

# An Agnostic and Practically Useful Estimator of the Stochastic Discount Factor

By

Kuntara Pukthuanthong and Richard Roll

May 11, 2020

## Abstract

We propose an estimator for the stochastic discount factor (SDF) that does not require macroeconomic proxies or preference assumptions. It depends only on observed asset returns, yet is immune to the form of the multivariate return distribution, including the distribution's factor structure. Using US equity data, our estimator satisfies the Hansen/Jagannathan bounds.

Authors' Coordinates	
Pukthuanthong <sup>1</sup>	Roll
Department of Finance University of Missouri Columbia, MO 65211	Caltech Pasadena, CA 91125 & Emeritus, UCLA
573-884-9785	626-395-3890
<a href="mailto:pukthuanthongk@missouri.edu">pukthuanthongk@missouri.edu</a>	<a href="mailto:rroll@caltech.edu">rroll@caltech.edu</a>

## Acknowledgement

For many helpful comments and suggestions on previous drafts, we are indebted to Viral Acharya, Tobias Adrian, Serena Agoro-Menyang, Yakov Amihud, Antonio Bernardo, Michael Brennan, Menachem Brenner, Mikhail Chernov, Bhagwan Chowdhry, John Cochrane, Harry DeAngelo, Eric DeBodt, Federico Echenique, Robert Engle, Michael Ewens, Wayne Ferson, Steve Figlewski, Ben Gillen, Andrei Gonçaves, Lars Hansen, Jens Jackwerth, Raymond Kan, Shiki Levy, Francis Longstaff, John McConnell, Hugh McCulloch, Tyler Muir, Andrew Patton, Michael O'Doherty, Stephen Ross, Grigory Vilkov, Kenneth Winston, Rui Yu, Guofu Zhou, and participants in the Caltech and UCLA Brown Bag seminars, the Amundi Paris Dauphine annual workshop and the Stern School Volatility Institute Seminar at New York University.

---

<sup>1</sup> Corresponding author

We propose an estimator of the Stochastic Discount Factor (SDF) inspired by a recognition that the SDF appears in a particular mathematical object, an integral equation.

The solution to this integral equation makes our proposed estimator novel in several respects. First, it does not depend on macroeconomic proxies or preferences, unlike some estimators in the previous literature. It is constructed from observed asset returns only, which is why we call it a “practical” and “agnostic” estimator. Second, in contrast to typical portfolio applications such as mean/variance analysis, our estimator requires an unconventional condition: the number of assets must exceed the number of time period observations, which allows for a broader application to shorter time samples. Third, although the estimator is a function of observed returns, it does not depend on the distribution of returns. It applies to single- and multi-factor data, any asset class, and for thick or thin tails. This is an important implication of the integral equation property of the SDF and is not simply window dressing, misplaced or irrelevant. Finally, our estimator is immune to the grouping of assets. If  $N$  assets share a common SDF, the SDF estimator will be statistically indistinguishable when derived from the  $N$  assets as a whole or from subsets of size  $N/2$ ,  $N/3$ , etc.

Although our estimator is quite easy to compute and is asymptotically consistent, its sampling distribution is complex for reasons that will be fully explained below. Hence, we conduct a battery of simulations to disclose its small sample properties and its accuracy. As with any estimator, noise in the data degrades accuracy; we show how larger sample sizes (of assets and time periods) can offset this degradation.

Before presenting a formal derivation of the estimator, we embark on a brief literature review that places it in the context of previous work. This is followed by sections containing the derivation, remarks on the sampling distribution, simulations, and finally empirical applications. A summary concludes.

## **I. Previous SDF Literature**

The Stochastic Discount Factor (SDF) has become a dominant paradigm in recent asset pricing literature. For example, Ferson (1995) shows how the main asset pricing results (mean/variance efficiency, multi-beta models) are special cases of the basic SDF relation. Cochrane (2005) begins with the SDF relation in chapter 1 and expands it into almost all other known models of assets. Bossaerts (2002) relies on the SDF paradigm in criticizing empirical work on asset pricing. The SDF 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. Excellent reviews are provided by Ferson (1995) and Cochrane and Culp (2003).

The empirical success of the SDF approach is less apparent. In many (but not all) previous empirical applications, the SDF is proxied by a construct that depends typically on aggregate consumption, but occasionally on some other macroeconomic quantity, combined with a risk aversion parameter. For example, 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 (2005, 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 (2001) add macro variables such as labor income and find that 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 several 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.) Da and Yun (2017) employ electricity generation as a proxy for aggregate consumption.<sup>1</sup> Adrian, Crump and Moench (2013) employ an exponential function of a grouping of state variables, which are themselves principal components of Treasury bond returns.

In research published just 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.)

Recognizing that aggregate consumption changes are too "smooth" to be well connected 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 Ait-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.

---

<sup>1</sup> See also the variety of specifications discussed by Cochrane and Hansen (1992) in Section III, "Other Candidate Discount Factors."

It is widely understood that SDF is (by definition) whatever process generates the equality in (1). 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 is unique. A sample of more recent literatures gives a better sense of how widespread this understanding is. Dew-Becker and Giglio (2016) adopt a frequency domain approach and a very general specification for the SDF process (their equation 1). Then they discuss “consumption-based models” with several different types of preferences including Power Utility (Section 2.1), Habit formation (2.2), and Epstein-Zin (2.3.) Sections 3 and 4 of their paper provide more general approaches but they use consumption growth as an observable factor and/or Epstein-Zin preferences.

Alvarez and Jermann (2005) use this agnostic representation to study how volatile permanent shocks to the SDF process must be; nonetheless, their objective is to derive a bound on the permanent component of the SDF. To estimate the SDF, they suggest a representative agent with recursive preferences and a particular consumption process (see p. 2004.) Martin (2017) places minimal restrictions on the features of the SDF such that option prices allow us to recover forward-looking expected returns. He makes a significant contribution by deducing certain properties of the SDF, particularly that it must be highly volatile and negatively associated with returns. To determine bounds or likely volatilities and correlations are valuable insights because they represent properties that the true SDF should possess and consequently suggest tests for any SDF estimator. Alvarez and Jermann (2005) and Martin (2017) offer such insights but do not constitute an estimator per se.

Ross (2015) proposes to recover the full forward-looking distribution of the SDF in an agnostic manner using derivative prices. He motivates his result economically in terms of the marginal rate of substitution between current and future consumption. This is completely compatible with our own approach, which is to propose an empirical estimator for the SDF that depends only on returns. Ross’s objective is different from ours. He succeeds in recovering the objective probability distribution while we propose an estimator of the SDF that depends only on asset returns and can be easily implemented. In addition, we propose to estimate the past realizations of the SDF during a given historical period, not an estimate of the future SDF as in Ross’s paper. Our estimator does not require option prices, preferences, state prices, macroeconomic proxies, or anything other than a sample of asset returns.

Campbell (1993) surmounts the annoyance of preference specifications 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 (2005, 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 (2011, 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 relatively low risk aversion parameters are consistent with their estimated SDFs. They are able to price some stocks successfully, but not stocks with low capitalization levels.

Both AIF and AI essentially assume that the SDF is unique, rejecting that proposition only indirectly in the case of AI with low cap stocks. Our SDF estimator does not depend on a factor model or a logarithmic approximation, or any other structural condition. Also, it applies regardless of the multivariate distribution of returns, whatever its form, provided that certain lower order moments exist.<sup>2</sup>Kozak et al. (2018) apply several characteristics and suggest different methodologies to obtain a robust SDF from various traded factors, which achieves the bound by Hansen and Jagannathan (1991).

## **II. An Agnostic Estimator for the SDF**

This section first shows (in Section II.A) how SDFs can be approximated by a transformation of returns, without any additional information about preferences, consumption or other macro-economic data. The following Section (II.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. Section II.C contains a proof that the estimator is asymptotically consistent and Section II.D shows an alternative proof of consistency. Section II.E provides demonstrations of concept in small samples; using simulations that are limited in scope, it illustrates that the estimator works well regardless of the underlying distributions of returns, including their factor structure. Finally, Section II.F provides more detailed simulations that provide insight into the estimator's likely accuracy in future applications.

### **II.A. Estimating the SDF from Returns Alone**

---

<sup>2</sup> We explain the required moments below.

Let  $p_{i,t}$  denote the cash value of asset  $i$  at time  $t$ . When markets are complete, the SDF paradigm implies the existence of a unique  $m_t$ , the SDF, such that

$$E_{t-1}(\tilde{m}_t \tilde{p}_{i,t}) = p_{i,t-1} \quad \forall i,t.^3 \quad (1)$$

Denoting a gross return between  $t-1$  and  $t$  by  $R_{i,t} \equiv p_{i,t}/p_{i,t-1}$ , equation (1) is the same as

$$E_{t-1}(\tilde{m}_t \tilde{R}_{i,t}) = 1 \quad \forall i,t.^4 \quad (2)$$

In the SDF literature on consumption-based asset pricing, Eq. (2) was originally derived and the SDF is considered a function of preferences and consumption, often as held by a representative investor. From a mathematical perspective, however, Eq. (2) is a special case of an Integral or Volterra equation and, indeed, it is the simplest of these, called a ‘‘Fredholm equation of the first type.’’<sup>5</sup> As such,  $m_t$  is simply an unknown mathematical function, not necessarily related to anything economic including the probability distribution of  $R$ . This is why it should be obvious from the outset that estimators of the SDF could, but need not, depend on the multivariate form of the distribution of returns.<sup>6</sup>

Integral equations often do not have analytic solutions, so mathematicians and physicists solve them numerically, typically by a ‘‘quadrature rule’’ whereby a system of equations with an equal number of unknowns provides a set of discrete values for the unknown function, which in our application would be some set of observations  $m_t$  for, say,  $t=1, \dots, T$ . We are proposing an analogous approach, discretizing as usual but with an over-identified system whose solution is rendered unique by a statistical restriction on the error of estimation.

We begin by noting that the expectation in Eq. (2), must correspond to a realization at time  $t$ ; i.e.,

$$m_t R_{i,t} = E_{t-1}(\tilde{m}_t \tilde{R}_{i,t}) + \varepsilon_{i,t} \quad (3)$$

where  $\varepsilon_{i,t}$  denotes the (complete) surprise in the  $mR$  product for asset  $i$  in period  $t$ . For each time period  $t$ , the realization in (3) is determined by whatever state occurs among the many encapsulated in the

<sup>3</sup> 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  $t-1$ .

<sup>4</sup> Eq. (2) is the only moment condition required by SDF theory. However, the basic SDF relation applies similarly to multiple periods; e.g.,  $E_t(\tilde{m}_{t+\tau} \tilde{R}_{i,t+\tau}) = 1$  for  $\tau > 1$  where the gross return and  $m$  span  $\tau$  periods. This could provide some interesting features involving a term structure of SDFs but we do not explore that possibility in this paper.

<sup>5</sup> We are grateful to Francis Longstaff for pointing out this isomorphism. See also Polyanin and Manzhirov (1998). It is implied in McCulloch (2003) who shows that the SDF (or ‘‘pricing kernel’’) has finite payoffs even when returns follow stable laws whose second moments are infinite.

<sup>6</sup> In terms of the asymptotic distribution of the estimator, we are not specific about it because there is no single distribution that would apply in all cases. Our estimator is a function of observed returns; thus, it would have their distribution. If returns had a Gaussian, or instance, so would the estimator. Returns might also have a more complex distribution such as being driven by a factor structure. Perhaps returns have fat tails, etc., etc

expectation (2). The surprise is complete if expectations are rational; i.e., if agents can freely change their expectation in response to new information.

We hasten to add that the surprise is complete only in a time series sense, for period  $t$ , but the surprises are very likely to be correlated across assets at  $t$ . There are two reasons: first, returns themselves are usually cross-sectionally correlated,  $\text{Cov}(R_{i,t}, R_{j,t}) \neq 0$ , for  $j \neq i$  and second, the products in (3) have a common element, the SDF itself, so  $\text{Cov}(\tilde{m}_t \tilde{R}_{i,t}, \tilde{m}_t \tilde{R}_{j,t}) \neq 0$ , unless there is some unusual correlation between  $m$  and the two return.<sup>7</sup> However, we mostly rely on the time series independence of the surprises; their cross-sectional correlation is not pertinent for this reliance. **Moreover, the purpose of our extended derivation on pp. 10-13** and of the simulations is to assess the consequences, if any, of this difficulty, for the standard deviation of our estimator (as opposed to its mean and its deviation from the true SDF.)

Since there is a state realization for each time  $t$ , over  $T$  time periods, we have, from Eqs. (3) and (2),

$$\frac{1}{T} \sum_{t=1}^T m_t R_{i,t} = \frac{1}{T} \sum_{t=1}^T [E_{t-1}(\tilde{m}_t \tilde{R}_{i,t})] + \frac{1}{T} \sum_{t=1}^T \varepsilon_{i,t} = 1 + \frac{1}{T} \sum_{t=1}^T \varepsilon_{i,t} \cong 1 \quad (4)$$

where the approximation indicates that the average surprise is not exactly zero in a finite sample, though it should vanish for each asset separately as  $T \rightarrow \infty$ . The summation on the left of Eq. (4) is essentially the sample analog of the expectation on the left of Eq. (2), which implies that its approximation will improve as  $T$  increases.

The approximation error in Eq. (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

$$\bar{\varepsilon}_i \equiv \frac{1}{T} \sum_{t=1}^T \varepsilon_{i,t}$$

Rational expectations rules out any serial dependence in the surprises,

$$\text{Cov}(\varepsilon_{i,t}, \varepsilon_{i,t-j}) = 0, \quad j \neq 0$$

but the surprises could be heteroscedastic. Hence,

$$\text{Var}(\bar{\varepsilon}_i) = \frac{1}{T^2} \sum_{t=1}^T \text{Var}_t(\varepsilon_{i,t}) = \frac{1}{T} \bar{\sigma}_i^2$$

where  $\bar{\sigma}_i^2$  denotes the mean variance of surprises for asset  $i$  over the particular sample period,  $t=1, \dots, T$ . Unless the mean variance is growing without bound, the approximation error should disappear as  $T$  grows larger.

