F U N D A Ç Ã O
GETULIO VARGAS

EPGE

Escola de Pós-Graduação
em Economia

Ensaios Econômicos

Escola de

Pós-Graduação

em Economia

da Fundação

Getulio Vargas

A Stochastic discount factor approach to asset pricing using panel data asymptotics

Fabio Araújo, João Victor Issler

Maio de 2011

Os artigos publicados são de inteira responsabilidade de seus autores. As opiniões neles emitidas não exprimem, necessariamente, o ponto de vista da Fundação Getulio Vargas.

A Stochastic Discount Factor Approach to Asset Pricing using Panel Data Asymptotics[*]

Fabio Araujo

Department of Economics

Princeton University

email: faraujo@princeton.edu

João Victor Issler

Graduate School of Economics – EPGE

Getulio Vargas Foundation

email: jissler@fgv.br[†]

This Draft: May, 2011.

Keywords: Stochastic Discount Factor, No-Arbitrage, Common Features, Panel-Data Econometrics.

J.E.L. Codes: C32, C33, E21, E44, G12.

Abstract

Using the Pricing Equation in a panel-data framework, we construct a novel consistent estimator of the stochastic discount factor (SDF) which relies on the fact that its logarithm is the "common feature" in every asset return of the economy. Our estimator is a simple function of asset returns and does not depend on any parametric function representing preferences.

The techniques discussed in this paper were applied to two relevant issues in macroeconomics and finance: the first asks what type of parametric preference-representation could be validated by asset-return data, and the second asks whether or not our SDF estimator can price returns in an out-of-sample forecasting exercise.

In formal testing, we cannot reject standard preference specifications used in the macro/finance literature. Estimates of the relative risk-aversion coefficient are between 1 and 2, and statistically equal to unity.

We also show that our SDF proxy can price reasonably well the returns of stocks with a higher capitalization level, whereas it shows some difficulty in pricing stocks with a lower level of capitalization.

190 s. 1100, Rio de Janeiro, RJ 22253-900, Brazil.

1 Introduction

In this paper, we derive a novel consistent estimator of the stochastic discount factor (or pricing kernel) that takes seriously the consequences of the Pricing Equation established by Harrison and Kreps (1979), Hansen and Richard (1987), and Hansen and Jagannathan (1991), where asset prices today are a function of their expected future payoffs discounted by the stochastic discount factor (SDF). If the Pricing Equation is valid for all assets at all times, it can serve as a basis to construct an estimator of the SDF in a panel-data framework when the number of assets and time periods is sufficiently large. This is exactly the approach taken here.

We start with a general Taylor Expansion of the Pricing Equation to derive the determinants of the *logarithm* of returns once we impose the moment restriction implied by the Pricing Equation. The identification strategy employed to recover the *logarithm* of the SDF relies on one of its basic properties – it is a "common feature," in the sense of Engle and Kozicki (1993), of every asset return of the economy. Under mild restrictions on the behavior of asset returns, used frequently elsewhere, we show how to construct a consistent estimator for the SDF which is a simple function of the arithmetic and geometric averages of asset returns alone, and does not depend on any parametric function used to characterize preferences.

A major benefit of our approach is that we are able to study intertemporal asset pricing without the need to characterize preferences or to use of consumption data; see a similar approach by Hansen and Jagannathan (1991, 1997). This yields several advantages of our SDF estimator over possible alternatives. First, since it does not depend on any parametric assumptions about preferences, there is no risk of misspecification in choosing an inappropriate functional form for the estimation of the SDF. Moreover, our estimator can be used to test directly different parametric-preference specifications commonly used in finance and macroeconomics. Second, since it does not depend on consumption data, our estimator does not inherit the smoothness observed in previous consumption-based estimates which generated important puzzles in finance and in macroeconomics, such as excess smoothness (excess sensitivity) in consumption, the equity-premium puzzle,

etc.; see Hansen and Singleton (1982, 1983, 1984), Mehra and Prescott (1985), Campbell (1987), Campbell and Deaton (1989), and Epstein and Zin (1991).

Our approach is related to research done in three different fields. From econometrics, it is related to the common-features literature after Engle and Kozicki (1993). Indeed, we attempt to bridge the gap between a large literature on serial-correlation common features applied to macroeconomics, e.g., Vahid and Engle (1993, 1997), Engle and Issler (1995), Issler and Vahid (2001, 2006), Vahid and Issler (2002), Hecq, Palm and Urbain (2005), Issler and Lima (2009), Athanasopoulos et al. (2011), and the financial econometrics literature related to the SDF approach, perhaps best represented by Chapman (1998), Aït-Sahalia and Lo (1998, 2000), Rosenberg and Engle (2002), Garcia, Luger, and Renault (2003), Garcia, Renault, and Semenov (2006), Hansen and Scheinkman (2009), and Hansen and Renault (2009). It is also related respectively to work on common factors in macroeconomics and in finance; see Geweke (1977), Stock and Watson (1989, 1993, 2002) Forni et al. (2000), and Bai and Ng (2004) as examples of the former, and a large literature in finance perhaps best exemplified by Fama and French (1992, 1993), Lettau and Ludvigson (2001), Sentana (2004), and Sentana, Calzolari, and Fiorentini (2008) as examples of the latter. From macroeconomics, it is related to the work using panel data for testing optimal behavior in consumption, e.g., Runkle (1991), Blundell, Browning, and Meghir (1994), Attanasio and Browning (1995), Attanasio and Weber (1995), and to the work of Mulligan (2002) on cross-sectional aggregation and intertemporal substitution.

The set of assumptions needed to derive our results are common to many papers in financial econometrics: the lack of arbitrage opportunities in pricing securities is assumed in virtually all studies estimating the SDF, and the restrictions (discipline) we impose on the stochastic behavior of asset returns are fairly standard. What we see as non-standard in our approach is an attempt to bridge the gap between economic and econometric theory in devising an econometric estimator of a random process which has a straightforward economic interpretation: it is the common feature of asset returns. Once the estimation problem is put in these terms, it is straightforward to apply panel-data techniques to construct a consistent estimator for the SDF. By construction, it will not depend on any

parametric function used to characterize preferences, which we see as a major benefit following the arguments in the seminal work of Hansen and Jagannathan (1991, 1997).

