), the combination \( N = 240, T = 120 \) having the smallest power. In contrast, for \( N = 960, 720, \) and 480, the opposite occurs in that power increases with \( T \). Evidently, there is a tradeoff between
sampling return perturbation and degrees-of-freedom. When N is relatively small (120), power is reduced if T is too close to N. This is a degrees-of-freedom issue. However, when N is relatively large, (N=480, 720, 960), a larger T reduces the influence of return perturbations and there are still sufficient remaining degrees-of-freedom.

In the second simulation, the tests reverse in strength. KW and WE have virtually no power because the locations are the same; their average powers are 2.34% and 1.52%, respectively.\textsuperscript{12} The BF test, in contrast, has excellent power which rises monotonically with N-T. As shown in Figure 3, BF(1%) has perfect (100%) power when N-T exceeds the mid-300 range. Even for the larger return perturbation variance, BF(2%) has power above 80% when N-T>600. However, for lower N-T, its power drops off precipitously and is only around 20% for N-T<150.

Clearly, return perturbation volatility has a large deleterious impact on power, but it appears that this can be overcome with a large enough collection of assets and a judicious choice of the time series sample size. It also seems clear that all three of these tests, (KW, WE, and BF), provide valuable information about the validity of SDF theory. WE and KW are similar, and in Figure 2 WE has slightly higher power, but WE has the disadvantage of being a parametric test; hence KW might be preferred when one is not sure about the distributions of returns or of the underlying SDFs.

In the next simulation, we allow both the mean and volatility of the true SDFs to differ and also introduce stochastic dominance by allowing the SDF with the larger mean to have a smaller volatility. Thus, the riskless rate is set to .001 (.05) for the first (second) SDF and the volatility is set to 0.02 (0.10). Figure 4 plots the resulting power for all three tests, KW, BF, and WE and for two different return perturbation volatilities, 1% and 2%. The numbers of assets in a group are N=240, 480, 720, and 960. For each N, the numbers of time periods are T=30, 60, 90, and 120. The figure is arranged so that for each test, the 16 N,T parameter combinations vary first by N and then by T within each N. In other words, from left to right they are: N,T = {240,30; 240,60; 240,90; 240,120; 480,30;…; 960,90; 960,120.}

\textsuperscript{12} They should be 5%, the type I error, but these average are only over 30 different parameter combinations and there is some sampling error. However, the observed values do appear rather smaller than they should be.
Given the results in Figures 2 and 3, there are no surprises in Figure 4. For the smaller return perturbation volatility (1%), the BF test has almost perfect power when N is 480 or above. For the smallest N, (120), however, BF(1%)’s power declines with T, again indicating the influence of fewer degrees-of-freedom. The power of BF(2%) is smaller throughout, as expected. It reaches 90% only for the largest N (960).

The KW and WE tests are almost tied for power at every parameter combination, one occasionally being slightly larger than the other and vice versa. For the smaller return perturbation volatility, KW(1%) and WE(1%) have excellent power except for the smallest value of T (30) where power is 60% to 70%. For T>30, power exceeds 80% in all cases and reaches almost 100% for N=720 and 960. For larger return perturbation volatility (2%), however, power is poor at N=240 but it exceeds 80% for the larger values of N and even reaches 90% with N>480 and T>60.

Finally, we document power with a larger number of asset groups. We choose five groups to match some of our later empirical tests. To make the tests face a tough challenge, we set up the experiment so that just one of the groups has a stochastically dominant SDF, the other four having SDFs with the same mean and variance. Asset group #1 has a stochastically dominant SDF with a riskless return of 0.001 and a standard deviation of .02. Groups #2 through #5 each have SDFs with a riskless return of .05 and a standard deviation of 0.1. Even though the last four groups have SDFs with the same mean and volatility, the sample values of the SDFs are generated independently across all groups.

Figure 5 presents the results. It is arranged identically to Figure 4, with 16 combinations of N and T for each of the six tests, two tests with differing return perturbation volatilities for each of the three test statistics, KW, BF and WE. We are surprised that the test power is actually larger in some cases for this five-group experiment than for the two-group tests reported in Figure 4. For example, the WE(1%) test is particularly impressive with more than 90% power for 15 of the 16 N,T combinations; (and the only exception, N=240, T=30 still has 85% power.) For nine of the 16 parameter combinations, the WE(1%) test has almost perfect power. Similarly, the KW(1%) also has higher power in the five-group tests and is close to perfect in the same nine cases. It also displays slightly higher power for the four T=30 combinations. The BF(1%) test
remains virtually perfect for N>240 and the power is similar to that observed with the two-group
tests for N=240.

For higher return perturbation volatility (2%), the KW(2%) and WE(2%) display mixed results in
this five-group test as compared with the two-group test in Figure 4. With large N, (720 and
960), power is higher but for small N, (240 and 480), power is lower, dramatically so for N=240.
Indeed, for N=240 and T=90 and 120, the power is practically zero. The BF(2%) test also shows
lower power except for N=960.

In summary, from all these simulations we have learned that very large N, close to 1,000,
provides robust power under a variety of conditions including the time series length, T, and the
return perturbation volatility. With low return perturbation volatility, all of the tests provide
good power except when T approaches N/2 and the degrees-of-freedom start to become
problematic. The power is generally very poor when the return perturbation variance is large and
T is a large fraction of N.

III. Data

We collect monthly return observations on U.S. bonds, stocks, currencies (per US$),
commodities and real estate (REITs, or real estate investment trusts), for July 2002 through
December 2013, 138 months in all. The data begin in July 2002 because the Trace data base
starts reporting bond returns in that month. Stocks are sampled randomly from those on the
CRSP database. We purposely select equities with low leverage to make them as different as
possible from bonds, although we also select an equal-size random sample of other equities for
later comparison. Currencies and commodities are drawn from the Datastream and Real Estate
Investment Trusts (REITs) from the CRSP database. In the cross-sectional sample, there are 956
low-leverage stocks, 123 bonds, 37 spot exchange rates per US$, 47 commodities, and 89 REITs
that have simultaneous observations for every month.