---

<sup>7</sup> We are grateful to a previous referee for pointing out this likely cross-sectional correlation.

As mentioned earlier, the individual surprises in (3) are likely to be cross-sectionally correlated at time  $t$ , so the average surprised are likely to be as well.

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  $\mathbf{R}$  (hereafter boldface denotes a matrix or vector). There are  $N$  columns in  $\mathbf{R}$  and the  $i^{\text{th}}$  column is  $[R_{i,1}:\dots:R_{i,T}]'$ . We also need a column vector  $\mathbf{m} \equiv [m_1:\dots:m_T]'$  to hold  $T$  realized values of the SDF and a  $N$ -element column unit vector  $\mathbf{1} \equiv [1:\dots:1]'$ . The entire SDF ensemble of realizations for all assets and periods can then be written compactly as

$$\mathbf{R}'\mathbf{m}/T \cong \mathbf{1}. \quad (5)$$

Pre-multiply Eq. (5) by  $\mathbf{R}$ , to obtain

$$(\mathbf{R}\mathbf{R}')\mathbf{m}/T \cong \mathbf{R}\mathbf{1}.$$

Since we have chosen  $N > T$ , the cross-sectional time-product matrix  $\mathbf{R}\mathbf{R}'$  is non-singular unless there are two or more periods with linearly dependent cross-sectional vectors of returns.<sup>8</sup> Hence, we can usually solve for a time-varying vector of estimated stochastic discount factors as

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

*N.B.: It is very important to emphasize that our solution (6) absolutely requires the number of assets to exceed the number of time periods; i.e.,  $N > T$ . Many comments on earlier drafts make it clear that this condition, which is unusual and perhaps unprecedented in finance, is hard to grasp. Yet it is essential. It is not possible to uncover a unique vector of SDF realizations if  $T > N$ , which is the familiar condition in most other contexts, such as computing non-singular covariance matrices. We MUST have  $N > T$  to obtain a unique  $\mathbf{m}$ . We hasten to add that this is merely a sample requirement and hence is easy to satisfy; e.g., by reducing  $T$  until it falls below  $N$ . The condition does not imply anything egregious such as the existence of an arbitrage because we are simply estimating  $T$  sample realizations of  $m$ , not the entire state space of  $m$  in each time period  $t$ .<sup>9</sup>*

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

$$(\mathbf{R}\mathbf{R}')^{-1}\mathbf{R}\bar{\boldsymbol{\varepsilon}}. \quad (7)$$

<sup>8</sup> That is, unless the return of every individual asset in a given period is a linear function of the returns on that asset in another or several other periods, (not that the returns are linearly dependent relative to each other in a given period.)

<sup>9</sup> A large cross-section of stocks is also advocated by Pelger and Lattau (2017) and used for a different purpose, estimating risk factors.

This error is not exactly zero because, for each  $t$ , there are related components in  $\mathbf{R}$  and  $\bar{\boldsymbol{\varepsilon}}$ . For very large  $N$  and  $T$ , these components should become immaterial, but they add sampling error to the estimated SDFs with smaller  $N$  and  $T$ . We investigate the consequences in the next section after presenting an alternative approach for deriving the same estimator.

Hansen and Jagannathan (1991, p. 233) derive an expression that appears similar to Eq. (6), but the resemblance is superficial.<sup>10</sup> Their expression involves a covariance matrix of payoffs (or returns). Our  $\mathbf{R}\mathbf{R}'$  is not a covariance matrix. They note that their solution involves the first and second moments of the future payoffs and prices. If  $\mathbf{R}\mathbf{R}'$  above were diagonal, Eq. (6) would also involve first and second moments but in this case the (sample) moments would be the cross-sectional mean return in each period divided by the cross-sectional mean of the individual squared returns in that period.

Kim and Korajczyk (2019), after seeing our estimator in an earlier draft of this paper, modify it by adding the assumption that returns and the SDF itself are all generated by the same linear factor model. They argue that this additional structure reduces estimation error and they provide some supportive simulation evidence. But the assumption that all assets and the SDF itself follow an identical factor process is obviously not the only possibility or even the most realistic one. Some assets might not conform to a factor structure at all and the number of factors need not be the same for each and every asset.<sup>11</sup> Moreover, there seems a little reason to think that the SDF would necessarily follow such a process. Even if it does not, the basic SDF integral equation (2) must prevail for all assets within an integrated market since in this case the SDF is unique.<sup>12</sup>

## II.B. The Minimum Sum of Squared Average Surprises

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

$$\mathbf{R}'\mathbf{m} / T = \mathbf{1} + \bar{\boldsymbol{\varepsilon}} \quad (8)$$

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

---

<sup>10</sup> The Hansen/Jagannathan approach is implemented for performance measurement by Chen and Knez (1996) and is further refined by He, Ng, and Zhang (1999.)

<sup>11</sup> This would be particularly problematic when testing whether different groups of assets are in the same (integrated) market, as we do later in this paper. Two groups of assets could have different factor structures yet still be integrated. The Kim Korajczyk approach, however, could conclude that they are not.

<sup>12</sup> See Back (2017, pp. 56-57), for the joint conditions about the existence of a unique SDF, no arbitrage opportunities, and the law of one price.

$$\min_{\mathbf{m}} (\bar{\boldsymbol{\varepsilon}}' \bar{\boldsymbol{\varepsilon}}) = (\mathbf{R}'\mathbf{m} / T - \mathbf{1})'(\mathbf{R}'\mathbf{m} / T - \mathbf{1}).$$

The first-order condition is

$$\frac{\partial}{\partial \mathbf{m}} (\mathbf{m}'\mathbf{R}\mathbf{R}'\mathbf{m} / T^2 - 2\mathbf{m}'\mathbf{R}\mathbf{1} / T) = 2\mathbf{R}\mathbf{R}'\mathbf{m} / T^2 - 2\mathbf{R}\mathbf{1} / T = \mathbf{0}$$

and the extremum is achieved for the  $\hat{\mathbf{m}}$  that satisfies

$$\hat{\mathbf{m}} / T = (\mathbf{R}\mathbf{R}')^{-1}\mathbf{R}\mathbf{1} \quad (9)$$

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

One may legitimately question why the estimator in Eqs. (6) or (9) should involve a cross-sectional sum of returns ( $\mathbf{R}\mathbf{1}$ ) in each period.<sup>13</sup> Obviously, it is different from the estimator of Hansen and Jagannathan (1991) who derive an SDF that is equal to a mean/variance tangency portfolio plus an error term. Our estimator is dictated by the mathematical fact that the basic SDF Eq. (2) has a 1.0 on the right side for every asset in every period while the tangency portfolio can change over time. Perhaps this confers a benefit in terms of pricing the entire vector of SDF realizations over time; this is an empirical issue for future research.

Another possible question might occur to some readers in that the estimator seems to use information across all observed time periods even though for any given period  $t$  within  $T$ , the SDF is a random variable. But the answer is simply that we are estimating the best fit to the entire vector  $\hat{\mathbf{m}}$  whose elements have already occurred as realizations of the random variable at each  $t$ . There is an associated surprise each  $t$  as well and the estimator simply minimizes the sum of squared average surprises. This puzzlement is similar to the one we would be faced if we run a market model regression,  $R_{i,t} = \alpha + \beta R_{M,t} + \varepsilon_{i,t}$  ( $t=1, \dots, T$ ) to estimate the intercept and slope. In this regression, we would also use past and future observations in the regression.

The least squares estimator in Eq. (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), the estimator could be biased in finite samples.

---

<sup>13</sup> In the robustness section A-3, we consider an alternative, a precision weighted sum as opposed to a simple sum.

This setup is reminiscent of Britten-Jones (1999), who shows that a no-intercept regression of the unit vector (as dependent variable) on a matrix of returns (as independent variables) yields slope coefficients that are proportional to the weights of the sample mean/variance efficient tangency portfolio from the riskless rate. Britten-Jones notes, as we do also, that a dependent variable of all 1's induces a connection between the regression residuals and the explanatory variables which raises finite sample concerns.

The big difference between Britten-Jones' regression and our superficially similar equation (9) is that his is time series and ours is cross-sectional. As a consequence, his coefficients are proportional to cross-sectional portfolio weights while our coefficients constitute a time series vector of estimated SDF realizations.<sup>14</sup>

To elucidate this issue, solve Eq. (8) for  $\mathbf{1}$  and substitute the result in Eq. (9), which simplifies to,

$$\hat{\mathbf{m}} - \mathbf{m} = -\mathbf{T}(\mathbf{R}\mathbf{R}')^{-1}\mathbf{R}\bar{\boldsymbol{\varepsilon}}. \quad (10)$$

The expected value of this expression is the bias. Expanding  $\mathbf{R}\bar{\boldsymbol{\varepsilon}}$  term by term, we observe that most are close to zero because they involve products such as  $(\varepsilon_{j,t} R_{i,t-k})$  for  $i \neq j$  and  $k \neq 0$ . However, there are a few elements that are unlikely to disappear. For period  $t$ , there is

$$R_{1,t}\varepsilon_{1,t} + R_{2,t}\varepsilon_{2,t} + \dots + R_{N,t}\varepsilon_{N,t} = m_t(R_{1,t}^2 + R_{2,t}^2 + \dots + R_{N,t}^2) - \sum_{j=1}^N R_{j,t}$$

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  $\varepsilon$ 's) is not linear.

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

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{R}'\hat{\mathbf{m}}/\mathbf{T} - \mathbf{1} = \mathbf{R}'(\mathbf{R}\mathbf{R}')^{-1}\mathbf{R}\mathbf{1} - \mathbf{1} = -[\mathbf{I} - \mathbf{R}'(\mathbf{R}\mathbf{R}')^{-1}\mathbf{R}]\mathbf{1} \quad (11)$$

The matrix in brackets in Eq. (11) is idempotent, so the sum of squared residuals divided by the degrees-of-freedom,  $N-T$ , is

$$\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{N-T} = \frac{\mathbf{1}'[\mathbf{I} - \mathbf{R}'(\mathbf{R}\mathbf{R}')^{-1}\mathbf{R}]\mathbf{1}}{N-T} = \frac{N}{N-T} - \frac{\mathbf{1}'[\mathbf{R}'(\mathbf{R}\mathbf{R}')^{-1}\mathbf{R}]\mathbf{1}}{N-T} \quad (12)$$

---

<sup>14</sup> Interestingly, Britten-Jones foresaw that a similar approach might one day help estimate the SDF. He said, "Future research could attempt to analyze and identify stochastic discount factors using the statistical inference procedures developed here," (p. 657, footnote 7.) He was right and he was very close.

The mean squared residual in (12) clearly depends on both N and T and on their difference. Because the mean surprises are likely to be cross-sectionally dependent (recall the discussion on page 7 above), we do not have a simplified analytic expression for (12), but we do investigate it later with simulations..

The square root of Eq. (11) gives the standard error of the estimate,

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

The covariance matrix of the estimated SDFs is given by

$$E[(\hat{\mathbf{m}} - \mathbf{m})(\hat{\mathbf{m}} - \mathbf{m})' | \mathbf{R}, T] = (\mathbf{R}\mathbf{R}')^{-1} \mathbf{R} E(\mathbf{V}_{\Sigma_{\boldsymbol{\varepsilon}}}) \mathbf{R}' (\mathbf{R}\mathbf{R}')^{-1} \quad (13)$$

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

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

Unlike the analogous covariance matrix of disturbances in standard OLS regressions, the diagonal elements of  $\mathbf{V}_{\Sigma_{\boldsymbol{\varepsilon}}}$  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  $j^{\text{th}}$  row and  $k^{\text{th}}$  column of  $\mathbf{V}_{\Sigma_{\boldsymbol{\varepsilon}}}$  reduces to  $\sum_{t=1}^T \varepsilon_{j,t} \varepsilon_{k,t}$ . Moreover, if the  $\varepsilon$ 's are not correlated across assets, an arguably unlikely condition, this sum has an expected value of zero for  $j \neq k$  and then  $E(\mathbf{V}_{\Sigma_{\boldsymbol{\varepsilon}}})$  becomes diagonal and equal to  $\mathbf{I} \boldsymbol{\sigma}_{\Sigma_{\boldsymbol{\varepsilon}}}^2$  where  $\mathbf{I}$  is the identity matrix and  $\boldsymbol{\sigma}_{\Sigma_{\boldsymbol{\varepsilon}}}^2$  is the N element column vector whose  $j^{\text{th}}$  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 Eq. (13) simplifies further to

$$E[(\hat{\mathbf{m}} - \mathbf{m})(\hat{\mathbf{m}} - \mathbf{m})' | \mathbf{R}, T] = T \sigma^2 (\mathbf{R}\mathbf{R}')^{-1} \quad (14)$$

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 Eqs. (13) or (14) 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.

### II.C. Asymptotic Consistency of the Estimator

The previous section notes that any dependence between  $\mathbf{R}$  and  $\bar{\boldsymbol{\varepsilon}}$  is non-linear. Consequently, since  $E(\bar{\boldsymbol{\varepsilon}}) = \mathbf{0}$  by assumption, for any T large enough that the non-linear terms become vanishingly small.

$$p \lim_{N \rightarrow \infty} (\mathbf{R}\bar{\boldsymbol{\varepsilon}} | T) = \mathbf{0},$$

and hence, from (10) and Slutsky's theorem,

$$p \lim_{N \rightarrow \infty} (\hat{\mathbf{m}} - \mathbf{m}) = p \lim_{N \rightarrow \infty} [-T(\mathbf{R}\mathbf{R}')^{-1} \mathbf{R}\bar{\boldsymbol{\varepsilon}}] = p \lim_{N \rightarrow \infty} [-T(\mathbf{R}\mathbf{R}')^{-1}] p \lim_{N \rightarrow \infty} (\mathbf{R}\bar{\boldsymbol{\varepsilon}}) = \mathbf{0},$$

so the estimator is consistent for a fixed (but very large) T as N increases.