In a first application, with quarterly data on U.S.\$ real returns, ultimately using thousands of assets available to the average U.S. investor, our estimator of the SDF is close to unity most of the time and bound by the interval $[0.85, 1.15]$, with an equivalent average annual discount factor of 0.9711, or an average annual real discount rate of 2.97%. When we examined the appropriateness of different functional forms to represent preferences, we concluded that standard preference representations cannot be rejected by the data. Moreover, estimates of the relative risk-aversion coefficient are close to what can be expected *a priori* – between 1 and 2, statistically significant, and not different from unity in statistical tests. In a second application, we tried to approximate the asymptotic environment directly, working with monthly U.S. time-series return data with $T = 336$ observations, collected for a total of $N = 16,193$ assets. Using the δ distance measure of Hansen and Jagannathan (1997), we show that our SDF proxy can price reasonably well the returns of stocks with a high capitalization value, although it shows some difficulty in pricing stocks of firms with a low level of capitalization.

The next Section presents basic theoretical results, our estimation techniques, and a discussion of our main result. Section 3 shows the results of empirical tests in macro-economics and finance using our estimator: estimating preference parameters using the Consumption-based Capital Asset-Pricing Model (CCAPM) and out-of-sample evaluation of the Asset-Pricing Equation. Section 4 concludes.

2 Economic Theory and SDF Estimator

2.1 A Simple Consistent Estimator

Harrison and Kreps (1979), Hansen and Richard (1987), and Hansen and Jagannathan (1991) describe a general framework to asset pricing, associated to the stochastic discount

factor (SDF), which relies on the Pricing Equation[1]:

$$\mathbb{E}_t \left\{ M_{t+1} x_{i,t+1} \right\} = p_{i,t}, \qquad i = 1, 2, \ldots, N, \ \text{or} \tag{1}$$

$$\mathbb{E}_t \left\{ M_{t+1} R_{i,t+1} \right\} = 1, \qquad i = 1, 2, \ldots, N, \tag{2}$$

where $\mathbb{E}_t(\cdot)$ denotes the conditional expectation given the information available at time t, M_t is the stochastic discount factor, $p_{i,t}$ denotes the price of the i-th asset at time t, $x_{i,t+1}$ denotes the payoff of the i-th asset in $t+1$, $R_{i,t+1} = \frac{x_{i,t+1}}{p_{i,t}}$ denotes the gross return of the i-th asset in $t+1$, and N is the number of assets in the economy.

The existence of a SDF M_{t+1} that prices assets in (1) is obtained under very mild conditions. In particular, there is no need to assume a complete set of security markets. Uniqueness of M_{t+1}, however, requires the existence of complete markets. If markets are incomplete, i.e., if they do not span the entire set of contingencies, there will be an infinite number of stochastic discount factors M_{t+1} pricing all traded securities. Despite that, there will still exist a unique discount factor M_{t+1}^*, which is an element of the payoff space, pricing all traded securities. Moreover, any discount factor M_{t+1} can be decomposed as the sum of M_{t+1}^* and an error term orthogonal to payoffs, i.e., $M_{t+1} = M_{t+1}^* + \nu_{t+1}$, where $\mathbb{E}_t \left(\nu_{t+1} x_{i,t+1} \right) = 0$. The important fact here is that the pricing implications of any M_{t+1} are the same as those of M_{t+1}^*, also known as the mimicking portfolio.

We now state the four basic assumptions needed to construct our estimator:

Assumption 1: We assume the absence of arbitrage opportunities in asset pricing, c.f., Ross (1976). This must hold instantaneously for all $t = 1, 2, ..., T$, i.e., it must hold at all times and for all lapses of time, however small.

Assumption 2: Let $\mathbf{R}_t = (R_{1,t}, R_{2,t}, ... \ R_{N,t})'$ be an $N \times 1$ vector stacking all asset returns in the economy and consider the vector process $\{\ln(M_t \mathbf{R}_t)\}$. In the time (t) dimension, we assume that $\{\ln(M_t \mathbf{R}_t)\}_{t=1}^{\infty}$ is covariance-stationary and ergodic with finite first and second moments uniformly across i.

---

[1] See also Rubinstein(1976) and Ross(1978).

At a basic level, Assumption 1 is a necessary and sufficient condition for the Pricing Equation (2) to hold; see Cochrane (2002). Under the assumptions in Hansen and Renault (2009), Assumption 1 implies (2). In any case, (2) is present, either implicitly or explicitly, in virtually all studies in finance and macroeconomics dealing with asset pricing and/or with intertemporal substitution; see, e.g., Hansen and Singleton (1982, 1983, 1984), Mehra and Prescott (1985), Epstein and Zin (1991), Fama and French (1992, 1993), Attanasio and Browning (1995), Lettau and Ludvigson (2001), Garcia, Renault, and Semenov (2006), Hansen and Scheinkman (2009) and Hansen and Renault (2009). It is essentially equivalent to the "law of one price" – where securities with identical payoffs in all states of the world must have the same price. We impose its validity instantaneously since we will derive a logarithmic representation for (2), which allows exact measure of instantaneous returns for all assets.

The absence of arbitrage opportunities has also two other important implications. The first is there exists at least one stochastic discount factor M_t, for which $M_t > 0$; see Hansen and Jagannathan (1997). This is due to the fact that, when we consider the existence derivatives on traded assets, arbitrage opportunities will arise if $M_t \leq 0$. Positivity of some M_t is required here because we will take logs of M_t in proving our asymptotic results[2]. The second is that the absence of arbitrage requires that a weak law-of-large numbers (WLLN) holds in the cross-sectional dimension for the level of gross returns $R_{i,t}$ (Ross (1976, p. 342)). This controls the degree of cross-sectional dependence in the data and constitutes the basis of the arbitrage pricing theory (APT). Applying the Ergodic Theorem in the cross-sectional dimension, implies that we should also expect a WLLN to hold for its logarithmic counterpart ($\ln R_{i,t}$), forming the basis of our asymptotic results.