13 The average leverage (book debt/total assets) ratio is 10.21% for the 956 low-leverage equities.
IV. SDF estimates and empirical tests of the SDF theory

IC.A. Tests among asset classes

The SDF theory should apply to any partition of the available assets, but we decided to begin with what could be a tough challenge. We estimate SDFs from each asset class independently and then test whether they are the same across asset classes. Our SDF estimator requires more assets than time periods, so we are limited to time series samples shorter than the number of individual assets in the smallest class, which is currencies with 37. Hence, the 138 available months are separated into roughly equal subsamples, 34 observations in the first two subsamples and 35 observations in the next two. We realize these tests probably lack power because N-T, the degrees-of-freedom, is quite small for some asset classes. Nonetheless, we believe they are worth reporting while recognizing their likely limitations. The results are in Table 2.

The Kruskal/Wallis (1952) (hereafter KW) test indicates whether one set of SDF estimates stochastically dominates any other and it also provides a test of the difference in medians. There are five sets of sample SDFs, one for each asset class, which implies that the KW Chi-Square variate under the null hypothesis (H₀: no SDF dominates another) has four degrees-of-freedom. According to KW test results reported in Table 2, there is no stochastic dominance in any of the four sub-periods. The sample medians are not significantly different from one another. Hence this test does not reject the SDF theory for these assets and time periods.

Table 2 also reports tests for the equality of means and variances across the five sets of SDF estimates, the Welch (1951) (WE) test for means and the Brown/Forsythe (1974) (BF) test for variances. In agreement with the non-parametric KW test, the WE test finds no evidence of a difference in means for the SDFs estimated independently from the five asset classes. None of the p-values indicates significance.

The WE test allows for unequal volatilities across asset classes. This is fortunate because the BF test for differences in variances rejects the null in every sub-period. Evidently, although the sample SDFs appear to be located with their means and medians close to one another,¹⁴ at least one asset class has sample SDFs with significantly larger or smaller variance than the others. In

¹⁴ The Welch test for equal means is valid even when variances are unequal.
order to ascertain which asset class (or classes) is responsible, Table 3 reports the time series standard deviations of the sample SDFs.\footnote{\footnotesize The simulations in section II.B suggest that test power might not be very good for small collections of assets such as 37 for currencies and 47 for commodities. However, since the BF test rejects strongly, power per se does not appear to be a problem.}

It appears that currencies and commodities have larger volatilities than equities, bonds, and real estate. One possible explanation is that the sampling error in estimated SDFs is larger for asset classes with fewer constituent members. There seems to be a strong negative connection between the number of available assets and the volatility. Currencies have the smallest number of individual assets (only 37) and commodities are next (with 47.) This explanation is buttressed by the simulation results in Section II.A, which reveals a material improvement in the quality of our estimator with the number of assets, holding constant the time series sample size.

\textbf{IV.B. Tests with larger samples of assets and time periods}

To investigate the possible confounding impact of sampling error, we conduct two further experiments. First, we compare stocks and bonds alone, without reference to the other three asset classes. Stocks and bonds dominate the sample with 956 and 123 individual assets, respectively. Recall that the number of assets N in a group must exceed the number of time series observations T. Since there are only 123 bonds available, we cannot use all 138 time series observations at once, so we simply divide them in half, 69 months in the first sub-sample (July 2002 – March 2008) and 69 in the second (April 2008 – December 2013.) Table 4 provides the results.

As the Table 4 Brown/Forsyth tests indicate, the increased sample sizes, both in number of asset and in time periods, does not overturn the previous result that the SDFs have divergent volatilities. Table 5 reports the volatilities, which are considerably larger for bonds than for stocks in both sub-periods. Evidently, the number of bonds remains too small compared to the number of equities, which probably implies more sampling error and hence higher estimated SDF volatility for bonds.

In the second experiment, we abandon a strict asset class categorization in order to estimate sample SDFs using all available monthly observations at once and roughly equal-sized groups of
assets. This increases the time series sample size from $T=34$ or $T=35$ (as in Table 2) or $T=69$ (in Table 4) to $T=138$. Since the number of assets $N$ in a group must exceed the number of time series observations $T$, it becomes necessary to mix stocks, which are the most numerous, in with the four other asset types in separate groups. There are 1252 individual assets of all types available for the 138 sample months, so five roughly equal-sized groups would contain, respectively, 250, 250, 250, 251, and 251 individual assets.

We compose the groups in the following manner: To group #1, we assign 250 equities, selected randomly; in group #2 we mingle 127 randomly-selected equities with all available (123) bonds; similarly, group #3 has 213 equities and 37 currencies; group #4 has 204 equities and 47 commodities; and group #5 has 162 equities and 89 REITs. The results are reported in Table 6. The Welch (1947) test for means and the Brown/Forsythe (1974) test for variances are in agreement with the non-parametric KW test. There is no evidence of a significant difference in the SDFs estimated from the different asset groupings. This supports the SDF theory completely. After correcting for possible sampling error differences across test groups, there is no evidence of SDF differences even though the groups are heterogeneous in the sense of including five distinct asset classes.

However, there is one caveat. Our simulations in section II.B reveal that test power might not be very large when there are only 240 assets in each group unless sampling error is rather small. Thus far, we have not attempted to disentangle sampling return perturbation volatility from volatility in the true SDF. The variance of the estimated SDF the sum of the two variances.

**IV.C. Tests with greater power**

In the hope of achieving more test power, we conduct two further empirical experiments. In the first, we divide the sample of low-leverage equities into two equal-sized groups of 478 stocks each and work with the entire 138 time series observations. Section II.B suggests that this choice of $N$ and $T$ should have good power. In the second test, we expand $N$ even further by collecting a second group of 956 equities, randomly sampled from remaining CRSP stocks that do have
significant levels of leverage.\textsuperscript{16} Conceptually, this second test should be an exacting hurdle for SDF theory because the two groups of equities differ markedly in their leverage ratios.

The results for both tests are reported in Table 7. In Panel A, we see that none of the three tests, KW, WE or BF, rejects the null hypothesis that the SDFs are different in the two groups of 478 low-leverage equities at a high level of significance, though the BF test is on the margin with a p-value of 0.084. Panel B reports a stronger inference; even with very different leverage (and, as consequence, likely different levels of riskiness), there is no evidence of a difference in SDFs. In all cases, the p-values are far from indicating significant rejection of the null hypothesis.