In a similar vein, make the simple rescaling,  $\bar{\mathbf{R}} \equiv \mathbf{R} / N$  and note that (10) can also be written as  $\hat{\mathbf{m}} = \mathbf{m} - (T/N)(\bar{\mathbf{R}}\bar{\mathbf{R}}')^{-1} \bar{\mathbf{R}}\bar{\boldsymbol{\varepsilon}}$ . In this case,  $p \lim_{T \rightarrow \infty} \frac{1}{N_T} \bar{\mathbf{R}}\bar{\boldsymbol{\varepsilon}} = \mathbf{0}$  as long as  $\bar{\mathbf{R}}$  and  $\bar{\boldsymbol{\varepsilon}}$  are linearly non-dependent (though they might share some non-linear dependencies.) Hence, again applying Slutsky's theorem,  $p \lim_{N, T \rightarrow \infty} (\hat{\mathbf{m}}) = \mathbf{m}$ .

It is noteworthy, if all asset returns are perfectly correlated with one another, then there is no guarantee that the required unbiasedness holds even in the limit. Such an assumption seems to be rarely stated in most of the finance literature simply because it's realistically impossible. On page 9, we note that there cannot be two or more periods with linearly dependent cross-sectional vectors of returns. See also footnote 8 on that page.

## II.D. Alternative Proof of Consistency

This alternative proof, presented in Kim et al (2019), also establishes the consistency of the agnostic estimator based on the following

**Assumptions:** As  $N, T \rightarrow \infty$ , (i)  $\frac{N}{T} \rightarrow \infty$ , (ii) there exists a positive constant  $c$  such that the minimum eigenvalue of  $\frac{\mathbf{R}'\mathbf{R}}{N}$  is larger than  $c$  and (iii)  $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} < \infty$ .

The first condition is that the size of cross-section increases faster than the size of the time-series, which is generally true for many data sets except, perhaps, with trade-to-trade observations. The second condition states that the minimum eigenvalue of  $\frac{\mathbf{R}'\mathbf{R}}{N}$  does not vanish, which can be rationalized as follows: Note that  $\frac{\mathbf{R}'\mathbf{R}}{N}$  is decomposed into a component due to systematic factors (whether this component exists or not) and a component due to idiosyncratic shocks. Since the minimum eigenvalue of  $\frac{\mathbf{R}'\mathbf{R}}{N}$  is larger than that of each component, the condition is guaranteed by simply imposing that the cross-sectional variance of idiosyncratic shocks do not vanish. The last condition is that aggregate mispricing in the economy is bounded, which is in parallel with APT of Ross (1976).

The agnostic SDF estimator  $\hat{\mathbf{m}}$  converges to the true SDF in mean squared error sense by exploiting a large panel return data.

**Theorem 2.1.** Under Assumption 1, as  $N, T \rightarrow \infty$ ,

$$\frac{1}{T}(\hat{\mathbf{m}} - \mathbf{m})'(\hat{\mathbf{m}} - \mathbf{m}) \xrightarrow{p} 0.$$

All proofs are in Appendix C or in Kim et al (2019).

## II.E. Demonstrations of Concept with Simple Examples

We have learned from comments on earlier drafts and in presentations that our proposed estimator is sometimes understood, erroneously, to be a projection on observed sample returns. Such intuition is understandable because the estimator does employ returns; hence, one could easily fall into the mistaken notion that the estimator is akin to a sample mean/variance efficient portfolio, which, of course, is composed differently across various sub-samples of assets.<sup>15</sup>

But a close examination of our estimator belies such intuition. Instead of a projection on asset returns, it is actually a projection on time periods. As a consequence, it does not depend on by the distributions of returns or even by their identity as long as a unique SDF prices all assets in the cross-section.

To demonstrate this property, we resort first to some unsophisticated (i.e., simplistic) simulations that subsume the potential small sample problems discussed in Section II.B above. These are not comprehensive but are merely intended to illustrate how our estimator performs with various levels of sampling volatility (noise). We then show that the estimator is immune to differences in the distributions of returns and extracts indistinguishable estimates of the SDF even from sub-samples of assets with different factor structures.

The next subsection (II.F) presents a more comprehensive battery of simulations.

Assuming that the SDF is unique, we generate “true” SDF realizations with a mean equal to the reciprocal of the gross riskless interest rate, as the SDF paradigm stipulates, and with a given level of time series variation about the mean. Specifically, we select a gross riskless rate,  $R_F$ , (1+the riskless return), and generate the SDF at time  $t$  as

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

---

<sup>15</sup> Such intuition is readily overturned by thinking about the SDF as an unknown function in an integral equation; see Section II.A.

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<sup>16</sup> and, in accordance with SDF theory and the absence of arbitrage, (15) provides a strictly positive  $m_t$ .

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

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

where  $\mu$  is the expected gross return (1 + the net return) and  $\sigma_\zeta$  is the standard deviation of the unscaled gross return  $\hat{R}$ .<sup>17</sup> We find in subsequent simulations that imposition of equal means and variances has an immaterial effect because the final scaled returns used in all subsequent calculations are computed as

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

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

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

However, and very importantly, because of the perturbations added as shown in (17), we emphasize that the sample average return and SDF product, (the expression within brackets above) is not exactly unity and differs from unity by an amount that varies across individual assets. As a consequence of this sampling perturbation, our estimator will not exactly match the true SDF in any time period or even on average except in the limit as the time series sample grows indefinitely larger. In other words, the expected SDF/Return product will be unity but not the sample SDF/Return product.

---

<sup>16</sup> In Appendix A.3, a robustness check considers non-normally distributed variation whose simulations are detailed later in this Appendix.

<sup>17</sup> Theoretically, the basic SDF equation usually implies negative correlation between the SDF and returns. This follows by noting that  $E(mR) = 1$  implies  $\text{Cov}(m,R) = 1 - E(R)E(m)$  and  $E(R)E(m) > 1$  because the risky return is usually larger than the riskless rate. However, Specification (16) does not take this into account because the correlation between  $R$  and  $m$  is zero. In our robustness section, however, Appendix Section A.1, we investigate an alternative specification wherein  $R$  and  $m$  are negatively correlated and find that our estimator is improved relative to the specification here.

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 in this simulation are uncorrelated with each other. We consider the consequences of this assumption below where we present analogous simulations with correlated returns that are generated by assets that conform to a factor structure.

The final simulation step uses the estimator (9) with the final returns from (17) to obtain  $\hat{m}_t$ ,  $(t=1, \dots, T)$ , for comparison with the true values from (15),  $m_t$ ,  $(t=1, \dots, T)$ .

Our first illustration of concept uses 120 assets and 60 time periods, (a modest degrees-of-freedom according to Section II.B), a riskless rate of .4% per period, and a true SDF standard deviation of 4% per period. In this first illustration, the standard deviation of the perturbation,  $\vartheta$  in (17), is intentionally small, .01% per period.

The minimum (maximum) individual return in (17) is -25.3% (35.6%). The average standard deviation is 8.1% over the 120 simulated assets with a minimum (maximum) individual asset standard deviation of 6.17% (11.2%). Thus, the returns have means and volatilities slightly larger but roughly comparable to actual US equity returns in percent per month.

Figure I plots the resulting estimated SDFs against the true SDFs for the 60 time periods. Their difference is trifling. Their correlation is 0.99946 and they are aligned with each other almost perfectly. This illustrates that the theoretical bias discussed in Section II.B is empirically trivial when the sampling perturbations are small.

In reality, of course, returns are correlated with one another and conceivably have heterogeneous factor structures across asset classes. For example, bond returns could be driven by different risk factors than equity returns. Nonetheless, if a unique SDF prices all asset expected returns in the cross-section, the basic SDF Eq. (1) is valid with the same SDF for all assets.

To consider this situation, we provide a further demonstration of concept by simulating returns that are not only correlated but also have diverse factor structures. In this simulation, we presume that there are two asset classes that share a common factor but that the second asset class is also driven by a second factor that has no influence on the first asset class.

Instead of the uncorrelated returns as in (16), we now have

$$\hat{R}_{i,t} = \exp[R_F + \beta_{i,1} f_{1,t} + \beta_{i,2} f_{2,t} + \zeta_{i,t} - \phi_i] \quad (18)$$

where the exponentiation correction factor is

$$\phi_i = \text{Var}(\beta_{i,1} f_{1,t} + \beta_{i,2} f_{2,t} + \zeta_{i,t}) / 2.$$

The mean return for each individual asset is dictated by the riskless rate,  $R_F$ , plus the mean of the first factor, which we assume is equal to a constant risk premium of .6% per period plus the riskless rate of .4% per period. The mean of the second factor is zero along with the mean of the idiosyncratic return,  $\zeta_{i,t}$ . The time series standard deviation is four percent per period for the factors and for the idiosyncratic return. The mean of the first factor  $\beta_{i,1}$  is 1.0 for both asset groups and the cross-sectional standard deviation is 0.1. For assets in the first class,  $\beta_{i,2} = 0, \dots \forall i$ , but only their mean is zero for assets in the second class and the cross-sectional standard deviation of their  $\beta_{i,2}$  is 0.1. The assets are arranged into two groups depending on their class and separate SDFs are estimated for each class.

The plots in Figure II illustrate the influence of two separate features: (a) noise and (b) an underlying factor structure of returns. Figure II, Panels A and B plots the estimated SDF against the true SDF in the left chart and the SDFs estimated separately for the two groups against each other in the right chart (for two different levels of return perturbation). As the figure shows, there is sampling variation, but the recovered estimate of the SDF is close to the true SDF and the estimated SDFs from the two divergent (by factor structure) are close to one another.

When the perturbation volatility is small (Panel A), the SDF equation holds extremely well in sample. When the volatility is larger (Panel B), there is substantial sampling error in the SDF estimator, though it appears to have little bias. In either case, regardless of the perturbation volatility, the SDFs estimated for the two asset classes with different factor structure are strongly related to one another.

Our third illustrative simulations assumes a two-factor structure for both groups of assets, but the factors are independent of each other across groups. In this case, both  $\beta$ 's in (18) are non-zero for most assets. The cross-sectional means of  $\beta_{i,1}$  and  $\beta_{i,2}$  are 1.0 and zero, respectively. Their cross-sectional standard deviations are both 0.1. Figure III shows the results using the higher level of return perturbation from Figure II, Panel B.

Again, despite the fact that the factors are entirely different in the two asset groups, there is a strong connection between the true and estimated SDFs and between the SDFs estimated from the two groups. This illustrates our contention that the distributions of returns are inconsequential for our SDF estimator provided that the true SDF is unique and prices all assets regardless of groupings.

Some might find these results surprising because our SDF estimator is unaffected by the return distribution. This could be particularly hard to fathom because a competing construct, a sample mean/variance efficient portfolio, also perfectly prices returns in the cross-section, but it obviously depends on the distribution of returns and has a different composition for various groups of assets. But examining carefully the basic SDF (2) reveals why our estimator is so robust. Equation (2) portrays a mathematical

object, an integral or “Volterra” equation, which says nothing about the distribution of returns other than the product of each return and the SDF has an expected value of unity. Consequently, every expected return obeys the same cross-sectional linear function of the covariance between the return and the SDF. So long as the first moment of the SDF/return product is finite and the SDF is unique, its estimator needs not be troubled by any other property of the multivariate distribution of returns.

## II.F. Augmented Simulations

To provide further insight about the performance of our SDF estimator, this section offers a series of simulations to compare true SDFs with estimated SDFs. Extending the examples in Section II.E above, we provide simulations for a wider set of parameters and sampling variation. The basic setup is identical to that in the previous section (II.E.)

In all cases, we compare the true and estimated SDFs using several criteria. The first criterion is the simple correlation between  $\mathbf{m}$  and  $\hat{\mathbf{m}}$ . Second is the Theil (1966)  $U_2$  statistic, which is closely related to the mean square prediction error, (MSE). Specifically,

$$\text{MSE} = \sum_{t=1}^T (\mathbf{m}_t - \hat{\mathbf{m}}_t)^2 / T, \text{ and}$$

$$U_2 = \text{MSE} / \left( \sum_{t=1}^T \mathbf{m}_t^2 / T \right).$$

The correlation is easy to understand but it can be a bit misleading because it fails to measure whether  $\mathbf{m}$  and  $\hat{\mathbf{m}}$  are congruent. For example, if  $\hat{\mathbf{m}} = 2\mathbf{m}$ , 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.,

$$\text{MSE} \equiv (\bar{\mathbf{m}} - \bar{\hat{\mathbf{m}}})^2 + (s_{\mathbf{m}} - s_{\hat{\mathbf{m}}})^2 + 2(1 - \rho)s_{\mathbf{m}}s_{\hat{\mathbf{m}}} \quad (19)$$

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

The third criteria is suggested by Kim and Korajczyk (2019). It involves a time series regression of the estimated SDF on the true SDF; i.e.,

$$\hat{\mathbf{m}}_t = a + b\mathbf{m}_t + \xi_t, \quad (20)$$

where  $\xi_t$  is a regression disturbance. A perfect estimator would exhibit  $a=0, b=1$  and a regression R-square of 1.0. The unadjusted regression R-square from (20) is the square of the simple correlation, so it need not be reported separately. A convenient estimate of relative bias is

$$(\hat{\bar{m}} - \bar{m}) / \bar{m} = a / \bar{m} + b - 1 = \hat{\bar{m}} / \bar{m} - 1. \quad (21)$$

In simulations with different levels of sampling 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, the volatility of the true SDF, and the risk-free rate. With these 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 generate the simulated  $\mathbf{m}$  and  $\hat{\mathbf{m}}$  with several different choices of the parameters and then present summary linear regressions of the correlations, Theil's  $U_2$ , and the regression measure of bias (21) against 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 likely minimal sample sizes and numbers of assets in subsequent empirical work by ourselves and others. Larger sample size generally improves the performance of any estimator, which is why we examine these and refer to them "minimal."