Assumption 2 controls the degree of time-series dependence in the data. Across time (t), asset returns have clear signs of heterogeneity: different means and variances, and conditional heteroskedasticity; as examples of the latter see Bollerslev, Engle and Wooldridge (1988) and Engle and Marcucci (2006). Of course, weak-stationary processes can display

---

[2]Recall that all CCAPM studies implicitly assume $M_t > 0$, since $M_t = \beta \frac{u'(c_t)}{u'(c_{t-1})} > 0$, where c_t is consumption, $\beta \in (0,1)$ and $u'(\cdot) > 0$.

conditional heteroskedasticity as long as second moments are finite; see Engle (1982). Therefore, Assumption 2 allows for heterogeneity in mean returns and conditional heteroskedasticity in returns used in computing our estimator. Uniformity across (i) is required for technical reasons, since we want the mean across first and second moments of returns to be defined.

To construct a consistent estimator for $\{M_t\}$ we consider a second-order Taylor Expansion of the exponential function around x, with increment h, as follows:

$$e^{x+h} = e^x + he^x + \frac{h^2 e^{x+\lambda(h)\cdot h}}{2}, \tag{3}$$

$$\text{with } \lambda(h) : \mathbb{R} \rightarrow (0,1). \tag{4}$$

It is important to stress that (3) is an exact relationship and not an approximation. This is due to the nature of the function $\lambda(h) : \mathbb{R} \rightarrow (0,1)$, which maps into the open unit interval. Thus, the last term is evaluated between x and $x+h$, making (3) to hold exactly.

For the expansion of a generic function, $\lambda(\cdot)$ would depend on x and h. However, dividing (3) by e^x:

$$e^h = 1 + h + \frac{h^2 e^{\lambda(h)\cdot h}}{2}, \tag{5}$$

shows that (5) does not depend on x. Therefore, we get a closed-form solution for $\lambda(\cdot)$ as function of h alone:

$$\lambda(h) = \begin{cases} \frac{1}{h} \times \ln\left[\frac{2\times\left(e^h - 1 - h\right)}{h^2}\right], & h \neq 0 \\ \\ 1/3, & h = 0, \end{cases}$$

where $\lambda(\cdot)$ maps from the real line into $(0,1)$. To connect (5) with the Pricing Equation (2), we impose $h = \ln(M_t R_{i,t})$ in (5) to obtain:

$$M_t R_{i,t} = 1 + \ln(M_t R_{i,t}) + \frac{[\ln(M_t R_{i,t})]^2\, e^{\lambda(\ln(M_t R_{i,t}))\cdot \ln(M_t R_{i,t})}}{2}, \tag{6}$$

which shows that the behavior of $M_t R_{i,t}$ will be governed solely by that of $\ln(M_t R_{i,t})$.

It is useful to define the random variable collecting the higher order term of (6):

$$z_{i,t} \equiv \frac{1}{2} \times [\ln(M_t R_{i,t})]^2 \, e^{\lambda(\ln(M_t R_{i,t}))\cdot \ln(M_t R_{i,t})}.$$

Taking the conditional expectation of both sides of (6) gives:

$$\mathbb{E}_{t-1}\left(M_t R_{i,t}\right) = 1 + \mathbb{E}_{t-1}\left(\ln(M_t R_{i,t})\right) + \mathbb{E}_{t-1}\left(z_{i,t}\right). \tag{7}$$

As a direct consequence of the Pricing Equation, the left-hand side cancels with the first term of the right-hand side of (7), yielding:

$$\mathbb{E}_{t-1}\left(z_{i,t}\right) = -\mathbb{E}_{t-1}\left\{\ln(M_t R_{i,t})\right\}. \tag{8}$$

This first shows that $\mathbb{E}_{t-1}\left(z_{i,t}\right)$ will be solely a function of $\mathbb{E}_{t-1}\left\{\ln(M_t R_{i,t})\right\}$ if the Pricing Equation holds, otherwise it will also be a function of $\mathbb{E}_{t-1}(M_t R_{i,t})$. Second, $z_{i,t} \geq 0$ for all (i,t). Therefore, $\mathbb{E}_{t-1}\left(z_{i,t}\right) \equiv \gamma_{i,t}^2 \geq 0$, and we denote it as $\gamma_{i,t}^2$ to stress the fact that it is non-negative.

Let $\boldsymbol{\gamma}_t^2 \equiv \left(\gamma_{1,t}^2, \gamma_{2,t}^2, ..., \gamma_{N,t}^2\right)'$ and $\boldsymbol{\varepsilon}_t \equiv \left(\varepsilon_{1,t}, \varepsilon_{2,t}, ..., \varepsilon_{N,t}\right)'$ stack respectively the conditional means $\gamma_{i,t}^2$ and the forecast errors $\varepsilon_{i,t}$. Then, from the definition of $\boldsymbol{\varepsilon}_t$ we have:

$$\begin{aligned}
\ln(M_t \mathbf{R}_t) &= \mathbb{E}_{t-1}\{\ln(M_t \mathbf{R}_t)\} + \boldsymbol{\varepsilon}_t \\
&= -\boldsymbol{\gamma}_t^2 + \boldsymbol{\varepsilon}_t.
\end{aligned} \tag{9}$$

Denoting by $\mathbf{r}_t = \ln\left(\mathbf{R}_t\right)$, which elements are denoted by $r_{i,t} = \ln\left(R_{i,t}\right)$, and by $m_t = \ln\left(M_t\right)$, (9) yields the following system of equations:

$$r_{i,t} = -m_t - \gamma_{i,t}^2 + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N. \tag{10}$$

The system (10) shows that the (log of the) SDF is a common feature, in the sense of Engle and Kozicki (1993), of all (logged) asset returns. For any two economic series, a common feature exists if it is present in both of them and can be removed by linear

combination. Hansen and Singleton (1983) were the first authors to exploit this property of (logged) asset returns, although the concept was only proposed 10 years later by Engle and Kozicki.