In conclusion, for a battery of tests with differing asset classes, differing group sizes, and diverse time series sample sizes, the SDF theory holds up well. It cannot be rejected after properly accounting for sampling variation. Of course, our tests are valid only for a recent decade of observations on US assets, so more comprehensive tests with longer samples and international collections of assets are clearly in order.

IV.D. Properties of estimated SDFs; disentangling sampling error and true SDF volatility and the Hansen/Jagannathan bounds

In the previous sub-section, we find with presumably powerful tests that SDF estimates from low- and higher-leveraged stocks are not significantly different. This does not prove that SDF theory is true, but the theory cannot be rejected by those tests. In this section, we temporarily assume that the theory is true, which enables us to shed light on the properties of SDF estimates. It also permits the disentanglement of volatility in the true SDF from sampling error volatility in the estimated SDF and it allows us to check whether or estimates satisfy the Hansen/Jagannathan bounds.

In agreement with previous notation, we now let $\hat{m}(L)$ denote the vector of estimated SDFs from the low-leverage stocks and $\hat{m}(H)$ denote the estimated SDFs from the higher-leverage stocks. Given SDF theory, an element of these vectors at time $t$ can be expressed as

\textsuperscript{16} The average leverage ratio (book debt/total assets) for this second group of stocks is 32.51%; the low leverage group has an average ratio of 10.21%.
\[ \hat{m}_{j,t} = E_{t-1}(m_t) + \nu_{m,t} + \nu_{j,t}, \quad j=L,H \]  

(19)

where \( \nu_{m,t} \) is the unexpected component of the true SDF at time \( t \) and \( \nu_{j,t} \) is the estimation error in the sample SDF for group \( j \) (\( j=L,H \)). No element on the right side of (19) is correlated with any other, so the time series variance of the estimated SDF is

\[ \text{Var}(\hat{m}_j) = \text{Var}[E(m)] + \text{Var}(\nu_m) + \text{Var}(\nu_j), \quad j=L,H. \]  

(20)

In addition, since the true SDF is common to both \( L \) and \( H \) by assumption, while the estimation errors are independent of each other, we have using (19)

\[ \text{Var}(\hat{m}_L - \hat{m}_H) = \text{Var}(\nu_L) + \text{Var}(\nu_H). \]  

(21)

Combining (20) for both \( L \) and \( H \) with (21), we obtain

\[ \text{Var}[E(m)] + \text{Var}(\nu_m) = \frac{1}{2} [\text{Var}(\hat{m}_L) + \text{Var}(\hat{m}_H) - \text{Var}(\hat{m}_L - \hat{m}_H)]. \]  

(22)

The left side of (22) the total volatility induced by the SDF, including the intertemporal evolution of its expectation and its period-by-period unexpected component. Subtracting this result from (20) provides estimation error variances for \( j=L \) and \( j=H \).

SDF theory requires that \( E_{t-1}(m_t) = 1/(1+R_{F,t}) \) for the riskless rate \( R_F \) at time \( t-1 \). During the time period of our sample, 2002-2013, the riskless rate had historically low variation over time, so \( \text{Var}[E(m)] \) should be relatively small compared to \( \text{Var}(\nu_m) \), which should dominate (22).

Estimated over July 2002 through December 2013, the standard deviations of \( \hat{m}_L \) and \( \hat{m}_H \) are, respectively, 0.580 and 0.696 per month. The standard deviation of the difference in the SDFs; i.e., the square root of equation (21), is 0.5786. This implies a standard deviation of SDF components, the square root of (22), equal to 0.3481. The standard deviations for the estimation errors for \( L \) and \( H \) are then, respectively, 0.4636 and 0.6031. Not surprisingly, higher leverage equities are associated with larger estimation errors.
The Hansen/Jagannathan bounds require that \( \frac{\sigma(m)}{E(m)} \) be larger than the largest possible Sharpe ratio. Recent opinions, Welch (2000), seem to be that the excess return on the best possible portfolio is no more than about 7% per annum (or even lower lately) and the portfolio’s standard deviation may be around 16% per annum, so the largest Sharpe ratio is no more than 0.44. The sample means of \( \hat{m}_L \) and \( \hat{m}_H \) are, respectively, 0.9945 and 0.9961, both approximately unity. Our annualized SDF standard deviation is \( 0.3481 \sqrt{12} \), which comfortably satisfies the HS bounds. This inference contrasts strongly with previous research that has specified SDF proxies that depend on macroeconomic data. Evidently, SDFs that depends on returns, such as ours and Long’s Numeraire portfolio, are sufficiently volatile. This is a puzzle that clearly deserves further investigation.

The means and standard deviations of the estimated SDFs can be used to conduct a simple test that the true SDFs are positive (and consequently there are no arbitrage opportunities.) The t-statistics for low- and higher-leveraged equities are, respectively, 20.2 and 16.8, thereby overwhelmingly indicating that the SDFs are not negative.\(^{17}\)

To get a visual image of the evolution of our SDF, it is appropriate to first expunge estimation error. This is not possible for each individual time series observation, but one can adjust the overall series to have the true SDF volatility as estimated by (22). We simply need to find an attenuation coefficient, \( \gamma \) such that \( \text{Var}(\gamma \hat{m}) = \text{Var}(\hat{m}_M) \), which assumes that the riskless rate’s variance is sufficiently small that it can be ignored; hence, 
\[
\gamma = \left[ \frac{\text{Var}(\hat{m}_M)}{\text{Var}(\hat{m})} \right]^{1/2}.
\]

The adjustment entails the transformation
\[
\hat{m} = \bar{m} + \gamma (\hat{m} - \bar{m}),
\]
where the double “chapeau” indicates the transformed SDF and \( \bar{m} \) is the sample mean. For the low and high leverage equity groups, the attenuation coefficients are .6005 and .4999, respectively.