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  $\sigma_g$  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.

Table I presents the results, Panel A for the correlation between  $\mathbf{m}$  and  $\hat{\mathbf{m}}$ , Panel B for Theil's  $U_2$  and Panel C for the relative bias (21) in percent. Panel A shows that the correlation falls with  $T$ , rises with  $N$ , rises with  $\sigma_\xi$ , the volatility of the true SDF, and falls with  $\sigma_g$ , the perturbation volatility, all with 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.<sup>18</sup>

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, after decomposing the MSE into its three components, (19), 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 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.

This is consistent with the results in Panel C of Table I, which discloses how the estimator's relative bias varies with the parameters. The bias is larger (smaller) for larger  $T$  ( $N$ ), again showing the impact of degrees of freedom. There is virtually no effect of either the riskless rate or the volatility of the true SDF, much as one would anticipate. However, the perturbation volatility has a strong and negative influence on bias. The bias measure from (21) agrees completely with these results. On average over 2880 parameters, the bias (in percent) is -0.0449% with a cross-parameter standard deviation of 0.167%. The maximum (minimum) percentage bias is 1.502% (-1.046%). Clearly, the potential bias discussed in Section II.B above is not a critical issue, at least for the parameter values used in our simulations.

However, although the means of the true and estimated SDFs are typically close, the individual period-by-period estimated and true SDFs display substantial divergence for some parameter combinations.

---

<sup>18</sup> In unreported results, we verify that this is also true of the mean and variance of the initial returns as generated by Equations 16-18.

The correlation averaged over all parameter values is only .189 and the maximum and minimum correlations over the 2,880 parameter combinations are 0.951 and -0.547, respectively. Moreover, the mean, maximum and minimum of Theil's  $U_2$ , are, respectively, 0.339, 0.787, and 0.0359; larger values indicate more disagreement between the estimated SDF and the true SDF. There is also substantial cross-parameter dispersion in the regression coefficients  $a$  and  $b$  from (20). The mean, minimum, and maximum of  $a$  ( $b$ ) are 0.688, -.988, and 2.62 (0.309, -1.60, and 2.00.) These results make it very clear that the estimator performs poorly for some combinations of parameters.

To gain further insights into when the estimator performs well and when it does not, Table II reports two sub-samples of parameters, one consisting of 74 different parameter combinations chosen because every correlation between the true and estimated SDF is at least 0.8, or an R-square in (20) of at least 0.64; (we call this the “good performance” sub-sample). A matching sub-sample with the same number of different parameter combinations (74) is selected from the smallest correlations which all turn out to be negative (the “poor performance” sub-sample).<sup>19</sup>

As reported in Table II,  $N$  is relatively large (small) compared to  $T$  within the good (poor) performance sub-sample, which agrees with previous results. The very best performance is turned in by  $N=960$  and  $T=30$  but this is not the end of the story because the perturbation and true SDF volatilities also play a role. Notice that the perturbation volatility is always considerable smaller than the true SDF volatility when performance is good and the reverse is true for the absolute worst performance.

The good news here is that our SDF estimator performs very well for at least some combinations of parameters. Parameters of choice include  $N$  and  $T$ , so future investigators who adopt our estimator would be well advised to collect a sample with a large number of assets,  $N$ , relative to the length of the time series,  $T$ . However, it is clear from these results that an assertion in the paper by Kim and Korajczyk (2019) is not correct. They allege (p. 26, footnote 6) that our estimator requires “millions” of assets before it performs well. They are wrong for two reasons. First, the performance of our estimator is quite good even for less than 1,000 assets (see Table I). Second, the performance does not depend very much on the number of assets but rather on the perturbation volatility relative to the volatility of the true SDF (see Table II).

Admittedly, performance is degraded when perturbation volatility is large relative true SDF volatility and in empirical work (rather than with simulations) neither volatility is known. One does know, though, that the estimated SDF must have larger volatility than the true SDF while the estimated perturbation volatility must also be larger than the true perturbation volatility. This would allow some rough inference about the likely degradation of the SDF estimator.

---

<sup>19</sup> The correlation is negative for 21.3% of the 2880 parameter combinations, those that generate larger sampling error.

A better approach would be to estimate the SDF from two separate samples of assets that are presumed to be traded in an integrated market. Section IV.B below shows how perturbation volatility can be estimated in such a circumstance, using two separate samples that are first tested for being jointly integrated.

Panel D of Table I 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 D in Table I, using  $\sqrt{T}$  instead of  $T$  as a regressor, we find that virtually nothing is altered except the coefficient.<sup>20</sup>

In Panel E, of Table I, 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 the true SDF are marginally significant, which might be explained by Jensen's inequality (since the implied riskless rate is obtained from a reciprocal of an estimated SDF.)

We have generated tables corresponding to Tables I and II with returns that are generated by various multi-factor models. There are virtually no difference between these tables and Tables I and II, thus again confirming that the return generating process has no material influence on the SDF estimator.<sup>21</sup> Appendix A, Sections A.2 and A.3, reports analogous and very similar results even for thick-tailed distributions. It also provides some cursory evidence about a variance-weighted alternative estimator.

### III. Data

We collect monthly return observations on U.S. stocks over an intentionally limited recent time period, July 2002 through December 2013, 138 months in all. Two groups are selected, (1) 956 equities

---

<sup>20</sup> The  $t$ -statistic for  $\sqrt{T}$  is 53.8 as opposed to the 53.9 reported for  $T$  in Table I. Everything else is similarly close; e.g., the adjusted R-square is 0.730 as opposed to 0.731.

<sup>21</sup> The results are available upon request.

with lower leverage and (2) an equal-sized sample of higher leverage equities.<sup>22</sup> The sample sizes are dictated mainly by the requirement that simultaneous return observations be available every month. A longer time period (T) reduces the number of stocks with complete records and our estimator works best when the number of stocks (N) is larger than T. With N=956 (for each group) and T=138, we can have some confidence in the estimated SDFs produced for each group.

Groups of lower- and higher-leveraged equities are chosen so that their volatilities and probably their industries and other characteristics are diverse. This erects a hard hurdle for the proposition that the SDF is unique and the same for all equities. Any other dichotomy of individual assets besides leverage could be used in an identical test. We are just illustrating one possible test. We investigate this uniqueness with a battery of standard tests and find no significant evidence to contradict it. See Appendix B for these results.

#### IV. Further Empirical Findings.

##### IV.A. Disentangling SDF sampling error volatility and true SDF volatility

Bolstered by the results in Appendix B, we now assume that the SDF is in fact unique and is the same for lower- and higher-leveraged groups of US equities during the 138 calendar months composing the sample. This enables us to present some further empirical implications in this and the next subsections.

To disentangle the volatility of the true SDF from the sampling error volatility in the estimated SDF, denote by  $\hat{\mathbf{m}}(L)$  the vector of SDFs estimated from lower-leveraged stocks and  $\hat{\mathbf{m}}(H)$  the estimated SDFs estimated from higher-leveraged stocks. When the H and L markets are integrated, an element of these vectors at time t can be expressed as

$$\hat{m}_{j,t} = E_{t-1}(m_t) + v_{m,t} + v_{j,t}, j=L,H \quad (22)$$

where  $v_{m,t}$  is the unexpected component of the true SDF at time t and  $v_{j,t}$  is the sampling error in the estimated SDF for group j (j=L,H). No element on the right side of Eq. (22) is correlated with any other, so the time series variance of the estimated SDF is

$$\text{Var}(\hat{m}_j) = \text{Var}[E(m)] + \text{Var}(v_m) + \text{Var}(v_j), j=L,H. \quad (23)$$

Assuming that the estimation errors for L and H are independent of each other,

---

<sup>22</sup> The average leverage (book debt/total assets) ratio is 10.21% for the 956 low-leverage equities and 32.51% for the second (random) group.

$$\text{Cov}(\hat{m}_L, \hat{m}_H) = \text{Var}[E(m)] + \text{Var}(v_m) . \quad (24)$$

The right side of (24) is the total volatility induced by the true SDF, including the intertemporal evolution of its expectation and its period-by-period unexpected component. Subtracting this result from (23) provides estimation error variances for  $j=L$  and  $j=H$ .

The SDF paradigm implies 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  $\text{Var}(v_m)$ , which should dominate (24).

Estimated over July 2002 through December 2013, the standard deviations of  $\hat{m}_L$  and  $\hat{m}_H$  are, respectively, 0.5797 and 0.6964 per month and the correlation of  $\hat{m}_L$  and  $\hat{m}_H$  is 0.3002. This implies a standard deviation of true SDF components, the square root of (24), equal to 0.3482. The standard deviations for the estimation errors for L and H are then, respectively, 0.4636 and 0.6031. Not surprisingly, higher leveraged equities are associated with more volatile estimation errors. Both  $\hat{m}_L$  and  $\hat{m}_H$  are slightly autocorrelated at the first lag, autocorrelations of .181 and .157, respectively, but neither is statistically significant. In other respects, they seem to possess no particularly bizarre properties; for example, their excess kurtoses are -0.266 and 0.436 and their skewnesses are 0.371 and 0.457, respectively.

#### IV.B. Estimating True Perturbation Volatility

For a given stock  $i$ , the sample perturbation at time  $t$ , using the sample SDF  $j$  for either higher- or lower-leveraged stocks, is given by  $\hat{m}_{j,t} R_{i,t} - 1$  ( $j=L,H$ ) and its time series sample variance is  $\text{Var}(\hat{m}_j R_i)$ . However, assuming that the estimation error in the sample SDF is unrelated to the returns of an individual stock,  $\text{Cov}(\hat{m}_L R_i, \hat{m}_H R_i) = \text{Var}(m R_i)$ , where again we assume that the true SDF is the same for both groups of stocks and for convenience we denote the true SDF as  $m \equiv E(m) + v_m$ . This follows because within the covariance, the estimation error in the SDF is unrelated to  $R$  and hence  $E(v_j R_i) = E(v_j)E(R_i)$ ,  $j=L,H$ . Consequently, for any stock  $i$ , the true perturbation variance is equal to the covariance between two products, the same return  $R_i$  multiplied by the two estimated SDFs for L and H.

Using our same estimated SDFs for the higher- and lower-leveraged groups of stocks, 956 stocks in each group, we compute the above covariance of products for each stock. Table III tabulates the results. As shown there, the estimated true perturbation standard deviation is 0.2977 (0.3003) for the lower- (higher-) leveraged stocks on average. As one would expect, there is substantial cross-stock variation in the

perturbation volatility, ranging from 0.2194 (0.1990) to 0.4862 (0.5774) for the lower- (higher-) leveraged groups.

The previous section (IV.A) reports that the true SDF has a volatility of 0.3482, so these perturbation volatilities are slightly smaller on average and considerably smaller for some stocks, though they are also substantially larger for other stocks. Again not surprisingly, more leveraged equities deviate further from the SDF equilibrium equation on a monthly basis.

#### IV.C. Checking the Hansen/Jagannathan Bounds

The Hansen/Jagannathan (1991) variance-related bound requires that  $\sigma(m)/E(m)$  be larger than the largest possible Sharpe ratio.<sup>23</sup> 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  $0.348\sqrt{12}$ , 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 provide a simple test that the true SDFs are strictly positive (and, consequently, that there are no arbitrage opportunities.) The monthly observed standard deviation, 0.3481, implies that a negative SDF realization would be about 2.37 standard deviations below the mean. This implies that the probability of a negative SDF is about 0.0088, less than one percent.

To obtain a visual image of the evolution of the 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 (24). We simply need to find an attenuation coefficient,  $\gamma$  such that  $\text{Var}(\gamma\hat{m}) \cong \text{Var}(v_M)$ , which assumes that the riskless rate's variance is sufficiently small that it can be ignored; hence,  $\gamma = [\text{Var}(v_m) / \text{Var}(\hat{m})]^{1/2}$ . The adjustment entails the transformation

$$\hat{\hat{m}} = \bar{m} + \gamma(\hat{m} - \bar{m}), \quad (25)$$

where the double "chapeau" denotes 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. This attenuation

---

<sup>23</sup> Hansen and Jagannathan also derived bounds involving moment other than the first and second. See Snow (1991) for empirical estimation with a variety of bounds.

provides an adjusted standard deviations of exactly 0.3481 for both the Low and Higher leverage equity groups.

Figure IV 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 toward the end 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.<sup>24</sup> Higher SDF values from 2006 through 2010 are, perhaps, not all that surprising as they precede and accompany the recent economic contraction.

#### IV.D. Estimated SDFs and Returns on a Market Index

In one further validation experiment, we estimate the relation over time between SDF estimates from low- and higher-leveraged stocks and observed returns on the S&P 500 index. This is motivated because the SDF paradigm stipulates that in each period there should be a relation between the aggregate market portfolio's return,  $R_{W,t}$ , and the SDF  $m_t$ , of the following form:

$$m_t = a_{t-1} - b_{t-1}R_{W,t} ,$$

where the coefficients are time varying and strictly positive; Cf. Cochrane (2005, pp. 139-140.)

Unfortunately, we have only estimates of the two variables in the relation above, our estimate  $\hat{m}$  for  $m$  and the S&P 500 return for  $R_W$ . Consequently, we really cannot estimate their time series variation and simply illustrate their average values below.

Operationally, we run two proximate regressions,

$$\hat{m}_{j,t} = a_j - b_j R_{S\&P,t}$$

with  $j=L$  (H) for Lower- (Higher-) leveraged equities, using 138 monthly observations, July 2002 through December 2013. Perhaps surprising, given the possible problems with this specification, we find  $-b_L=-3.48$  (t-statistic=-3.17) and  $-b_H=-2.68$  (t-statistic=-1.98.) Both slope coefficients have the right sign and are significant, though  $b_H$ 's significance could be regarded as marginal. Clearly, there is more estimation error in  $\hat{m}$  for the higher-leveraged equities.<sup>25</sup> As one would expect, the intercept terms are both very close to 1.0 and are highly significant, t-statistics of 21.0 and 17.1, respectively. However, the explanatory power is rather low, adjusted R-squares of 6.20% and 2.10%, respectively.