Looking at (10), asset returns are decomposed into three terms, but we focus on the first – the logarithm of the SDF, m_t, which is common to all returns and has random variation only across time. Notice that m_t can be removed by linearly combining returns: for any two assets i and j, $r_{i,t} - r_{j,t}$ will not contain the feature m_t, which makes $(1, -1)$ a "cofeature vector" for all asset pairs.

We label (10) as a *quasi-structural system* for logged returns, since its foundation is the Asset-Pricing Equation (1). Equation (10) can be thought as a factor model for $r_{i,t}$, where the common factor m_t has only time-series variation. Indeed, this is the logarithmic counterpart of the common-factor model assumed by Ross (1976) for the level of returns $R_{i,t}$, where here the Pricing Equation (1) provides a solid *structural* foundation to it.

The sources of cross-sectional variation in every equation of the system (10) are $\varepsilon_{i,t}$ and $\gamma_{i,t}^2$. However, as we show next, the terms $\gamma_{i,t}^2$ are a linear function of lagged $\varepsilon_{i,t}$, tying the cross-sectional variation in (10) ultimately to $\varepsilon_{i,t}$.

Start with Assumption 2. Because $\ln(M_t \mathbf{R}_t)$ is weakly stationary, for every one of its elements $\ln(M_t R_{i,t})$, there exists a Wold representation, which is a linear function of the innovation in $\ln(M_t R_{i,t})$, defined as $\varepsilon_{i,t} = \ln(M_t R_{i,t}) - \mathbb{E}_{t-1}\{\ln(M_t R_{i,t})\}$ and stacked in $\boldsymbol{\varepsilon}_t \equiv (\varepsilon_{1,t}, \varepsilon_{2,t}, \ ..., \varepsilon_{N,t})'$. Therefore, the individual Wold representations can be written as:

$$\ln(M_t R_{i,t}) = \mu_i + \sum_{j=0}^{\infty} b_{i,j} \varepsilon_{i,t-j}, \qquad i = 1, 2, \dots, N, \tag{11}$$

where, for all i, $b_{i,0} = 1$, $|\mu_i| < \infty$, $\sum_{j=0}^{\infty} b_{i,j}^2 < \infty$, and $\varepsilon_{i,t}$ is a multivariate white noise. Using (8), in light of (11), leads to:

$$\gamma_i^2 \equiv \mathbb{E}(z_{i,t}) = -\mathbb{E}\{\ln(M_t R_{i,t})\} = -\mu_i, \tag{12}$$

which is well defined and time-invariant under Assumption 2. Taking conditional expectations $\mathbb{E}_{t-1}(\cdot)$ of (11), allows computing $\gamma_{i,t}^2 = \mathbb{E}_{t-1}(z_{i,t}) = -\mathbb{E}\{\ln(M_t R_{i,t})\}$, leading to

the following system, once we consider (10):

$$r_{i,t} = -m_t - \gamma_i^2 + \varepsilon_{i,t} - \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}, \qquad i = 1, 2, \ldots, N. \tag{13}$$

This is just a different way of writing $(10)^3$. Because m_t is devoid of cross-sectional variation, (13) shows that the ultimate source of cross-sectional variation for $r_{i,t}$ is $\varepsilon_{i,t}$ (and its lags). This paves the way to derive a consistent estimator for M_t based on the existence of a WLLN for $\{\varepsilon_{i,t}\}_{i=1}^N$. This is consistent with $\lim_{N\to\infty} VAR\left(\frac{1}{N}\sum_{i=1}^N \varepsilon_{i,t}\right) = 0$, but the critical issue is whether or not $\frac{1}{N}\sum_{i=1}^N \varepsilon_{i,t} \xrightarrow{p} 0$. If that were the case, it would be straightforward to compute $\plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N r_{i,t} + m_t$ and then construct a a consistent estimator for M_t.

Convergence in probability for logged returns $r_{i,t}$ is not surprising, given the assumption of convergence in probability for the levels of returns $R_{i,t}$ behind the APT. After all, $r_{i,t} = \ln(R_{i,t})$ is a measurable transformation of $R_{i,t}$. By applying the Ergodic Theorem in the cross-sectional dimension, we should also expect that a WLLN holds for $\{r_{i,t}\}_{i=1}^N$ as well. Despite that, one may be skeptical of:

$$\frac{1}{N}\sum_{i=1}^N \varepsilon_{i,t} \xrightarrow{p} 0. \tag{14}$$

Equation (14) may seem restrictive because we can always decompose $\varepsilon_{i,t}$ as:

$$\varepsilon_{i,t} = \ln(M_t R_{i,t}) - \mathbb{E}_{t-1}\{\ln(M_t R_{i,t})\} \tag{15}$$

$$= [m_t - \mathbb{E}_{t-1}(m_t)] + [r_{i,t} - \mathbb{E}_{t-1}(r_{i,t})] = q_t + v_{i,t}, \tag{16}$$

---

[3] Here it becomes obvious that:

$$\gamma_{i,t}^2 = \gamma_i^2 + \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}$$

$$= -\mu_i + \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}.$$

where $q_t = [m_t - \mathbb{E}_{t-1}(m_t)]$ is the innovation in m_t and $v_{i,t} = [r_{i,t} - \mathbb{E}_{t-1}(r_{i,t})]$ is the innovation in $r_{i,t}$. Therefore, to get $\plim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t} = 0$, we need,

$$\plim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} v_{i,t} = -q_t, \tag{17}$$

which may seem like a knife-edge restriction on the cross-sectional distribution of $v_{i,t}$. Indeed, it is not. To show it, consider the argument of projecting $v_{i,t}$ into q_t, collecting terms, and decomposing $\varepsilon_{i,t}$ as follows:

$$\varepsilon_{i,t} = \delta_i q_t + \xi_{i,t}, \text{ where } \delta_i \equiv \frac{\text{COV}(\varepsilon_{i,t}, q_t)}{\text{VAR}(q_t)} = 1 + \frac{\text{COV}(v_{i,t}, q_t)}{\text{VAR}(q_t)}. \tag{18}$$

Here, we collect all that is pervasive in q_t and thus it is reasonable to assume that $\plim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \xi_{i,t} = 0$. In this context of the factor model (18), in order to get $\plim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t} = 0$, we must have:

$$\plim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \xi_{i,t} = -\delta q_t, \text{ where } \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_i = \delta. \text{ Thus,} \tag{19}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_i = \delta = 0, \text{ or } \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\text{COV}(v_{i,t}, q_t)}{\text{VAR}(q_t)} = -1. \tag{20}$$

Equation (20) highlights that the issue is not one of a knife-edge restriction. In order to obtain $\plim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t} = 0$, and use $\plim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} r_{i,t} + m_t$ to construct a consistent estimator for M_t, the average factor loading must obey $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_i = 0$. Notice that $v_{i,t}$ is an innovation coming from data $(r_{i,t})$, but q_t is an innovation coming from the latent variable m_t, which makes this an issue of separate identification of the factor (q_t) and of its respective factor loadings (δ_i).