\(^{17}\) There is a slight degree of autocorrelation in the estimated SDFs but it is too small to overcome this inference.
Figure 6 plots the two adjusted SDF series using a 12-month moving average to smooth out short-term fluctuations. There is clearly a connection between the two series, which is not a surprise because our test above could not reject the hypothesis that they are the same. There is, however, something of a puzzle here in that the SDF is larger than 1.0 in the middle of the 2000 decade for both series. Of course, this is the *ex post* SDF, including the unexpected component. The expected SDF would presumably be much smoother.\(^{18}\)

**V. Robustness Checks**

In this section, we investigate the qualities of our SDF estimator with alternative assumptions about returns. Sub-section V.A. examines the consequences of thick tails, a phenomenon that is seemingly ubiquitous for financial asset returns. Sub-section V.B. looks at the impact of returns that are cross-sectionally correlated and have different means and variances.

**V.A. Thick-tailed returns.**

In the simulations of section II, returns are log-normally distributed, so a natural question is whether our SDF estimator behaves as well when returns are characterized by very large or very small returns, well beyond those typically observed under a Gaussian regime. Our estimator does involve a cross-product matrix that contains squared returns, so it might be sensitive to extreme observations.

To examine this issue, we repeat the simulations of II.A holding everything the same except for the return perturbations. In other words, we generate “true” SDFs as before with a lognormal distribution as in equation (15) and the same panoply of parameters. Initial gross returns are also generated in the same way, as in equation (16).

But equation (17) is replaced by

\[
R_{i,t} = \frac{\hat{R}_{i,t}}{\sum_{t=1}^{T} \hat{R}_{i,t} / T} + \vartheta_{i,t}
\]  

\(^{18}\) Neither series has a unit root according to the usual tests.
In which the zero mean IID return perturbation $\vartheta$ is now additive and is distributed according to a truncated Cauchy distribution with a scale parameter that varies from .005 to .045 in .005 increments (i.e., nine different values.) The scale parameter is a measure of the Cauchy distribution’s spread; it replaces the standard deviation used for the same purpose with the Gaussian. However, it is not associated with a second moment because the Cauchy has an infinite mean and all higher moments are also infinite.

A truncated Cauchy possesses finite moments but still has very thick tails relative to a Gaussian. In the simulations here, we truncate the extremes, retaining only the middle 95% of simulated Cauchy values. With a 95% truncation and the scale parameters listed above, gross returns are guaranteed to remain strictly positive.

The return perturbation in (23) is additive, in contrast to the previously multiplicative lognormal return perturbation as in (17). This choice is necessitated by the extremely large positive values, even with truncation, that would result from taking the exponential of a Cauchy variate. We are not aware of a satisfactory method of correcting for the induced bias. In the Gaussian case, one simply subtracts half of the variance (see equations (15) through (17)), but there is no corresponding correction using the Cauchy scale for the same purpose. An additive return perturbation fineses this difficulty because it is symmetric and not exposed to the amplification of exponentiation.

Table 8, which corresponds to Table 1, presents the results with truncated Cauchy return perturbations. Comparing Panels A and B of the two tables, one observes that the results are virtually unchanged qualitatively and are even more significant with thick-tailed return perturbations. All the variables have the same signs and all the significant variables (which is everything except the riskless rate) are still significant.

There is one change in Panel C, which shows the influence of various parameters on the volatility of the estimated SDFs. In Table 8, the true SDF’s volatility has become significant. In

---

19 The simulations first select a cumulative distribution function p-value, a number between zero and 1.0, and then calculates the inverse Cauchy corresponding to that p. If the p is less than .025 or greater than .975, it is discarded and another p is randomly chosen.

20 Since the Cauchy mean does not exist, one often uses the median, but a Cauchy with median of zero always has an exponentiated median of 1.0. However, the exponentiated truncated Cauchy can have an extremely large mean.
Panel D, which explains the inferred riskless rate, the SDF volatility and the Cauchy return perturbation scale parameter are not significant while the true riskless rate is more significant. Earlier, we speculated that the volatilities might be showing up in Panel D of Table 1 because of Jensen’s inequality in the riskless rate’s reciprocal estimation, but instead, that result appears to be related to multiplicative return perturbations.

We again find no effect from the difference in means fractional component of the MSE. The averages of the standard deviation difference fractional component and the correlation fractional component are similar, 0.191 and 0.808, respectively.

As for the 2,880 means of true and estimated SDFs, they are still very close, with even a slightly higher correlation, 0.9998, and almost identical averages and standard deviations. The average correlation has risen to 0.439 and the maximum and minimum correlations over the 2,880 parameter combinations are now, respectively, 0.995 and -0.409.

In summary, thick-tailed returns do not seem to compromise the qualities of our estimator. Its seeming improvement with thick tails, however, may be partly attributable to the return perturbation being additive rather than multiplicative and to a set of Cauchy scale parameters that rendered the return perturbations less severe. Regardless of such caveats, however, there seems to be little cause for concern when returns exhibit thick tails.

V.B. Correlated Returns with Unequal Means and Volatilities

The simulated returns in section II are independent of one another and have the same expected values and volatilities. We now relax these conditions and generate correlated returns that have disparate means and divergent standard deviations. Perhaps the simplest way, and the way we choose, to simulate returns with such characteristics is to use the venerable market model. Hence, we assume that each initial gross return is obtained from the following model

\[ 1 + \hat{R}_{i,t} = \exp[R_F + \beta_i(R_{M,t} - R_F) + \zeta_{i,t} - (\beta_i^2 \sigma_M^2 + \sigma_\zeta^2) / 2] \]

where \( R_F \) is the net risk-free rate (not 1+R), \( R_{M,t} \) is a normally distributed “market” common return in period t, \( \beta_i \) is the slope coefficient or “beta” for asset i and \( \zeta_{i,t} \) is a normally distributed
IID “idiosyncratic” return for asset i in period t. The last term on the right of (24), in parentheses, is a volatility correction for exponentiation.

For each set of parameters, we generate a new set of market returns, idiosyncratic returns, and betas. Then, the simulation proceeds as before, making sure that the average initial gross return from (24), multiplied by the SDF, is equal to 1.0 and then adding sampling return perturbation as in equation (17) of Section II.A.

The betas are assumed to be cross-sectionally normally distributed with a mean of unity and a standard deviation of 0.1, which implies that most betas fall between 0.8 and 1.2. Since the beta is different for each asset, the expected returns vary cross-sectionally as well.