<sup>24</sup> Neither series has a unit root according to the usual tests.

<sup>25</sup> See equation (24), which implies that the slope in the regression above is less significant.

## V. Conclusions

The stochastic discount factor (SDF) paradigm predicts that the same SDF should price all assets in a given period when markets are complete. This implies that an SDF estimated from one subset of assets can be used to price assets from other subsets within the same (complete) market.

We derive an SDF estimator that can be calculated from observed returns only and is not a function of macroeconomic state variables, preferences, and the form of the multivariate distribution of returns, including its factor structure. We emphasize that an SDF estimator need not depend on the above-mentioned traditional elements because the SDF is a mathematical function within an integral equation. This property of the SDF implies that it does not depend on the distribution of returns, including their factor structure, if any.

Our SDF estimator is consistent but is 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. According our simulation, the estimator is accurate when  $N-T$  is relatively large with  $N > 2T$  and  $N$  near 1,000. There is very little bias for samples of this size.

Using a sample of U.S. equities, which exist in large numbers, we find that two large groups of equities, sorted into two groups based on leverage, are priced with SDFs that are not statistically distinguishable over a recent period of 138 months, 2002-2013. We also find that the estimated SDFs comfortably satisfy the Hansen/Jagannathan variance bound during this period.

Some caveats about our results and suggestions for future research include, though are not necessarily limited to the following:

- We use individual stocks. Future researchers might consider forming portfolios by sorting stocks on their characteristics, and computing our SDF estimator with portfolios instead, which have less noise. However, this has the disadvantage of reducing  $N$  relative to  $T$ , which may or may not become an overriding negative consequence.
- For some individual stocks that are thinly traded, the SDF may not be estimated to be unique even when such stocks are traded in an integrated market. Researchers might examine this issue with the methods in Pelger and Lattau (2017).
- We have not, but future research should, address the issue of survivorship bias especially for stocks listed on Nasdaq using the approach in Shumway (1997) and Shumway and Warther (1999).

- Our estimator requires  $T$  simultaneous observations for  $N$  assets. Many assets do not have continuous records, which suggest using methods outlined in Kim and Koraczyk (2019). They claim that our estimator “...requires very large samples (*larger than available in most empirical studies*) to converge to the true SDF; it requires a balanced panel with its inherent survivorship biases; and it is biased for finite time-series samples.” Yes, it is biased in small samples but we show simulations the bias is not large for  $N$  less than 1,000 provided that  $N > 2T$ . An investigator with a sufficient number of assets can generally simply reduce the time series sample size to meet this condition. For example, we test 2 groups of 956 stocks each that have simultaneous monthly observations for 10 years. A balanced panel seems rather unobjectionable for short periods particularly because the assets to include can be chosen freely. Researchers and practitioners can choose a set of assets that has simultaneous sample observations, Our estimator does not rely on a particular sample of assets.
- The estimated SDF in Equation (6) can be used to explain the cross-sectional variation in equity returns. Kim and Korajczyk (2019) show this application with their semi-agnostic estimator.
- It is difficult to provide a direct comparison between our measure and the other SDF measures. To the best of our knowledge, there have been no other approaches that do not use macro variables or preferences. Korajczyk and Kim (2019) adopt our idea and then impose a factor structure on returns, which we do not require. Of the several hundred papers that use macros and/or preferences, none have been directly comparable to us.
- Useful future research would estimate the SDF from other asset classes and with international data.

Table I  
Simulated Performance of the SDF Estimator Under Various Conditions

True and estimated SDFs are simulated with various sample sizes, time series volatilities and riskless rates. The SDF estimator's performance is measured by the correlation between the true and estimated SDFs, by Theil's (1966)  $U_2$  statistic, and by the relative bias from a regression of the estimated SDF on the true SDF (see Eq (21)). Each panel reports a cross-parameter multiple linear regression between the dependent variable given in the panel heading and the five explanatory variables (parameters) in the left-most column. There are 2,880 parameter combinations as listed in the text, each one with an independently-simulated set of true SDFs, returns, and estimated SDFs.

Variable	Coefficient	T-Statistic
A: Correlation between true and estimated SDFs		
T, Time Periods	-1.104E-03	-11.034
N, Assets	1.505E-04	12.036
True SDF Volatility	1.295	18.581
Perturbation Volatility	-1.744	-49.302
Riskless Rate	5.799E-01	0.244
Adjusted R <sup>2</sup>	0.488	
B: Theil's $U_2$ from comparing true and estimated SDFs		
T, Time Periods	2.068E-03	54.927
N, Assets	-2.964E-04	-62.989
True SDF Volatility	1.344E-01	5.126
Perturbation Volatility	6.237E-01	46.860
Riskless Rate	2.423E-01	0.271
Adjusted R <sup>2</sup>	0.782	
C: Relative Bias of SDF Estimator (in %)		
T, Time Periods	1.110E-05	12.914
N, Assets	-6.235E-07	-5.8052
True SDF Volatility	-1.512E-04	-0.2526
Perturbation Volatility	-4.318E-03	-14.216
Riskless Rate	7.255E-03	0.3560
Adjusted R <sup>2</sup>	0.141	
D: Standard Deviation of Estimated SDFs		
T, Time Periods	2.917E-03	53.945
N, Assets	-4.270E-04	-63.166
True SDF Volatility	5.180E-02	1.376
Perturbation Volatility	4.899E-01	25.622
Riskless Rate	2.360E-01	0.184
Adjusted R <sup>2</sup>	0.731	
E: Riskless Rate Inferred from Estimated SDFs		
T, Time Periods	-3.106E-06	-0.229
N, Assets	2.401E-06	1.415
True SDF Volatility	1.950E-02	2.063
Perturbation Volatility	1.164E-02	2.427
Riskless Rate	7.797E-01	2.422
Adjusted R <sup>2</sup>	0.008	

Table II

## Parameters for Good and Poor Performance of the Estimated SDF

From the 2880 parameters reported in the results of Table I, the 74 parameter combinations that produce a correlation of at least 80% between the estimated and true SDF are chosen as a “good performance” sub-sample and are contrasted with 74 combination that have the smallest explanatory power, a “poor performance” sub-sample. The sub-sample averages for T, N, the true SDF’s volatility and the Perturbation volatility are reported here. The “Best” and “Worst” lines report the parameter values that produce, within each sub-sample, the largest and smallest explanatory power.

	T	N	SDF Volatility	Perturbation Volatility	Correlation
Good Performance					
Sub-Sample Average	40.5	775.1	0.16	0.03	0.8623
Best S-S <sup>26</sup> Parameters	30	960	0.1	0.02	0.9510
Worst S-S Parameters	60	960	0.1	0.02	0.8004
Poor Performance					
Sub-Sample Average	50.27	564.32	0.19	0.27	-0.2493
Best S-S Parameters	30	480	0.2	0.16	-0.1805
Worst S-S Parameters	30	240	0.2	0.32	-0.5474

<sup>26</sup> “S-S” refers to Sub-sample; the parameters listed in adjacent lines are the best or worst within the sub-sample.

Table III  
Empirical Estimates of True Perturbation Volatility

Perturbation volatility is the error in the exact SDF pricing relation. In month  $t$  it is equal to  $m_t R_{i,t} - 1$ , where  $m$  is the true SDF realization for month  $t$  and  $R$  is the gross return (1+the net return) for asset  $i$ . Given an assumption that two groups of assets are traded in an integrated complete market, the true SDF is the same for both groups in every month. This permits the computation of the true SDF variance from the covariance between the two sample SDFs, one estimated from each group. Similarly, the true perturbation variance can be obtained for each asset from the covariance between two products, the asset's return multiplied by each sample SDF.<sup>27</sup> These computations are carried out for monthly returns on U.S. stocks over, July 2002 through December 2013, 138 months in all. Two groups are included: (1) 956 equities with lower leverage and (2) an equal-sized sample of higher-leveraged equities. See section III for data description. For this sample period, the time series standard deviation of the true SDF is 0.3482 per month, (Section IV.A.) For the 956 individual stocks in each leverage group, cross-stock statistics for the resulting perturbation volatility (standard deviation per month) are given in the table below.

Leverage Group	Mean	Standard Deviation	Maximum	Minimum
Lower	0.2977	0.0273	0.4862	0.2195
Higher	0.3003	0.0350	0.5774	0.1990

---

<sup>27</sup> This is explained in more detail in sections IV.A and IV.B of the text.

## References

- Adrian, Tobias, Richard K. Crump, and Emanuel Moench. "Pricing the Term Structure with Linear Regressions," *Journal of Financial Economics* 110(1) (2013), 110-138.
- Aït-Sahalia, Yacine and Andrew W. Lo. "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance* 53 (1998), 499-547.
- Aït-Sahalia, Yacine and Andrew W. Lo. "Nonparametric Risk Management and Implied Risk Aversion," *Journal of Econometrics* 94 (2000), 9-51.
- Alvarez, Fernando, and Urban J. Jermann. 2005. "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth." *Econometrica*, 73(6): 1977-2016.
- Araujo, Fabio, Joao Victor Issler, and Marcelo Fernandes. "Estimating the Stochastic Discount Factor without a Utility Function," Working Paper, Graduate School of Economics EPGE Getulio Vargas Foundation, March 2005.
- Araujo, Fabio and Joao Victor Issler. "A Stochastic Discount Factor Approach to Asset Pricing Using Panel Data Asymptotics," Graduate School of Economics EPGE Getulio Vargas Foundation, 2011.
- Back, Kerry, *Asset Pricing and Portfolio Choice Theory*, 2<sup>nd</sup> edition, Oxford University Press.
- Bossaerts, Peter L., *The Paradox of Asset Pricing*, Princeton, Princeton University Press, 2002.
- Britten-Jones, Mark, "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights," *Journal of Finance* 54 (1999), 655-671.
- Brown, Morton B. and Alan B. Forsythe. "Robust Tests for Equality of Variances," *Journal of the American Statistical Association* 69 (1974), 364-367.
- Brown, Stephen J. and Mark I. Weinstein. "A New Approach to Testing Asset Pricing Models: The Bilinear Paradigm," *Journal of Finance* 38(3) (1983), 711-43.
- Campbell, John Y. "Intertemporal Asset Pricing without Consumption Data," *American Economic Review* 83(1993), 487-512.
- Campbell, John Y. "Empirical Asset Pricing: Eugene Fama, Lars Peter Hansen, and Robert Shiller," *the Scandinavian Journal of Economics* 116 (2014), 593-634.
- Chapman, David A. "Approximating the Asset Pricing Kernel," *Journal of Finance* 52 (1997), 1383-1410.
- Chen, Xiaohong and Sydney Ludvigson, "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Journal of Applied Econometrics* 24 (2009), 1057-1093.
- Chen, Zhiwu, and Peter J. Knez, "A Pricing Operator-Based Testing Foundation for a Class of Factor Pricing Models," *Mathematical Finance* 4 (1994), 121-141.
- Chen, Zhiwu, and Peter J. Knez. "Measurement of Market Integration and Arbitrage," *Review of Financial Studies* 8 (1995), 287-325.

Chen, Zhiwu, and Peter J. Knez. "Portfolio Performance Measurement: Theory and Applications," *Review of Financial Studies* 9 (1996), 511-555.

Cochrane, John H. "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy* 104 (1996), 241-255.

Cochrane, John H. "Asset Pricing," Revised Edition, Princeton: Princeton University Press (2005).

Cochrane, John H., and Christopher L. Culp. "Equilibrium Asset Pricing and Discount Factors: Overview and Implications for Derivatives Valuation and Risk Management," ch. 5, in *Modern Risk Management: A History*. Peter Field, ed. London: Risk Books (2003).

Cochrane, John H., and Lars Peter Hansen. "Asset Pricing Explorations for Macroeconomics," Working Paper No. 4408, National Bureau of Economic Research (June 1992.)

Da, Zhi and Hayong Yun. "Industrial Electricity Usage and Stock Returns," *Journal of Financial and Quantitative Analysis* 52 (2017), 37-69.

DeSantis, Gerard. "Volatility Bounds for Stochastic Discount Factors: Tests and Implications from International Financial Markets," Ph.D. dissertation, Department of Economics, University of Chicago (1993).

Dew-Becker, Ian, and Stefano Giglio, 2016. "Asset Pricing in the Frequency Domain: Theory and Empirics." *Review of Financial Studies*, 29(8): 2029-2068.

Engle, Robert F., and Sharon Kozicki. "Testing for Common Features," *Journal of Business and Economic Statistics* 11 (1993), 369-80.

Ferson, Wayne. "Theory and Empirical Testing of Asset Pricing Models," in Jarrow, R. A., V. Maksimovic, and W. T. Ziemba (eds.), *Handbooks in OR and MS*, Vol. 9, Finance, Amsterdam: Elsevier Science, 1995.

Hansen, Lars Peter, and Jose A. Scheink,an. 2009. "Long-term Risk: An Operator Approach." *Econometrica*, 77(1): 177-234.

Hansen, Lars Peter, and Ravi Jagannathan. "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy* 99 (1991), 225-262.

Hausman, J.A. "Specification Tests in Econometrics," *Econometrica* 46 (6) (1978), 1251-1271.

He, Jia, Lilian Ng and Chu Zhang. "Asset Pricing Specification Errors and Performance Evaluation," *European Finance Review* 3 (1999), 205-232.

Hotelling, Harold. "The Generalization of Student's Ratio," *Annals of Mathematical Statistics* 2 (3) (1931), 360-378.