Next, we state our most important result: a novel consistent estimator of the stochastic process $\{M_t\}_{t=1}^{\infty}$. Instead of using the Ergodic Theorem, we chose a more intuitive asymptotic approach based on no-arbitrage, where the *quasi-structural* system (10) serves as a basis to measure instantaneous returns of no-arbitrage portfolios. In our proof, we

12

use directly the projection argument in (18) to show that no-arbitrage will indeed deliver the seemingly knife-edge restriction $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_i = 0$. In our discussion of the main result below, we exploit further the econometric identification issue raised above.

Theorem 1 *Under Assumptions 1 and 2, as $N, T \to \infty$, with N diverging at a rate at least as fast as T, the realization of the SDF at time t, denoted by M_t, can be consistently estimated using:*

$$\widehat{M_t} = \frac{\overline{R}_t^G}{\frac{1}{T} \sum_{j=1}^{T} \left(\overline{R}_j^G \overline{R}_j^A \right)},$$

where $\overline{R}_t^G = \prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}}$ and $\overline{R}_t^A = \frac{1}{N} \sum_{i=1}^{N} R_{i,t}$ are respectively the geometric average of the reciprocal of all asset returns and the arithmetic average of all asset returns.

Proof. Consider a cross-sectional average of (13):

$$\frac{1}{N} \sum_{i=1}^{N} r_{i,t} + m_t = -\frac{1}{N} \sum_{i=1}^{N} \gamma_i^2 + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t} - \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{\infty} b_{i,j} \varepsilon_{i,t-j}, \tag{21}$$

and examine convergence in probability of $\frac{1}{N} \sum_{i=1}^{N} r_{i,t} + m_t$ using (21).

First, because every term $\ln(M_t R_{i,t})$ has a finite mean $\mu_i = -\gamma_i^2$, uniformly across i, the limit of their average must be finite, i.e.,

$$\lim_{N\to\infty} -\frac{1}{N} \sum_{i=1}^{N} \gamma_i^2 \equiv -\gamma^2 < \infty. \tag{22}$$

Second, there is no correlation across time for the elements in $\boldsymbol{\varepsilon}_t \equiv \left(\varepsilon_{1,t} \quad \varepsilon_{2,t} \quad ... \quad \varepsilon_{N,t} \right)'$, due to the assumption of weak stationarity for the vector process $\{\ln(M_t \mathbf{R}_t)\}$. Hence, $\mathbb{E} \left(\varepsilon_{i,t} \quad \varepsilon_{h,t-j} \right) = 0$, for all i and h, and all $j \geq 1$. Therefore, the asymptotic variance of $\frac{1}{N} \sum_{i=1}^{N} r_{i,t} + m_t$ in the cross-sectional dimension has the following form:

$$\lim_{N\to\infty} VAR \left(\frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t} \right) + \lim_{N\to\infty} VAR \left(\frac{1}{N} \sum_{i=1}^{N} b_{i,1} \varepsilon_{i,t-1} \right) + \lim_{N\to\infty} VAR \left(\frac{1}{N} \sum_{i=1}^{N} b_{i,2} \varepsilon_{i,t-2} \right) + \cdots. \tag{23}$$

13

Below, we will exploit the form of (23) in proving consistency of our estimator.

Notice that we have assumed that the absence of arbitrage opportunities must hold instantaneously, where the level of returns $R_{i,t}$ and its instantaneous counterpart $r_{i,t}$ are identical. It is then intuitive that if a WLLN applies to $\{R_{i,t}\}_{i=1}^{N}$ it should apply to $\{r_{i,t}\}_{i=1}^{N}$ as well.

Large-sample arbitrage portfolios are characterized by weights w_i, all of order N^{-1} in absolute value, stacked in a vector $W = (w_1, w_2, \ ..., w_N)'$, with the following properties:

$$\text{(a)} \lim_{N\to\infty} W' \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = 0, \text{ and (b)} \lim_{N\to\infty} \text{VAR} \left[W' \begin{pmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N,t} \end{pmatrix} \right] = 0. \qquad (24)$$

Condition (a) implies that these portfolios cost nothing. Condition (b) implies that their return is not random. In this context, no-arbitrage requires that all large-sample portfolios W must also have a zero limit return, in probability:

$$\plim_{N\to\infty} W' \begin{pmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N,t} \end{pmatrix} = 0. \qquad (25)$$

Notice that we need strict equality in (25). Condition $\plim_{N\to\infty} W' \begin{pmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N,t} \end{pmatrix} \leq 0$ does not

work because if we find a portfolio W for which $\plim_{N\to\infty} W' \begin{pmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N,t} \end{pmatrix} < 0$, we could violate no

14

arbitrage by using portfolio $-W$: it obeys (24) and would have $\plim_{N \to \infty} -W' \begin{pmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N,t} \end{pmatrix} > 0.$

Start with the stacked *quasi-structural form* for logged returns:

$$\begin{pmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N,t} \end{pmatrix} = -m_t \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} - \begin{pmatrix} \gamma_1^2 \\ \gamma_2^2 \\ \vdots \\ \gamma_N^2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{N,t} \end{pmatrix} + \begin{pmatrix} \sum_{j=1}^\infty b_{1,j}\varepsilon_{1,t-j} \\ \sum_{j=1}^\infty b_{2,j}\varepsilon_{2,t-j} \\ \vdots \\ \sum_{j=1}^\infty b_{N,j}\varepsilon_{N,t-j} \end{pmatrix}$$

From condition (a) in (24), *every* large-sample arbitrage portfolios removes the term m_t from the linear combination. From condition (b), in the limit, the variance of the arbitrage portfolio must be zero, which poses a constraint on the cross-sectional dependence of $\{\varepsilon_{i,t}\}_{i=1}^N$.