The market returns are assumed to have a mean equal to the risk free rate plus a premium equal to 0.6% per month and a standard deviation of 4% per month, approximately 13.9% per annum. The idiosyncratic returns are assumed to have a standard deviation of 8% per month, so the market model R-square is 20%, which is in the usual range for equities.

Results are reported in Table 9. They are virtually identical in Panels A and B with the earlier results in Table 1 of Section II.A. Thus, inducing correlation and different mean returns and volatilities has no impact whatever on the correlations between true and estimated SDFs and on Theil’s $U_2$ statistic. There are some minor differences in Panels C and D, however. The standard deviation of estimated SDFs (Panel C) now shows significance for the true SDF volatility. The inferred riskless rate (Panel D) shows more significance for the true riskless rate and the number of assets and less significance for the return perturbation volatility. However, these differences are relatively small in magnitude.

The other indicators are also very similar, as one would expect given the similar results in Tables 1 and 8. For example, the correlations between true and estimated SDFs range from a maximum of 0.961 to a minimum of -0.567. The mean difference fractional component of the MSE is very close to zero in all cases (it’s maximum is only 0.0015), which implies that there is no material bias in the estimated SDFs.
In summary, returns that are correlated and differ in their means and volatilities present no difficulties for our SDF estimator.

VI. Conclusions

The stochastic discount factor (SDF) theory predicts that the same SDF should price all assets in a given period when markets are complete. We develop tests of this theory by first deriving an SDF estimator that depends only on observed returns and is agnostic with respect to macroeconomic state variables and preferences, on which is does not depend at all.

Our SDF estimator is theoretically biased in finite samples and has a standard error that depends on both the number of asset, N, and the number of time periods, T, used in its construction. Hence, to examine the estimator’s qualities, we resort to simulations. We find that the estimator is accurate when N-T is relatively large with N>2T and N near 1,000.

Equipped with an agnostic SDF estimator, we suggest three different tests of SDF theory that can potentially reject the theory when sample SDFs differ significantly across groups of assets. Simulations are presented to assess the power of these tests. For large N relative to T, the suggested tests have excellent power that approaches 100% depending on various parameters such as the volatility of the true SDFs and the sampling variation in returns. We also present evidence that our SDF estimator works well when return have thick tails and differ significantly in their means, volatilities and correlations with each other.

We apply our estimator and tests to data on U.S. equities, bonds, commodities, currencies, and real estate (REITs) over a common time period, 138 months from July 2002 through December 2013. As theory and the simulations predict, asset classes with few individual assets (a low N), such as commodities and currencies, produce sample SDFs with larger volatilities. However, even in this case, there is no evidence that SDF means are different across asset classes. This result suggests that excessive SDF volatility in smaller asset classes might be attributable to sampling variation.

This explanation is corroborated by reorganizing the individual assets into larger grouping; sampling error is thereby reduced and there are no longer any rejections of the SDF theory. The same result, no rejection, is obtained in a further test with larger numbers of individual assets.
(close to 1,000). Owing to data availability, such a test can be done only with equities. We find that two large groups of equities, one group with minimal leverage and the other with average leverage, are priced with SDFs that are not statistically distinguishable. We also find that these SDFs comfortably satisfy the Hansen/Jagannathan bounds and are very significantly non-negative.

Overall, the SDF theory’s main prediction, that the same SDF prices all assets during the same time period, cannot be rejected with our tests, data, or time periods. In addition, SDFs are positive with a high degree of statistical reliability. These results are consistent with complete markets and an absence of arbitrage. Future research will determine whether the same inferences are obtained with international data and with samples from other time periods.
Table 1
Simulated Performance Information for the SDF Estimator

To assess our SDF estimator, we simulate true SDFs with mean=$1/(1 + \text{riskless interest rate})$ and various time series volatilities. Gross asset returns are simulated so that their mean values multiplied by the SDFs are equal to 1.0 but errors perturb their sample values. The performance of the SDF estimator is measured by the correlation between true and sample SDFs and by Theil’s (1966) $U_2$ statistic, which is closely related to the mean square prediction error. Linear regressions are reported in Panel A where the dependent variable is the correlation and in Panel B where the dependent variable is $U_2$. In Panel C, the dependent variable is the sample time series standard deviation of the estimated SDFs. Panel D reports the implied riskless rate from the reciprocals of the estimated SDFs. There are 2,880 parameter combinations, each with an independently-simulated set of true SDFs and returns.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Correlation between true and estimated SDFs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>-1.104E-03</td>
<td>-11.034</td>
</tr>
<tr>
<td>N, Assets</td>
<td>1.505E-04</td>
<td>12.036</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>1.295</td>
<td>18.581</td>
</tr>
<tr>
<td>Perturbation Volatility</td>
<td>-1.744</td>
<td>-49.302</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>5.799E-01</td>
<td>0.244</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.488</td>
</tr>
<tr>
<td>B: $U_2$ from comparing true and estimated SDFs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>2.068E-03</td>
<td>54.927</td>
</tr>
<tr>
<td>N, Assets</td>
<td>-2.964E-04</td>
<td>-62.989</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>1.344E-01</td>
<td>5.126</td>
</tr>
<tr>
<td>Perturbation Volatility</td>
<td>6.237E-01</td>
<td>46.860</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>2.423E-01</td>
<td>0.271</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.782</td>
</tr>
<tr>
<td>C: Standard Deviation of Estimated SDFs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>2.917E-03</td>
<td>53.945</td>
</tr>
<tr>
<td>N, Assets</td>
<td>-4.270E-04</td>
<td>-63.166</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>5.180E-02</td>
<td>1.376</td>
</tr>
<tr>
<td>Perturbation Volatility</td>
<td>4.899E-01</td>
<td>25.622</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>2.360E-01</td>
<td>0.184</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.731</td>
</tr>
<tr>
<td>D: Riskless Rate Inferred from Estimated SDFs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>-3.106E-06</td>
<td>-0.229</td>
</tr>
<tr>
<td>N, Assets</td>
<td>2.401E-06</td>
<td>1.415</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>1.950E-02</td>
<td>2.063</td>
</tr>
<tr>
<td>Perturbation Volatility</td>
<td>1.164E-02</td>
<td>2.427</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>7.797E-01</td>
<td>2.422</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.008</td>
</tr>
</tbody>
</table>
Table 2
Tests of the SDF Theory With Five Asset Classes

Stochastic discount factors (SDFs) are estimated for five different asset classes, equities, bonds, currencies, commodities, and real estate (REITs), using simultaneous monthly observations for individual assets, July 2002 through December 2013, (138 months.) The total sample is divided into four similarly-sized subsamples with 34 monthly observations in the first two subsamples and 35 observations in last two. Differences across asset classes in the estimated SDFs are tested for stochastic dominance with the non-parametric Kruskal/Wallis (1952) statistic. Means and variances are compared with, respectively, the Welch (1951) and Brown/Forsythe (1974) tests. P-values are for the null hypothesis that the asset classes are all priced with the same SDFs. A low p-value rejects the null.