Kim, Soohun, and Robert Korajczyk, "Large Sample Estimator of the Stochastic Discount Factor," Georgia Institute of Technology, Scheller College of Business working paper, 2019.

Kim, Soohun, Robert Korajczyk, Kuntara Pukthuanthong, and Richard Roll, “Agnostic and Semi-Agnostic Estimators of the Stochastic Discount Factor,” Georgia Institute of Technology, Scheller College of Business working paper, 2019.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. “Interpreting Factor Models,” *Journal of Finance* 73 (2017), 1183-1223.

Lettau, Martin, and Sydney Ludvigson. “Consumption, Aggregate Wealth and Expected Stock Returns,” *Journal of Finance* 56 (2001), 815-849.

Long, John. “The Numeraire Portfolio,” *Journal of Financial Economics* 26(1) (1990), 29-69.

Martin, Ian. 2017. “What Is the Expected Return on the Market?” *Quarterly Journal of Economics*, 132(1): 367-433.

McCulloch, J. Huston, “The Risk-Neutral Measure and Option Pricing under Log-Stable Uncertainty,” Ohio State University Economics Department working paper, (2003).

Mehra, Raj, and Edward Prescott. “The Equity Premium: A Puzzle,” *Journal of Monetary Economics* 15 (1985), 145-161.

Pelger, Markus and Martin Lettau. “Estimating Latent Asset-Pricing Factors,” forthcoming *Journal of Econometrics*, 2019.

Polyanin, Andrei D., and Alexander V. Manzhirov, *Handbook of Integral Equations*, Boca Raton, CRC Press, 1998.

Roll, Richard. “Evidence on the Growth Optimum Model,” *Journal of Finance* 28(3) (1973), 551-566.

Rosenberg, Joshua, and Robert F. Engle. “Empirical Pricing Kernels,” *Journal of Financial Economics* 64 (2002), 341-372.

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

Ross, Stephen A. “The Recovery Theorem.” *Journal of Finance* 70(2): 615-648.

Shumway, Tyler, The Delisting Bias in CRSP Data, *Journal of Finance* 52 (1997), 327-340.

Shumway, Tyler and Vincent Warther, 1999, The Delisting Bias in CRSP’s Nasdaq Data and Its Implications for the Size Effect, *Journal of Finance* 54 (1999), 2361-2379.

Singleton, Kenneth J. “Specification and Estimation of Intertemporal Asset Pricing Models,” in Briedman, B., and F. Hahn (eds.), *Handbook of Monetary Economics*, Vol, 1, Amsterdam, Elsevier Science, 1990.

Singleton, Kenneth J. “*Empirical Dynamic Asset Pricing*,” Princeton: Princeton University Press, 2006.

Snow, Karl N. “Diagnosing Asset Pricing Models Using the Distribution of Asset Returns.” *Journal of Finance*, Papers and Proceedings 46 (1991), 955-983.

Theil, Henri. “*Applied Economic Forecasting*,” Chicago: Rand McNally, 1966.

Welch, B. L. “On the Comparison of Several Mean Values: An Alternative Approach,” *Biometrika* 38

(1951), 330-336.

Welch, Ivo, "View of Financial Economists on the Equity Premium and Other Issues," *Journal of Business* 73 (2000), 501-537.

Figure I  
The Estimated and True SDFs with Small Return Perturbations

To demonstrate the SDF estimator, the standard deviation of the perturbation in Eq. (17) 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.

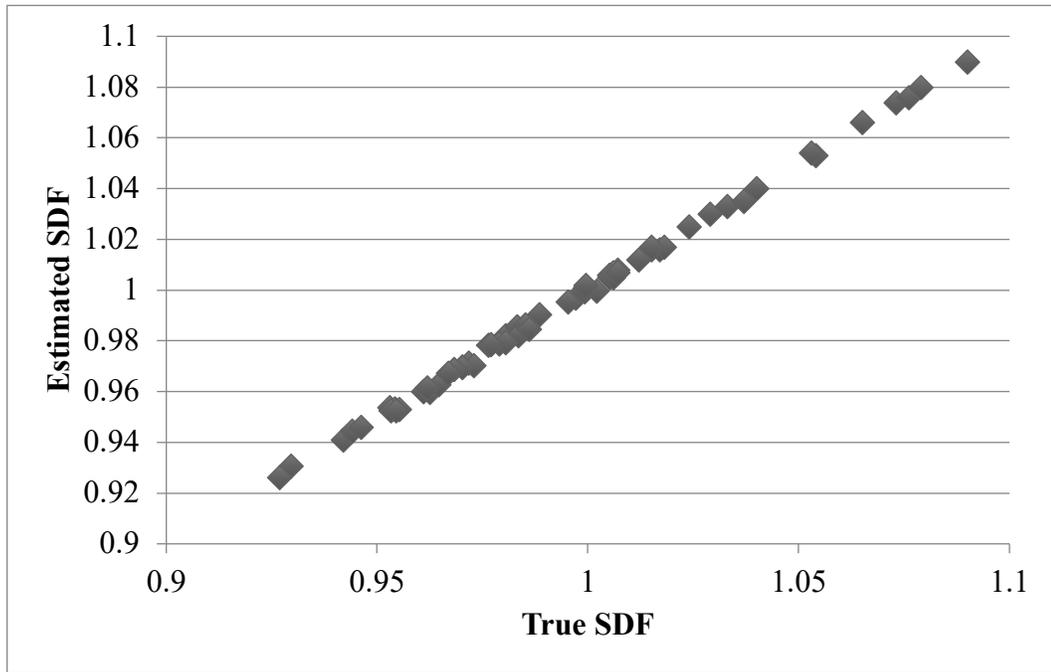
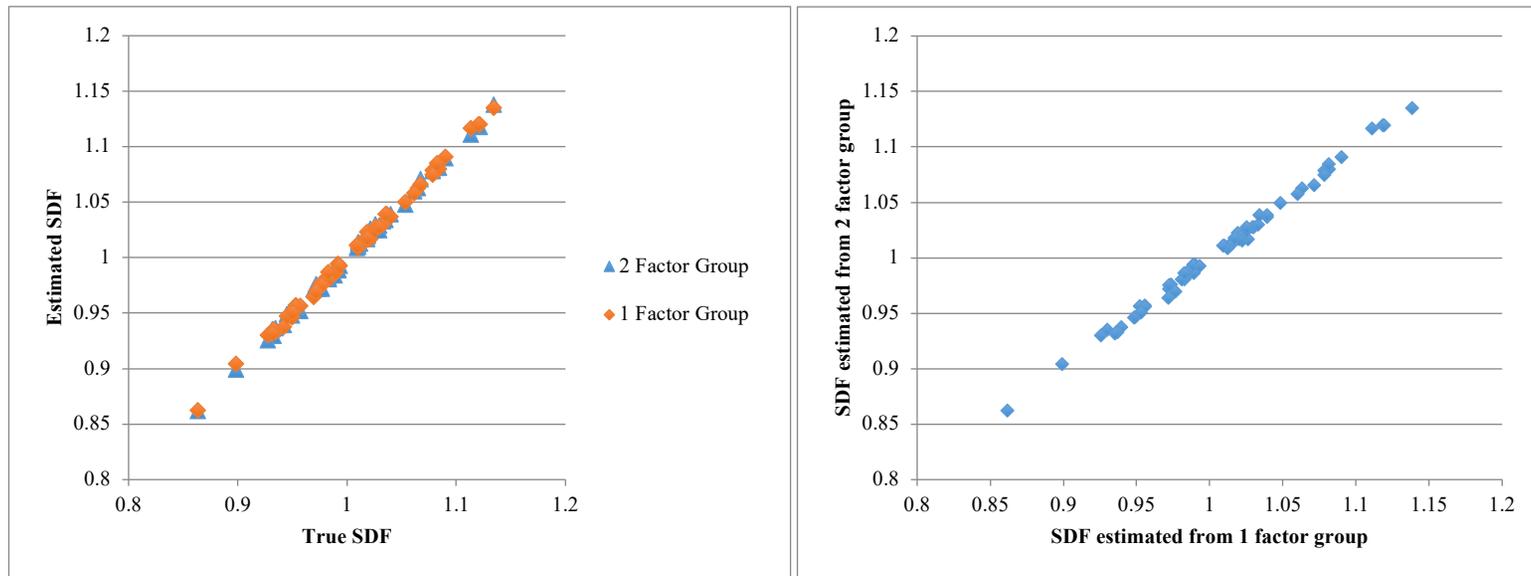


Figure II

Estimated and True SDFs for Asset Groups  
with Diverse Factor Structures and Levels of Return Perturbation Volatility

There is a unique SDF that prices all assets. It has a mean dictated by a riskless rate of .4% per period and a standard deviation is 4% per period. One group of assets has returns driven by a two-factor structure while the other group of assets has a single-factor structure. 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. In the Panel A, the return perturbations are very small, a standard deviation of 0.01% per period. Panel B has return perturbations with ten times as much volatility, a standard deviation of 0.1% per period. All other parameter values for the simulations are specified the text, Section II.E. The first plot below shows each group's estimated SDF plotted against the true SDF. The second plot shows the estimated SDFs for the two asset groups plotted against each other.

Panel A, Small Perturbations



Panel B, Larger Perturbations

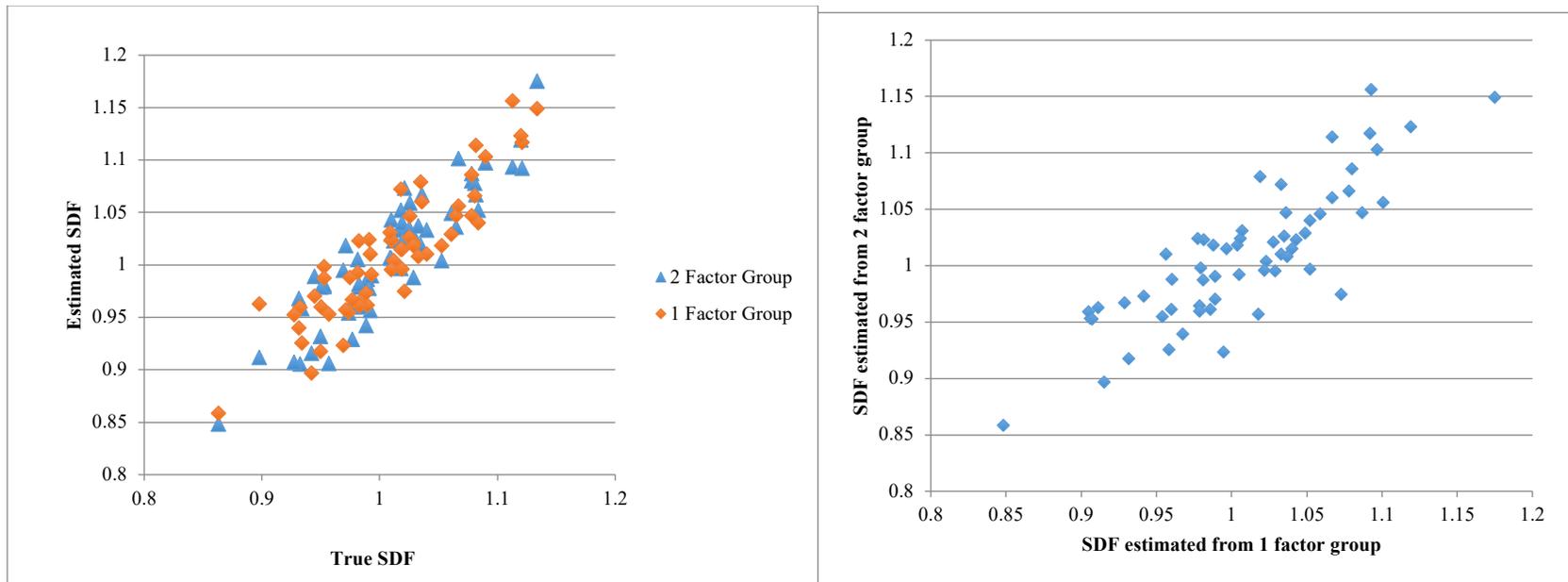


Figure III

Estimated and True SDFs for Two Asset Groups,  
Both with Two-Factor Structures but Whose Factors are Unrelated

There is a unique SDF that prices all assets. It has a mean dictated by a riskless rate of .4% per period and a standard deviation is 4% per period. Both groups of assets have returns driven by a two-factor structure but the factors are unrelated across groups. The groups are denoted 2 Factor A and 2 Factor B, respectively. See Section II.E for details. 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. The return perturbations are relatively large, a standard deviation of 0.1% per period, the same as in Panel B of Figure II above. All other parameter values for the simulations are specified Section II.E. The first plot below shows each group's estimated SDF plotted against the true SDF. The second plot shows the estimated SDFs for the two asset groups plotted against each other.