In what follows, we will prove that (23) is zero. Moreover, we will also prove that $\plim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \varepsilon_{i,t} = 0$, $\plim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N b_{i,1}\varepsilon_{i,t-1} = 0$, etc., using the factor model (18) for $\varepsilon_{i,t}$. To do so, we construct no-arbitrage portfolios and investigate what type of restriction they impose on the cross-sectional dependence of $\{\varepsilon_{i,t}\}_{i=1}^N$. We also show that portfolios W, which obey (24) and for which $\plim_{N \to \infty} W' \begin{pmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N,t} \end{pmatrix} = 0$, are inconsistent with:

$$\varepsilon_{i,t} = \delta_i q_t + \xi_{i,t}, \qquad \text{where} \qquad \frac{1}{N} \sum_{i=1}^N \xi_{i,t} \xrightarrow{p} 0. \tag{26}$$

Thus, a necessary condition for no-arbitrage is that $\varepsilon_{i,t}$ does not contain a factor q_t as in (26) above.

We start with the simplest form of limit arbitrage portfolios – buying $1/N$ units of even assets and selling $1/N$ units of odd assets; see the example in Chamberlain and

15

Rothschild (1983). We have two equally weighted portfolios (bought and sold assets) whose instantaneous returns are, respectively:

$$r_{e,t} = -m_t - \frac{1}{N/2}\sum_{i=1}^{N/2}\gamma_{2i}^2 + \frac{1}{N/2}\sum_{i=1}^{N/2}\varepsilon_{2i,t} - \frac{1}{N/2}\sum_{i=1}^{N/2}\sum_{j=1}^{\infty}b_{2i,j}\varepsilon_{2i,t-j}.$$

$$r_{o,t} = -m_t - \frac{1}{N/2}\sum_{i=1}^{N/2}\gamma_{2i-1}^2 + \frac{1}{N/2}\sum_{i=1}^{N/2}\varepsilon_{2i-1,t} - \frac{1}{N/2}\sum_{i=1}^{N/2}\sum_{j=1}^{\infty}b_{2i-1,j}\varepsilon_{2i-1,t-j}$$

The instantaneous return of the arbitrage portfolio is:

$$
\begin{aligned}
r_{e,t} - r_{o,t} &= -\frac{1}{N/2}\sum_{i=1}^{N/2}\left(\gamma_{2i}^2 - \gamma_{2i-1}^2\right) + \frac{1}{N/2}\sum_{i=1}^{N/2}\left(\varepsilon_{2i,t} - \varepsilon_{2i-1,t}\right) \\
&\quad - \frac{1}{N/2}\sum_{i=1}^{N/2}\sum_{j=1}^{\infty}\left(b_{2i,j}\varepsilon_{2i,t-j} - b_{2i-1,j}\varepsilon_{2i-1,t-j}\right),
\end{aligned}
\tag{27}
$$

which clearly eliminates the common-factor m_t in the linear combination of instantaneous returns. From (25), no arbitrage in large samples implies:

$$
\begin{aligned}
0 &= \plim_{N\to\infty}\frac{1}{N/2}\sum_{i=1}^{N/2}\left(\varepsilon_{2i,t} - \varepsilon_{2i-1,t}\right), \\
0 &= \plim_{N\to\infty}\frac{1}{N/2}\sum_{i=1}^{N/2}\left(b_{2i,1}\varepsilon_{2i,t-1} - b_{2i-1,1}\varepsilon_{2i-1,t-1}\right), \\
0 &= \plim_{N\to\infty}\frac{1}{N/2}\sum_{i=1}^{N/2}\left(b_{2i,2}\varepsilon_{2i,t-2} - b_{2i-1,2}\varepsilon_{2i-1,t-2}\right), \cdots \text{ etc.}
\end{aligned}
\tag{28}
$$

Notice that (28) requires convergence in probability for all stochastic terms in (27), since there is no cross-correlation of errors across lags of $\varepsilon_{i,t}$. Indeed, this is the only way their sum could converge to zero, in probability.

We look now at the first term of (28) in isolation, accounting for the factor structure

in (26):

$$
\begin{aligned}
0 &= \operatorname*{plim}_{N\to\infty} \frac{1}{N/2} \sum_{i=1}^{N/2} \left(\varepsilon_{2i,t} - \varepsilon_{2i-1,t} \right) \\
&= \operatorname*{plim}_{N\to\infty} \frac{1}{N/2} \sum_{i=1}^{N/2} \left[\left(\delta_{2i} q_t + \xi_{2i,t} \right) - \left(\delta_{2i-1} q_t + \xi_{2i-1,t} \right) \right] \\
&= \left[\lim_{N\to\infty} \frac{1}{N/2} \sum_{i=1}^{N/2} \left(\delta_{2i} - \delta_{2i-1} \right) \right] q_t + \operatorname*{plim}_{N\to\infty} \frac{1}{N/2} \sum_{i=1}^{N/2} \left(\xi_{2i,t} - \xi_{2i-1,t} \right) \\
&= \left[\lim_{N\to\infty} \frac{1}{N/2} \sum_{i=1}^{N/2} \left(\delta_{2i} - \delta_{2i-1} \right) \right] q_t. &(29)
\end{aligned}
$$

The cross-sectional dimension offers no natural order of assets, which is taken to be arbitrary here. Since (29) must hold for every possible permutation of odd and even assets, and for all possible realizations of q_t, in order to (29) to hold, we must have:

$$
0 = \lim_{N\to\infty} \frac{1}{N/2} \sum_{i=1}^{N/2} \left(\delta_{2i} - \delta_{2i-1} \right), \tag{30}
$$

i.e., limit weights of all permutations of odd and even assets must cancel out. Notice that this condition does not preclude the existence of a factor model as in (26) above. However, the factor model must have the following structure:

$$
\varepsilon_{i,t} = \delta q_t + \xi_{i,t},
$$

i.e., we must have $\delta_i = \delta$ across all assets. In this context, in order to rule out a factor structure we must have $\delta = 0$. This will indeed be the case, as we show below.