<table>
<thead>
<tr>
<th>Sub-Period</th>
<th>Stochastic Dominance (Kruskal/Wallis)</th>
<th>Equal Means (Welch)</th>
<th>Equal Variances (Brown/Forsythe)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul '02-Apr '05</td>
<td>0.976</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>May '05-Feb '08</td>
<td>0.956</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Mar ‘08-Jan ‘11</td>
<td>0.756</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Feb ‘11-Dec ’13</td>
<td>0.817</td>
<td>1.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table 3
Volatility of Sample SDFs by Asset Class and Sub-Period

The time series standard deviation is reported for sample SDFs estimated simultaneously with five different asset classes in four sequential sub-periods. The number of available assets, N, is reported in the second line.

<table>
<thead>
<tr>
<th>Sub-Period</th>
<th>Time Series Standard Deviation of Estimated SDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul ‘02-Apr ‘05</td>
<td>0.429 0.785 2.504 3.391 1.117</td>
</tr>
<tr>
<td>May ‘05-Feb ‘08</td>
<td>0.406 0.692 5.464 4.286 1.039</td>
</tr>
<tr>
<td>Mar ‘08-Jan ‘11</td>
<td>0.335 0.402 7.630 2.432 1.068</td>
</tr>
<tr>
<td>Feb ‘11-Dec ‘13</td>
<td>0.430 0.620 11.767 2.123 0.931</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Equities</th>
<th>Bonds</th>
<th>Currencies</th>
<th>Commodities</th>
<th>Real Estate</th>
</tr>
</thead>
<tbody>
<tr>
<td>956</td>
<td>123</td>
<td>37</td>
<td>47</td>
<td>89</td>
</tr>
</tbody>
</table>
Stochastic discount factors (SDFs) are estimated for equities and bonds using simultaneous monthly observations for individual assets, July 2002 through December 2013, (138 months.) The total sample is divided into roughly two equal sub-samples with 69 monthly observations each. Differences between stocks and bonds in the estimated SDFs are tested for stochastic dominance with the non-parametric Kruskal/Wallis (1952) statistic. Means and variances are compared with, respectively, the Welch (1951) and Brown/Forsythe (1974) tests. P-values are for the null hypothesis that bonds and stocks are priced with the same SDFs. A low p-value rejects the null.

<table>
<thead>
<tr>
<th>Sub-Period</th>
<th>Stochastic Dominance (Kruskal/Wallis)</th>
<th>Equal Means (Welch)</th>
<th>Equal Variances (Brown/Forsythe)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul ‘02-Mar ‘08</td>
<td>0.578</td>
<td>0.944</td>
<td>0.0000</td>
</tr>
<tr>
<td>April ‘08-Dec ‘13</td>
<td>0.927</td>
<td>0.968</td>
<td>0.0002</td>
</tr>
</tbody>
</table>
Table 5

Volatilities of Estimated SDFs for Stocks and Bonds

The time series standard deviation is reported for SDFs estimated simultaneously with stocks and bonds in two sequential sub-periods. The number of available assets, N, is reported in the second line.

<table>
<thead>
<tr>
<th>Sub-Period</th>
<th>SDF Volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul ‘02-Mar ‘08</td>
<td>0.480</td>
</tr>
<tr>
<td>April ‘08-Dec ‘13</td>
<td>0.420</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Equities</th>
<th>Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>956</td>
<td>123</td>
</tr>
</tbody>
</table>
Table 6

Tests of the SDF Theory With Mingled Groups of Assets in Different Classes

Stochastic discount factors (SDFs) are estimated for five groupings of asset from different classes using simultaneous monthly gross return observations, July 2002 through December 2013, (138 months.) There are 1252 assets of all types available; they are assigned to five roughly equal sized groups of 250, 250, 250, 251, and 251 so that the number of assets in each group exceeds the time series sample size, which permits the calculation of estimated SDFs for each group separately. The composition of each group is reported in the second part of the table. The 956 available equities are assigned randomly to groups and mingled with all available assets of another type in groups 2-5. Differences in estimated SDFs across asset groups are tested for stochastic dominance with the non-parametric Kruskal/Wallis (1952) statistic. Means and variances are compared with, respectively, the Welch (1951) and Brown/Forsythe (1974) tests. P-values are for the null hypothesis that the asset classes are all priced with the same SDFs. A low p-value rejects the null.

<table>
<thead>
<tr>
<th>Sample Period</th>
<th>Stochastic Dominance (Kruskal/Wallis)</th>
<th>Equal Means (Welch)</th>
<th>Equal Variances (Brown/Forsythe)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul ‘02-Dec ‘13</td>
<td>0.996</td>
<td>1.000</td>
<td>0.370</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>250 Equities</td>
</tr>
<tr>
<td>2</td>
<td>127 Equities and 123 Bonds</td>
</tr>
<tr>
<td>3</td>
<td>213 Equities and 37 Currencies</td>
</tr>
<tr>
<td>4</td>
<td>204 Equities and 47 Commodities</td>
</tr>
<tr>
<td>5</td>
<td>162 Equities and 89 REITs</td>
</tr>
</tbody>
</table>
Tests of the SDF Theory With Larger Samples of Equities

Stochastic discount factors (SDFs) are estimated for two groupings of equities using simultaneous monthly gross return observations, July 2002 through December 2013, (138 months.) In Panel A, 956 low-leverage equities are randomly assigned to two groups of 478 each. In Panel B, 956 low-leverage equities are compared with 956 randomly-selected equities with typical leverage in their capital structures. Differences in estimated SDFs across asset groups are tested for stochastic dominance with the non-parametric Kruskal/Wallis (1952) statistic. Means and variances are compared with, respectively, the Welch (1951) and Brown/Forsythe (1974) tests. Low p-values in the table would reject the null hypothesis that all groups are priced with the same SDFs.