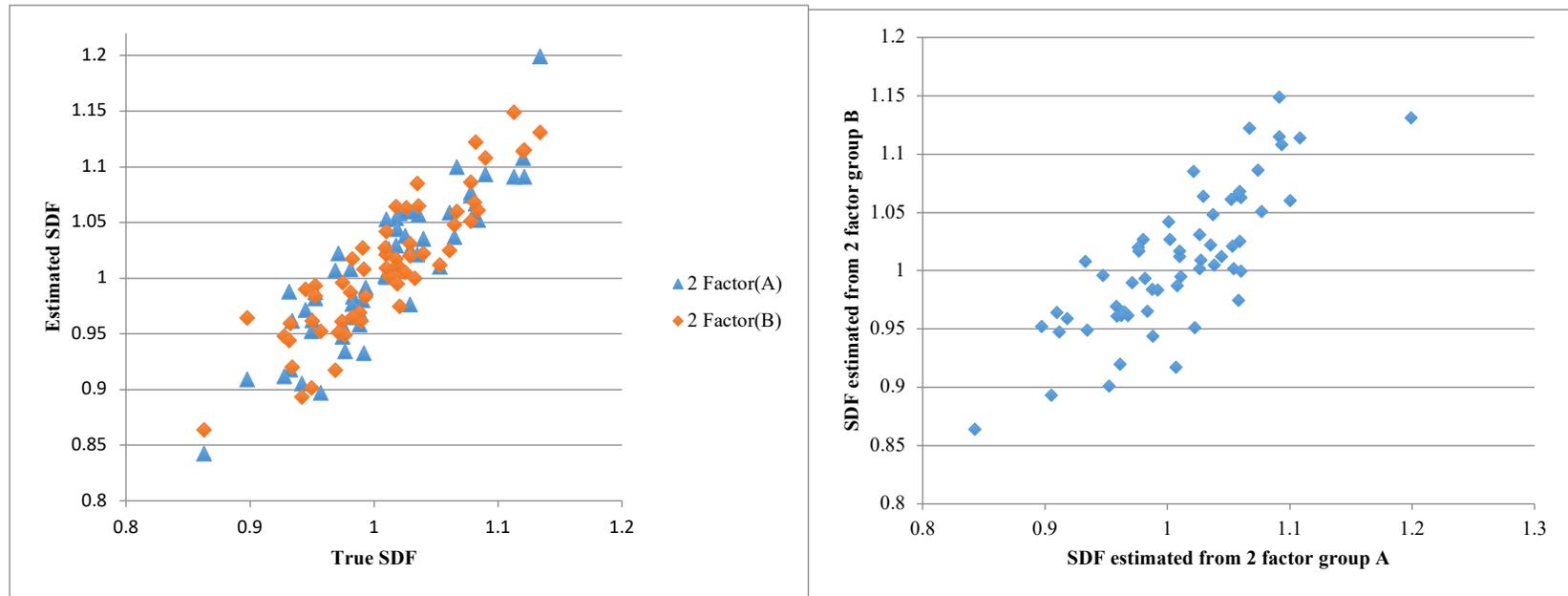
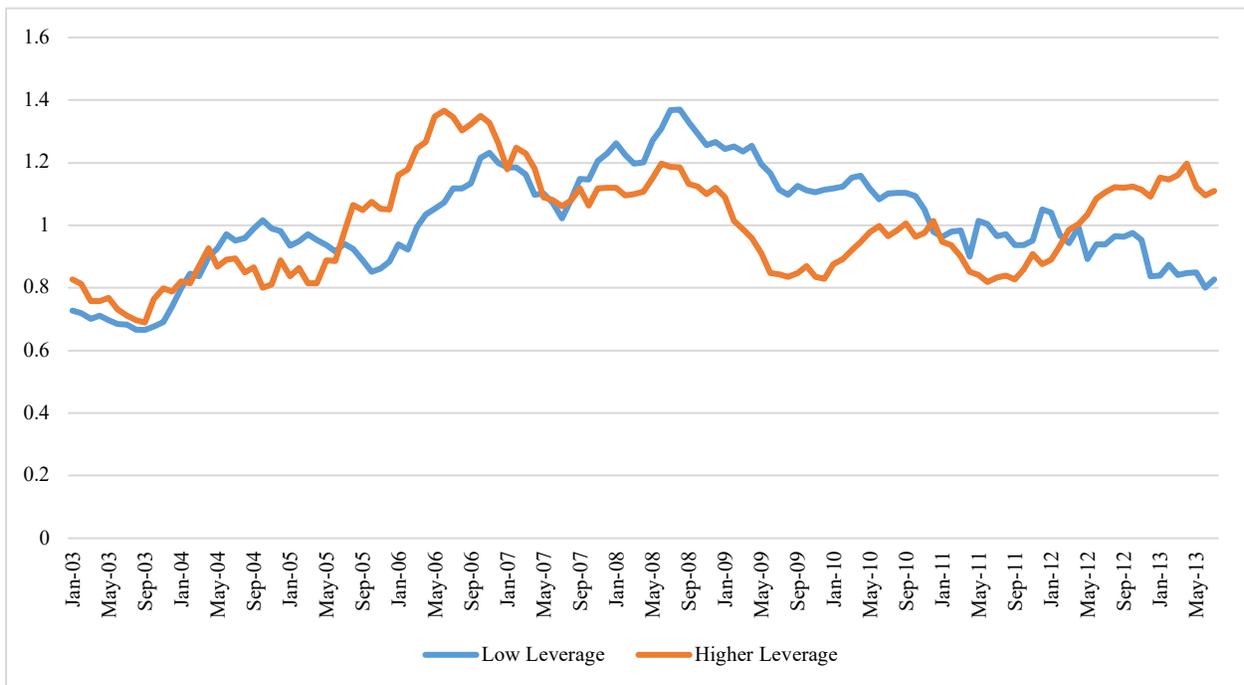


Figure IV

Time Series Plots of Estimated SDFs from Low- and Higher-Leveraged Equities

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 is selected from firms with the lowest average leverage ratios 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.



## Appendix A Robustness Checks

Section A.1 provides alternative simulation results for the situation when returns are negatively correlated with the stochastic discount factor. Section A.2 reports on the consequences of thick tails, a phenomenon that is seemingly ubiquitous for financial asset returns. Section A.3 proposes a modified estimator that may be useful when estimation error is highly heterogeneous.

### A.1. Negative Correlation Between the SDF and Returns

Theoretically, the basic SDF equation usually implies negative correlation between the SDF and returns. This follows by noting that  $E(mR) = 1$  implies  $\text{Cov}(m, R) = 1 - E(R)E(m)$  and  $E(R)E(m) > 1$  because the risky return is usually larger than the riskless rate.<sup>28</sup> However, the simulations in Table I of the text do not take this into account because the disturbance in the SDF generating equation (15) is unrelated to the disturbance in the return generating equation (16.) To ascertain the impact of this specification, if any, we change the specification in (16) to the following

$$\hat{R}_{i,t} = \mu \exp \{ \zeta_{i,t} + \rho [m_t - E(m)] - (\sigma_\zeta^2 + \sigma_\xi^2) / 2 \}, (t = 1, \dots, T; i = 1, \dots, N). \quad (\text{A.1})$$

In this appendix sub-section, we illustrate the effect by choosing a rather large negative coefficient,  $\rho = -.9$ , which is approximately the correlation between returns and the SDF (but not exactly because of the exponentiation.)<sup>29</sup>

The results, reported in Table A-1, should be compared to those in Table I of the text. Before reporting the highlights, however, we now find that the accuracy of our estimator has improved in response to the specification of the negative relation between  $m$  and  $R$ . For instance, in the previous simulations (with zero correlation between  $m$  and  $R$ ), the correlation between the estimated SDF and the true SDF was 0.189 on average over the 2,880 parameter values in Tables I or A-1, with a standard deviation of 0.252 and a maximum (minimum) of 0.941 (-.547). With the negative correlation now underlying Table A-1, the corresponding mean is 0.380 with a standard deviation of 0.219 and a maximum (minimum) of 0.969 (-.177).

A similar improvement is evident in both Theil's statistic and the relative bias (unreported.)

In some cases, there is also an improvement in the explanatory power of the cross-sectional regressions on parameters. For example, in Panel A of Table A-1, where the dependent variable is the correlation between true and estimated SDFs,  $T$ ,  $N$ , and the true SDF volatility all have larger coefficients

---

<sup>28</sup> Recall the  $R$  is the gross return; i.e.,  $(1 + \text{the net return})$ , and  $E(m) = 1 / (1 + \text{the riskless return})$ .

<sup>29</sup> The term in parentheses is a correction for the exponentiated mean.

and t-statistics (in absolute value.) The perturbation volatility is less significant, in contrast, which suggests a more accurate estimator less affected by noise.

There is not much change in Theil's statistic (Panel B), but the relative bias (Panel C) reveals an interesting pattern. Note the smaller levels of significance for everything except the true SDF volatility. It's more significant here (and also in Panel D) probably because the SDF now enters the return generating equation (A-1) directly. But the general impression is that the relative bias has been reduced by allowing for negative correlation between R and m.

Overall, if R and m really are negatively related, our estimator performs better.

## A.2. Thick-tailed returns

In the simulations of Sections II and A.1, 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 previous simulations holding everything the same except for the return perturbations, which are now assumed to follow a truncated Cauchy distribution. We again generate "true" SDFs with a lognormal distribution as in (15) and the same panoply of parameters. Initial gross returns are also generated in the same way, as in (16).

But (17) is replaced by

$$R_{i,t} = \frac{\hat{R}_{i,t}}{\sum_{t=1}^T m_t \hat{R}_{i,t} / T} + \vartheta_{i,t} \quad (\text{A-2})$$

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 exhibits extreme outcomes compared to a Gaussian. In the simulations here, we truncate the Cauchy tails, retaining only the middle 95% of simulated values.<sup>30</sup> With a 95% truncation and the scale parameters listed above, gross returns are guaranteed to remain strictly positive.

---

<sup>30</sup> The simulations first select a cumulative distribution function p-value, a number between zero and 1.0, and then calculate 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.

The return perturbation in (A-2) is additive, in contrast to the previously multiplicative lognormal return perturbation 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 (15) through (17)), but there is no corresponding correction using the Cauchy scale for the same purpose. An additive return perturbation finesses this difficulty because it is symmetric and not exposed to the amplification of exponentiation.<sup>31</sup>

Table A.2, which corresponds to Table I of the text, 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.

In Panel C, which reports the SDF percentage relative bias, the cross-sectional explanatory variable are generally less significant though just one, the number of assets  $N$ , has fallen from positive an significant in Table I to insignificant in Table A-2.

In Panel D, which shows the influence of various parameters on the volatility of the estimated SDFs, the true SDF's volatility has now become significant whereas it was not in Table I.

In Panel E, for 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 E of Table I 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 agreement with the mean component of the MSE, the relative bias (in %) is only -0.0105% with a maximum across 2880 parameter combinations of 0.249% and a minimum of -0.352%. As with lognormal returns, bias in the SDF estimator is not a material concern even with thick-tailed returns.

---

<sup>31</sup> 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.

Overall, in summary, thick-tailed returns do not appear to compromise the qualities of our estimator. Its occasional apparent 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.

### A.3. An Improved SDF Estimator for Cross-Sectionally Heteroscedastic Errors<sup>32</sup>

As explained in Section II, our SDF estimator minimizes the sum of squared average surprises in the SDF\*Gross Return product. The surprise for asset  $j$  in period  $t$  is  $\varepsilon_{j,t} = m_t R_{j,t} - 1$  (See Eq. (3)), its

average over  $T$  periods is  $\bar{\varepsilon}_j = (1/T) \sum_{t=1}^T \varepsilon_{j,t}$  and the variance of its average, assuming no serial correlation

but allowing for non-stationarity, is  $\text{Var}(\bar{\varepsilon}_j) = (1/T^2) \sum_{t=1}^T \text{Var}_t(\varepsilon_{j,t})$ . This suggests that a weighted estimator, with weights proportional to the precisions of each asset's average surprise, might very well have smaller sampling error, particularly when it is suspected to be a wide discrepancy across assets in  $\text{Var}(\bar{\varepsilon}_j)$ .

The only problem is that the surprise for each asset cannot be observed without knowing the value of the SDF in each period; hence, an iterative approach is required. We implement the iteration as follows: (1) calculate an SDF with the unweighted estimator in (9); (2) calculate an estimated  $\text{Var}(\bar{\varepsilon}_j)$  for each asset using this initial SDF; (3) weight the returns for asset  $j$  and the  $j$ th position in the vector of 1.0's by the precision,  $1/\sqrt{\text{Var}(\bar{\varepsilon}_j)}$ ; (4) obtain the new SDF and new measures of surprise with the weighted observations; (5) repeat until there are miniscule changes in the resulting SDF.

We apply this approach with the simulated two-factor structure described in Section II, while adding cross-sectional heteroscedasticity to the perturbation volatility in (17.) Specifically, the perturbation volatility for  $j$  is  $\sigma_{j,t}$ , where  $\bar{\sigma}_g = .01\%$  is the average perturbation volatility across assets. We vary the volatility of over a range of multiples of  $\bar{\sigma}_g$  and find very little improvement for multiples from 10 to 1000. For example, with a multiple of 500, the correlations between the true SDF and an estimated SDFs improve from .5984 for original unweighted estimator to .5998 with the weighted one.

Of course, things could be different with larger average perturbation volatility and with assets whose SDF estimation errors are grossly heteroscedastic.

---

<sup>32</sup> We are indebted to Robert Engle for suggesting this refinement.

Table A-1.  
 Simulated Performance of the SDF Estimator Under Various Conditions  
 With Negative Correlation between the SDF and Returns

True and estimated SDFs are simulated with various sample sizes, time series volatilities and riskless rates. The SDF estimator's performance is measured by the correlation between the true and estimated SDFs, by Theil's (1966)  $U_2$  statistic, and by the relative bias from a regression of the estimated SDF on the true SDF. Each panel reports a cross-parameter multiple linear regression between the dependent variable given in the panel heading and the five explanatory variables (parameters) in the left-most column. There are 2,880 parameter combinations as listed in the text, each one with an independently-simulated set of true SDFs, returns, and estimated SDFs.