To exclude a factor structure for $\varepsilon_{i,t}$, we now look into the all the other (infinite) terms in (27). For lag one and for higher lags of $\varepsilon_{i,t}$, notice that we have potentially different loadings for the odd and even error terms in (32) above, due to the existence of the double

array $\{b_{i,j}\}$. This requires:

$$0 = \left[\lim_{N\to\infty} \frac{1}{N/2} \sum_{i=1}^{N/2} \left(\delta_{2i} b_{2i,1} - \delta_{2i-1} b_{2i-1,1} \right) \right] q_{t-1},$$

$$0 = \left[\lim_{N\to\infty} \frac{1}{N/2} \sum_{i=1}^{N/2} \left(\delta_{2i} b_{2i,2} - \delta_{2i-1} b_{2i-1,2} \right) \right] q_{t-2},$$

$$\vdots$$

$$\text{etc.} \tag{31}$$

Notice that, if $\varepsilon_{i,t}$ contains a common factor q_t, even if is eliminated for a given lag of $\varepsilon_{i,t}$, and all permutations of assets, it will not be eliminated at other lags, because the limit loadings will not necessarily match[4]. In this case,

$$\plim_{N\to\infty} \left(r_{e,t} - r_{o,t} \right)$$

will necessarily be a linear function of q_t and (of some or all) of its lags. Hence, for some realization of the random process $\{q_t\}_{t=1}^{\infty}$, we could not prevent that

$$\plim_{N\to\infty} \left(r_{e,t} - r_{o,t} \right) > 0 \text{ or } \plim_{N\to\infty} \left(r_{e,t} - r_{o,t} \right) < 0 \text{ holds.}$$

However, this violates no arbitrage: there exists a portfolio W (or $-W$), which obeys (24) – cost nothing and have no uncertain return – and for which $\plim_{N\to\infty} W' \begin{pmatrix} r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N,t} \end{pmatrix} > 0.$

Considering all possible realizations $\{q_t\}_{t=1}^{\infty}$, the only way to get $\plim_{N\to\infty} \left(r_{e,t} - r_{o,t} \right) = 0$

---

[4] Of course, we can always impose a structure to the double array $\{b_{i,j}\}$ such that the terms in brackets in (31) all cancel out. However, the $\{b_{i,j}\}$ come from the Wold decomposition, so we must treat them as given.

is to rule out completely any common factor q_t in $\varepsilon_{i,t}$. This leads to:

$$\varepsilon_{i,t} = \xi_{i,t}, \qquad \text{with } \frac{1}{N} \sum_{i=1}^{N} \xi_{i,t} \xrightarrow{p} 0,$$

implying:

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t} &= \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \xi_{i,t} = 0, \\
\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} b_{i,1}\varepsilon_{i,t-1} &= \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} b_{i,1}\xi_{i,t-1} = 0, \\
&\vdots \\
&\text{etc.}
\end{aligned}
\tag{32}
$$

Up to now, we only discussed one possible large-sample arbitrage portfolio – buying $1/N$ units of even assets and selling $1/N$ units of odd assets. But this is sufficient to show that (32) holds and we need not discuss any further other no-arbitrage portfolios[5].

Indeed, (32) proves that:

$$\frac{1}{N} \sum_{i=1}^{N} r_{i,t} + m_t \xrightarrow{p} \lim_{N\to\infty} - \frac{1}{N} \sum_{i=1}^{N} \gamma_i^2 \equiv \gamma^2. \tag{33}$$

In excluding the factor structure for $\varepsilon_{i,t}$, we had to resort to the restrictions implied by $\varepsilon_{i,t-1}$ and by higher lags of $\varepsilon_{i,t}$. However, even for the special case where the Wold representation has an $MA\,(0)$ structure, i.e.,

$$r_{i,t} = -m_t - \gamma_i^2 + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \tag{34}$$

---

[5]Considering all possible arbitrage portfolios only reinforces the previous result of ruling out a common factor model for $\varepsilon_{i,t}$, since we will necessarlily have to consider alternative weighting schemes to $\frac{1}{N}$ and $-\frac{1}{N}$ for even and odd assets, respectively. If the number of assets is "large," there is an infinite number of arbitrage portfolios.

our result still holds[6].

As before, our starting point is the fact that $r_{i,t} + m_t$ is weakly stationary, which allows writing it as a linear function of the innovation $\varepsilon_{i,t}$ as in (11) or (34) above as a consequence of Wold's Decomposition. As is well known, this proposition relies on the existence of stable second moments, i.e., Assumption 2. Because the relationship between $r_{i,t} + m_t$ and $\varepsilon_{i,t}$ is solely linear, we first eliminate the dependence of $\varepsilon_{i,t}$ on m_t by projecting $\varepsilon_{i,t}$ onto m_t. Using (34) and collecting terms leads to:

$$r_{i,t} = -\beta_i m_t - \gamma_i^2 + \eta_{i,t}, \qquad i = 1, 2, \ldots, N, \tag{35}$$

where, by construction, $\beta_i \equiv 1 - \frac{\text{COV}(\varepsilon_{i,t}, m_t)}{\text{VAR}(m_t)}$ and it becomes clear that $\plim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \eta_{i,t} = 0$, since $\eta_{i,t}$ is devoid of any pervasive factor. Notice that β_i is non-random for all i. Recall the Pricing Equation using the unconditional expectation operator:

$$\mathbb{E}\left[M_t R_{i,t}\right] = 1, \qquad i = 1, 2, \ldots, N. \tag{36}$$

Assume the usual regularity conditions and partially differentiate (36) with respect to m_t:

$$\frac{\partial}{\partial m_t} \mathbb{E}\left[\exp\left(r_{i,t} + m_t\right)\right] = \mathbb{E}\left[\frac{\partial}{\partial m_t} \exp\left(r_{i,t} + m_t\right)\right] =$$
$$\mathbb{E}\left[\exp\left(r_{i,t} + m_t\right) \times \left(\frac{\partial r_{i,t}}{\partial m_t} + 1\right)\right] = 0, \qquad i = 1, 2, \ldots, N.$$