<table>
<thead>
<tr>
<th>Sample Period</th>
<th>Stochastic Dominance (Kruskal/Wallis)</th>
<th>Equal Means (Welch)</th>
<th>Equal Variances (Brown/Forsythe)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul ‘02-Dec ‘13</td>
<td>A: Two groups of 478 low-leverage equities</td>
<td>0.547</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>B: Low- vs. high-leverage groups of 956 equities each</td>
<td>0.679</td>
<td>0.995</td>
</tr>
</tbody>
</table>
Table 8
Simulated Performance Information for the SDF Estimator with Thick-Tailed Returns

We simulate true SDFs with mean = 1/(1+riskless interest rate) and various time series volatilities. Gross asset returns are simulated so that their mean values multiplied by the SDFs are equal to 1.0, but errors perturb their sample values. The errors are generated from a Cauchy distribution with various scale parameters and truncation that retains only the middle 95%. The performance of the SDF estimator is measured by the correlation between true and sample SDFs and by Theil’s (1966) $U_2$ statistic, which is closely related to the mean square prediction perturbation. Linear regressions are reported in Panel A where the dependent variable is the correlation and in Panel B where the dependent variable is $U_2$. In Panel C, the dependent variable is the sample time series standard deviation of the estimated SDFs. Panel D reports the implied riskless rate from the reciprocals of the estimated SDFs. There are 2,880 parameter combinations, each with an independently-simulated set of true SDFs and returns.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Correlation between true and estimated SDFs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>-2.129E-03</td>
<td>-30.089</td>
</tr>
<tr>
<td>N, Assets</td>
<td>2.767E-04</td>
<td>31.286</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>1.474</td>
<td>34.709</td>
</tr>
<tr>
<td>Perturbation Scale</td>
<td>-0.1788</td>
<td>-97.287</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>-0.5534</td>
<td>-0.330</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.813</td>
</tr>
<tr>
<td>B: $U_2$ from comparing true and estimated SDFs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>1.772E-03</td>
<td>59.132</td>
</tr>
<tr>
<td>N, Assets</td>
<td>-2.483E-04</td>
<td>-66.289</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>0.1337</td>
<td>7.434</td>
</tr>
<tr>
<td>Perturbation Scale</td>
<td>6.895</td>
<td>88.548</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>-0.4715</td>
<td>-0.663</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.846</td>
</tr>
<tr>
<td>C: Standard Deviation of Estimated SDFs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>2.049E-03</td>
<td>47.071</td>
</tr>
<tr>
<td>N, Assets</td>
<td>-2.946E-04</td>
<td>-54.145</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>0.3602</td>
<td>13.793</td>
</tr>
<tr>
<td>Perturbation Scale</td>
<td>4.646</td>
<td>41.088</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>-1.518</td>
<td>-1.471</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.709</td>
</tr>
<tr>
<td>D: Riskless Rate Inferred from Estimated SDFs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>-1.783E-05</td>
<td>-1.277</td>
</tr>
<tr>
<td>N, Assets</td>
<td>1.906E-06</td>
<td>1.092</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>6.402E-03</td>
<td>0.764</td>
</tr>
<tr>
<td>Perturbation Scale</td>
<td>-2.588E-02</td>
<td>-0.713</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>1.395</td>
<td>4.212</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.00600</td>
</tr>
</tbody>
</table>
We simulate true SDFs with mean $= 1/(1 + \text{riskless interest rate})$ and various time series volatilities. Gross asset returns are simulated so that their mean values multiplied by the SDFs are equal to 1.0, but errors perturb their sample values. The initial returns are lognormal and generated by an underlying one-factor market model with a dispersion in betas, a market index whose mean exceeds the riskfree rate by 0.6% per month and has a volatility of 4% per month. The market model R-square is 0.2. The performance of the SDF estimator is measured by the correlation between true and sample SDFs and by Theil’s (1966) $U_2$ statistic, which is closely related to the mean square prediction error. Linear regressions are reported in Panel A where the dependent variable is the correlation and in Panel B where the dependent variable is $U_2$. In Panel C, the dependent variable is the sample time series standard deviation of the estimated SDFs. Panel D reports the implied riskless rate from the reciprocals of the estimated SDFs. There are 2,880 parameter combinations, each with an independently-simulated set of true SDFs and returns, including betas, market returns and idiosyncratic returns.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A: Correlation between true and estimated SDFs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>-1.160E-03</td>
<td>-11.621</td>
</tr>
<tr>
<td>N, Assets</td>
<td>1.459E-04</td>
<td>11.692</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>1.263</td>
<td>18.161</td>
</tr>
<tr>
<td>Perturbation Volatility</td>
<td>-1.702</td>
<td>-48.219</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>-1.532</td>
<td>-0.647</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.479</td>
</tr>
<tr>
<td><strong>B: $U_2$ from comparing true and estimated SDFs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>2.093E-03</td>
<td>55.377</td>
</tr>
<tr>
<td>N, Assets</td>
<td>-2.912E-04</td>
<td>-61.646</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>0.162</td>
<td>6.171</td>
</tr>
<tr>
<td>Perturbation Volatility</td>
<td>0.620</td>
<td>46.421</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>0.308</td>
<td>0.344</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.780</td>
</tr>
<tr>
<td><strong>C: Standard Deviation of Estimated SDFs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>2.948E-03</td>
<td>54.193</td>
</tr>
<tr>
<td>N, Assets</td>
<td>-4.210E-04</td>
<td>-61.925</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>8.702E-02</td>
<td>2.297</td>
</tr>
<tr>
<td>Perturbation Volatility</td>
<td>0.496</td>
<td>25.778</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>-0.605</td>
<td>-0.469</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.729</td>
</tr>
<tr>
<td><strong>D: Riskless Rate Inferred from Estimated SDFs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, Time Periods</td>
<td>-1.796E-05</td>
<td>-1.330</td>
</tr>
<tr>
<td>N, Assets</td>
<td>3.682E-06</td>
<td>2.181</td>
</tr>
<tr>
<td>True SDF Volatility</td>
<td>1.946E-02</td>
<td>2.069</td>
</tr>
<tr>
<td>Perturbation Volatility</td>
<td>5.584E-03</td>
<td>1.169</td>
</tr>
<tr>
<td>Riskless Rate</td>
<td>1.419</td>
<td>4.430</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td>0.011</td>
</tr>
</tbody>
</table>