Variable	Coefficient	T-Statistic
A: Correlation between true and estimated SDFs		
T, Time Periods	-2.462E-03	-29.084
N, Assets	3.090E-04	29.745
True SDF Volatility	2.315	35.764
Perturbation Volatility	-1.139	-28.372
Riskless Rate	-9.955E-01	-0.496
Adjusted R <sup>2</sup>	0.516	
B: Theil's $U_2$ from comparing true and estimated SDFs		
T, Time Periods	2.301E-03	55.615
N, Assets	-2.683E-04	-52.844
True SDF Volatility	9.717E-02	3.071
Perturbation Volatility	8.182E-01	41.688
Riskless Rate	-6.981E-02	-0.071
Adjusted R <sup>2</sup>	0.757	
C: Relative Bias of SDF Estimator (in %)		
T, Time Periods	5.062E-06	1.370
N, Assets	-2.356E-07	-0.520
True SDF Volatility	-1.039E-02	-3.675
Perturbation Volatility	-6.287E-03	-3.587
Riskless Rate	4.556E-03	0.052
Adjusted R <sup>2</sup>	0.023	
D: Standard Deviation of Estimated SDFs		
T, Time Periods	2.777E-03	49.174
N, Assets	-3.162E-04	-45.619
True SDF Volatility	5.227E-01	12.101
Perturbation Volatility	9.136E-01	34.103
Riskless Rate	-9.500E-01	-0.709
Adjusted R <sup>2</sup>	0.721	
E: Riskless Rate Inferred from Estimated SDFs		
T, Time Periods	-5.271E-06	-0.500
N, Assets	-1.547E-06	-1.195
True SDF Volatility	-1.103E-03	-0.137
Perturbation Volatility	1.081E-02	2.162
Riskless Rate	8.879E-01	3.551
Adjusted R <sup>2</sup>	0.006	

Table A.2  
Simulated Performance Information for the SDF Estimator Applied to Thick-Tailed Returns

We simulate true SDFs with  $\text{mean}=1/(1+\text{riskless interest rate})$  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 perturbation scale is 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, by Theil's (1966)  $U_2$  statistic, which is closely related to the mean square prediction perturbation, and by the relative bias from a regression of the estimated SDF on the true SDF. Each panel reports a cross-parameter multiple linear regression between the dependent variable given in the panel heading and the five explanatory variables (parameters) in the left-most column. There are 2,880 parameter combinations as listed in the text, each one with an independently-simulated set of true SDFs, returns, and estimated SDFs.

Variable	Coefficient	T-Statistic
A: Correlation between true and estimated SDFs		
T, Time Periods	-2.129E-03	-30.09
N, Assets	2.767E-04	31.29
True SDF Volatility	1.474	34.71
Perturbation Scale	-0.1788	-97.29
Riskless Rate	-0.5534	-0.330
Adjusted R <sup>2</sup>	0.813	
B: $U_2$ from comparing true and estimated SDFs		
T, Time Periods	1.772E-03	59.13
N, Assets	-2.483E-04	-66.29
True SDF Volatility	0.1337	7.434
Perturbation Scale	6.895	88.55
Riskless Rate	-0.4715	-0.663
Adjusted R <sup>2</sup>	0.846	
C: Relative Bias of SDF Estimator (in %)		
T, Time Periods	1.738E-06	6.572
N, Assets	-1.259E-08	-0.381
True SDF Volatility	-1.742E-04	-1.098
Perturbation Scale	-7.150E-03	-10.41
Riskless Rate	7.063E-04	0.113
Adjusted R <sup>2</sup>	0.0489	
D: Standard Deviation of Estimated SDFs		
T, Time Periods	2.049E-03	47.07
N, Assets	-2.946E-04	-54.15
True SDF Volatility	0.3602	13.79
Perturbation Scale	4.646	41.09
Riskless Rate	-1.518	-1.471
Adjusted R <sup>2</sup>	0.709	
E: Riskless Rate Inferred from Estimated SDFs		
T, Time Periods	-1.783E-05	-1.277
N, Assets	1.906E-06	1.092
True SDF Volatility	6.402E-03	0.764
Perturbation Scale	-2.588E-02	-0.713
Riskless Rate	1.395	4.212
Adjusted R <sup>2</sup>	0.00600	

**Appendix B**  
**Tests of Equality of Estimated SDFs from Two Groups of U.S. Equities**

The vector on the right side of (9) is an estimate based on  $N$  assets and a sample period of length  $T$ , a combination of cross-sectional and time series observations. The SDF paradigm contends that any **other** set of assets within the same integrated market should produce, aside from sampling variation, the same  $\hat{\mathbf{m}}$  from concurrent time series observations. Hence, if we denote by  $\hat{\mathbf{m}}(k)$  a sample  $\hat{\mathbf{m}}$  computed according to (9) (where  $k$  indicates a set of  $K$  assets,  $K > T$ ) and then, from the same calendar observations, choose a complement set  $j \not\subset k$  with  $J$  assets (and  $J > T$ ), the SDF null hypothesis of market integration can be expressed as

$$H_0: E[\hat{\mathbf{m}}(k) - \hat{\mathbf{m}}(j)] = \mathbf{0}. \quad (\text{B-1})$$

Notice that  $K$  and  $J$  need not be equal, but both must be larger than  $T$ .

This test is reminiscent of DeSantis (1993) and Ferson (1995), who suggest comparing SDFs derived from a subset of assets to SDFs derived from all available assets. Testing for the equivalence of pricing operators across two groups of assets is also explored by Chen and Knez (1994)<sup>33</sup> and, in the context of the APT, by Brown and Weinstein (1983).<sup>34</sup>

It is important to emphasize that the philosophy of the above test is standard; i.e., we will never be able to prove that two SDFs are exactly the same and that compared markets are indeed completely integrated, but we do have the possibility of rejecting these implications. If markets are not complete and integrated, an infinite number of stochastic discount factors satisfy (1) because  $E_{t-1}[(\tilde{m}_t + \tilde{\omega}_t)\tilde{p}_t] = E_{t-1}(\tilde{m}_t\tilde{p}_t)$  whenever  $\omega$  and  $p$  are orthogonal; Cf. Cochrane (2001a, Section 4.1). But  $\tilde{m}_t + \tilde{\omega}_t$  looks just like the true SDF plus an estimation error. Indeed, if markets are complete,  $\tilde{\omega}_t$  is an estimation error because  $\tilde{m}_t$  is unique. On the contrary, if markets are incomplete  $\tilde{\omega}_t$  can differ across groups of assets and hence the null hypothesis in (B-1) can potentially be rejected.

Many standard tests of equality could be employed for (B-1). For example, the Hotelling (1931)  $T^2$  test could check whether the means of  $\hat{\mathbf{m}}(k)$  and  $\hat{\mathbf{m}}(j)$  are statistically indistinguishable. The non-parametric Kruskal-Wallis (1952) test (hereafter KW) is designed for this purpose and will reject the null hypothesis if  $\hat{\mathbf{m}}(k)$  stochastically dominates  $\hat{\mathbf{m}}(j)$  or vice versa. This also provides a test of the equality of medians.

---

<sup>33</sup> Chen and Knez (1995) derive a measure of market integration as the minimal amount that two pricing operators differ. Their measure too does not rely on any restrictive pricing model assumptions and its applications do not depend on the validity of any asset pricing model. They use a similar framework to develop a general approach to portfolio performance measurement in Chen and Knez (1996).

<sup>34</sup> The Arbitrage Pricing Theory due to Ross (1976).

The Welch (1951) test (hereafter WE) would serve nicely to check whether the means of the two vectors are the same (i.e., that the both imply the same riskless rate.) 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 (hereafter BF) is designed specifically to check for unequal volatilities using absolute deviations.

The KW, WE and BF tests involve necessary conditions for the SDF paradigm. They can detect differences in, respectively, the medians, means and volatilities two estimated SDF vectors, but they are not capable of detecting time-dependent patterns of differences in the individual elements of the two vectors. For example, one vector might be increasing over time and the other decreasing but they could still have the same mean and variance.

The SDF paradigm stipulates not only that the location and volatility in SDFs are the same across groups of assets but also that SDF estimated realizations are the same in every time period. A sufficient condition is that the entire vectors  $\hat{\mathbf{m}}(k)$  and  $\hat{\mathbf{m}}(j)$  are congruent. Thus, we consider also a test that compares the two vectors element by element, a Hausman (1978) type Chi-Square test (hereafter CH.)<sup>35</sup>

To explain the Hausman type test in our application, let  $\hat{m}_{j,t}$  and  $\hat{m}_{k,t}$  denote the estimated SDF observation from asset groups  $j$  and  $k$  at time  $t$ . Under the null SDF hypothesis, they have the same expected value,  $\mu$ , and a common standard deviation,  $\sigma_t$ . Their correlation is  $\rho_t$ . Note that the correlation is not perfect because these are estimates of  $m$ , not the true values.

Under the null hypothesis, the variance of  $\hat{m}_{j,t} - \hat{m}_{k,t}$  is  $2\sigma_t^2(1-\rho_t)$ . Consequently, the standardized variate,

$$z_t = \frac{\hat{m}_{j,t} - \hat{m}_{k,t}}{\sigma_t \sqrt{2(1-\rho_t)}}$$

has mean zero and unit variance. When  $z$  is not autocorrelated,

$$\chi_T^2 = \sum_{t=1}^T z_t^2$$

converges asymptotically to a Chi-Square distribution with  $T$  degrees of freedom.<sup>36</sup>

The main implementation problem is, of course, that  $\sigma_t$  and  $\rho_t$  are unknown parameters that have to be estimated. Ignoring their time variation, this can be accomplished with the usual estimates over the sample of size  $T$ . However, since there are two estimated SDF vectors, even with this simplifying

---

<sup>35</sup> We are indebted to Ben Gillen for suggesting this test.

<sup>36</sup> If the SDF estimates are normally distributed and independent across time, the Chi-Square distribution is exact for any sample size.

assumption there would be two different estimates of  $\sigma$ . The most straightforward and sensible expedient is simply to average the two.

To implement these four tests, we employ the two groups of equities described in Section III of the paper, which have continuous simultaneous monthly observations for  $T=138$  months, July 2001 through December 2013. Each group consists of  $N=956$  stocks sorted into a lower- and higher-leveraged partition.

P-Values for the four tests of SDF equality between the two groups are reported in Table B-1. Since none of them is even close to a critical (low) value, there is no evidence against the null hypothesis that all 1912 stocks are traded in an integrated market with the same SDF.

Table B-1  
Tests of Equal Estimated SDFs from Two Groups of U.S. Equities

Stochastic discount factors (SDFs) are estimated independently for two groups of U.S. equities using simultaneous monthly gross return observations, July 2002 through December 2013, (T=138 months.) The two groups are, respectively, N=956 lower-leveraged equities and N=956 higher-leveraged equities.<sup>37</sup> These stocks all have continuous records over the sample period. 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. A Hausman (1978) type Chi-Square tests whether estimated SDF vectors are equal element by element. Low p-values in the table would reject the null hypothesis that all groups are priced with the same SDFs.

Stochastic Dominance (Kruskal/Wallis)	Equal Means (Welch)	Equal Variances (Brown/Forsythe)	Equal Elements (Chi-Square)
p-values: Lower- vs. Higher-Leverage Groups of 956 Equities Each			
0.679	0.995	0.808	0.457

---

<sup>37</sup> The Lower (Higher) leveraged group has an average book debt to assets ratio of 10.21% (32.51%).

**Appendix C**  
**Proof of Theorem 2.1, from Kim et al (2019)**

The pricing equation of  $\frac{1}{T} \mathbf{R} \mathbf{m} = \mathbf{1}_N + \varepsilon$  implies that

$$\hat{\mathbf{m}} - \mathbf{m} = -T(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\varepsilon.$$

Let  $\lambda_{max}(\mathbf{A})$ ,  $\lambda_{min}(\mathbf{A})$  and  $\text{tr}(\mathbf{A})$  denote the maximum eigenvalue, the minimum eigenvalue, and the trace of a square matrix  $\mathbf{A}$ , respectively. Then, we have that

$$\begin{aligned} \frac{1}{T}(\hat{\mathbf{m}} - \mathbf{m})'(\hat{\mathbf{m}} - \mathbf{m}) &= -T\varepsilon'\mathbf{R}(\mathbf{R}'\mathbf{R})^{-2}\mathbf{R}'\varepsilon = T\text{tr}(\mathbf{R}(\mathbf{R}'\mathbf{R})^{-2}\mathbf{R}'\varepsilon\varepsilon') \\ &\leq T\lambda_{max}(\mathbf{R}(\mathbf{R}'\mathbf{R})^{-2}\mathbf{R}')\text{tr}(\varepsilon\varepsilon') \\ &= T\lambda_{max}(((\mathbf{R}'\mathbf{R})^{-1})\varepsilon'\varepsilon) = \frac{T}{N}(\lambda_{min}\left(\frac{\mathbf{R}'\mathbf{R}}{N}\right)(\varepsilon\varepsilon')\left(\lambda_{min}\left(\frac{\mathbf{R}'\mathbf{R}}{N}\right)\right)^{-1})\varepsilon'\varepsilon \end{aligned}$$

where the second equality is from (A.1), the inequality is from (A.2) and the third equality is from (A.1) and the property of eigenvalue (Greene (2008) page 270.)

In addition, Assumption 1 (i), (ii), and (iii) imply  $\frac{T}{N} \rightarrow 0$ ,  $\lambda_{min}\left(\frac{\mathbf{R}'\mathbf{R}}{N}\right) > c$ , the boundedness of  $\varepsilon'\varepsilon$ , respectively, which, in conjunction with (B.1), yield

$$\frac{1}{T}(\hat{\mathbf{m}} - \mathbf{m})'(\hat{\mathbf{m}} - \mathbf{m}) \rightarrow 0$$

This completes the proof of the theorem.

**Supplementary Proofs:**

Let  $\lambda_{max}(\mathbf{A})$  and  $\text{tr}(\mathbf{A})$  denote the maximum eigenvalue and the trace of a square matrix  $\mathbf{A}$ , respectively. The following properties of eigenvalues and trace operator, vectorize operator are useful for the proof of the lemmas:

(A.1) Consider a  $(L \times M)$  matrix of  $\mathbf{A}$  and  $(M \times L)$  matrix of  $\mathbf{B}$ . Then, it holds that

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA}).$$

(A.2) Consider  $(L \times L)$  positive semidefinite matrices of  $\mathbf{A}$ ,  $\mathbf{B}$ . Then, it holds

$$\text{tr}(\mathbf{AB}) \leq \lambda_{max}(\mathbf{A})\text{tr}(\mathbf{B})$$