Araujo, Fabio, Joao Victor Issler, and Marcelo Fernandes, 2005, Estimating the Stochastic Discount Factor without a Utility Function, working paper, (Graduate School of Economics EPGE Getulio Vargas Foundation, March.)

Araujo, Fabio, and Joao Victor Issler, 2011, A Stochastic discount factor approach to asset pricing using panel data asymptotics, (Graduate School of Economics EPGE Getulio Vargas Foundation.)


Welch, B. L., 1951, On the Comparison of Several Mean Values: An Alternative Approach, Biometrika 38, 330-336.

To demonstrate the SDF estimator, the perturbation in equation (17) of the text is set to a very small value, .01% per period. The true SDF has a mean dictated by a riskless rate of .4% per period and its standard deviation is 4% per period. Returns have a mean and standard deviation per period of .8% and 8%, respectively. The number of assets, N, is 120 and the number of time periods, T, is 60, so there are sixty estimated and true SDFs plotted.
Two groups, each with \(N\) individual assets, have \(T\) simultaneous time series observations. The true stochastic discount factors (SDFs) are not the same for the two groups, so the SDF theory is false. Our SDF estimator is computed from the sample return observations in each asset group and the two SDF estimates are compared with the Kruskal/Wallis (KW) test, the Welch (WE) test and the Brown/Forsythe (BF) test. Test power is the frequency of correct rejections of the null hypothesis (\(H_0: \text{no difference in the SDFs}\)) in 1,000 replications with a type I error of five percent. The figure reports power for the KW(\(k\%\)) and WE(\(k\%\)) tests, where \(k\%\) is the perturbation volatility in standard deviation per period. The SDFs have different means determined by riskless rates of .001 and .05 per period. The SDFs have the same volatility, so the BF test has no power.
Figure 3

Test Power for Stochastic Discount Factors with the Same Mean but Different Volatilities

Two groups, each with N individual assets, have T simultaneous time series observations. The true stochastic discount factors (SDFs) are not the same for the two groups, so the SDF theory is false. Our SDF estimator is computed from the sample return observations in each asset group and the two SDF estimates are compared with the Kruskak/Wallis (KW) test, the Welch (WE) test and the Brown/Forsythe (BF) test. Test power is the frequency of correct rejections of the null hypothesis ($H_0$: no difference in the SDFs) in 1,000 replications with a type I error of five percent. The figure reports power for the BF(k%) where k% is the perturbation volatility in standard deviation per period. The SDFs have the same mean, .001, but volatilities of 0.01 and 0.002 (standard deviations per period.) The KW and WE tests have no power in this case.
Two groups, each with $N$ individual assets, have $T$ simultaneous time series observations. The true stochastic discount factors (SDFs) are not the same for the two groups, so the SDF theory is false. Our SDF estimator is computed from the sample return observations in each asset group and the two SDF estimates are compared with the Kruskal/Wallis (KW) test, the Brown/Forsythe (BF) test and the Welch (WE) test. Test power is the frequency of correct rejections of the null hypothesis ($H_0$: no difference in the SDFs) in 1,000 replications with a type I error of five percent. The figure reports power for the KW($k\%$), BF($k\%$), WE($k\%$) tests, where $k\%$ is the perturbation volatility in standard deviation per period. The SDFs have different means determined by the reciprocals of unity plus riskless rates of .001 and .05 per period and different volatilities of 0.02 and 0.1 (standard deviations per period.) The second SDF has a lower mean and higher volatility, so it is stochastically dominated by the first SDF. There are 16 choices of $N$ and $T$ for each test, arranged from left to right as follows: \{(T=30, 60, 90, 120), for each $N=240, 480, 720, 960$\}
Figure 5

Test Power for Five Stochastic Discount Factors, One with a Different Mean and Volatility

Five groups, each with N individual assets, have T simultaneous time series observations. The true stochastic discount factors (SDFs) are the same for four of the groups, but one group has a stochastically dominant SDF, so the SDF theory is false. Our SDF estimator is computed from the sample return observations in each asset group and the five SDF estimates are compared with the Kruskal/Wallis (KW) test, the Brown/Forsythe (BF) test and the Welch (WE) test. Test power is the frequency of correct rejections of the null hypothesis (H₀: no difference in the SDFs) in 1,000 replications with a type I error of five percent. The figure reports power for the KW(k%), BF(k%), WE(k%) tests, where k% is the perturbation volatility in standard deviation per period. Four groups have SDFs with means equal to the reciprocal of unity plus a riskless rate of .05 per period and a volatility of .1 (standard deviation per period.) One group has an SDF with mean equal to the reciprocal of unity plus a riskless rate of .001 per period and a volatility of .02 (standard deviation per period), which makes it stochastically dominant. There are 16 choices of N and T for each test, arranged from left to right as follows: \{(T=30, 60, 90, 120), for each N=240, 480, 720, 960.\}
Two groups of equities, each with 956 individual firms, are used to estimate Stochastic Discount Factors (SDFs) with data from July 2002 through December 2013. One group was selected to have the lowest leverage ratios among all available firms with full information over the 138 sample months. The other group is randomly selected from other firms and hence has higher leverage. The average leverage ratio for the first (second) group is 10.2% (32.5%) book debt divided by total assets. The estimated SDFs from each group are adjusted so that their time series standard deviations are equal to the implied standard deviation of the true SDF, which according to SDF theory and consistent with the tests in section IV.C, is the same for the two groups. The plot depicts 12-month moving averages centered on the first day of the labeled month.