Now, partially differentiate (35) with respect to m_t, recalling that $\eta_{i,t}$ does not depend on m_t. The result is the non-random coefficient $\frac{\partial r_{i,t}}{\partial m_t} = -\beta_i$. It then follows that, for $i = 1, 2, \ldots, N$:

$$\mathbb{E}\left[\exp\left(r_{i,t} + m_t\right) \times \left(\frac{\partial r_{i,t}}{\partial m_t} + 1\right)\right] = \mathbb{E}\left[M_t R_{i,t}\right] \times \left(\frac{\partial r_{i,t}}{\partial m_t} + 1\right) = 1 \times \left(\frac{\partial r_{i,t}}{\partial m_t} + 1\right) = 0.$$

---

[6]It is important to stress that (34) encompasses the canonical log-Normal, homoskedastic case, for $\left(\begin{array}{cccc} M_t, & R_{1,t}, & R_{2,t}, & \cdots & R_{N,t} \end{array}\right)'$, which is so prevalent in macroeconomics, but it is not constrained by these restrictive assumptions, including as well for the more general heteroskedastic case where log-Normality is dispensed with.

Thus:

$$\frac{\partial r_{i,t}}{\partial m_t} = -\beta_i = -1, \qquad i = 1, 2, \ldots, N,$$

leading to,

$$r_{i,t} = -m_t - \gamma_i^2 + \eta_{i,t}, \qquad i = 1, 2, \ldots, N. \tag{37}$$

Compare now (34) with (37) to conclude that $\varepsilon_{i,t} = \eta_{i,t}$, which is devoid of any pervasive factor[7], and for which $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t} \xrightarrow{p} 0$ holds. As before, this proves that (33) holds[8].

From (33), using Slutsky's Theorem, we can then propose a consistent estimator for a tilted version of M_t ($e^{\gamma^2} \times M_t = \widetilde{M_t}$):

$$\widehat{\widetilde{M_t}} = \prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}}. \tag{39}$$

We now show how to estimate e^{γ^2} consistently and therefore how to find a consistent estimator for M_t. Multiply the Pricing Equation for every asset by e^{γ^2} to get:

$$e^{\gamma^2} = \mathbb{E}_{t-1}\left\{ e^{\frac{\gamma^2}{2}} M_t R_{i,t} \right\} = \mathbb{E}_{t-1}\left\{ \widetilde{M_t} R_{i,t} \right\}.$$

Take now the unconditional expectation, use the law-of-iterated expectations, and average across $i = 1, 2, \ldots, N$ to get:

$$e^{\gamma^2} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left\{ \widetilde{M_t} R_{i,t} \right\}.$$

Because of Assumption 2, where $\{\ln(M_t \mathbf{R}_t)\}_{t=1}^{\infty}$ is covariance-stationary and ergodic, $\widetilde{M_t} R_{i,t}$ will keep these properties due to the Ergodic Theorem. Thus, it is straightforward

---

[7]From (37), it is straightforward to obtain a factor model for innovations as in (16). Take conditional expectations of (37). Subtracting it from (37) yields:

$$v_{i,t} = -q_t + \eta_{i,t}, \tag{38}$$

which makes clear that $\varepsilon_{i,t} = \eta_{i,t}$ and that $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t} \xrightarrow{p} 0$.

[8]Going back to the canonical log-Normal, homoskedastic case, if the conditional distribution of $r_{i,t} + m_t$ is $\mathcal{N}(\tau_i, \sigma_i^2)$, $i = 1, 2, \ldots, N$, then $\gamma_i^2 = \frac{\sigma_i^2}{2}$. Still, $\varepsilon_{i,t} = \eta_{i,t}$ and $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t} \xrightarrow{p} 0$.

to obtain a consistent estimator for e^{γ^2} using (39):

$$\widehat{e^{\gamma^2}} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\widetilde{M}}_t R_{i,t} \right) = \frac{1}{T} \sum_{t=1}^{T} \left(\prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}} \frac{1}{N} \sum_{i=1}^{N} R_{i,t} \right) = \frac{1}{T} \sum_{t=1}^{T} \overline{R}_t^G \overline{R}_t^A,$$

where, in this last step, N must diverge at a rate at least as fast as T, otherwise we would not be able to exchange $\widetilde{M}_t$ by $\widehat{\widetilde{M}}_t$.

We can finally propose a consistent estimator for M_t:

$$\widehat{M}_t = \frac{\widehat{\widetilde{M}}_t}{\widehat{e^{\gamma^2}}} = \frac{\overline{R}_t^G}{\frac{1}{T} \sum_{j=1}^{T} \overline{R}_j^G \overline{R}_j^A},$$

which is a simple function of asset returns. ■

2.2 Discussion

The Asset-Pricing Equation is a non-linear function of the SDF and of returns, which may question the assumption of the existence of a linear factor model relating returns to SDF factors. We show above how to derive an exact log-linear relationship between returns and the SDF, which allows a natural one-factor model linking $r_{i,t}$, $i = 1, 2, \cdots$ and m_t. Under the assumption that no-arbitrage holds instantaneously for all periods of time, large-sample arbitrage portfolios may be constructed using this one-factor model. They remove the common-factor component of returns, but must also remove any common component of the pricing errors $\varepsilon_{i,t}$, since their returns must be non-random in the limit and their limit returns must be zero. Hence, a WLLN applies to the simple average of the cross-sectional errors of the exact log-linear models for returns. It is key to our proof to assume that no-arbitrage holds instantaneously. Indeed, there is no reason why one should dispense with this assumption.

Although our discussion in the previous section points out some skepticism regarding whether or not one should expect $\frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t} \xrightarrow{p} 0$ to hold, since a natural decomposition of $\varepsilon_{i,t}$ entails the factor q_t, we show that, the weights of q_t on this decomposition must all be nil, otherwise we violate no-arbitrage. It is perhaps more instructive to discuss this

issue using the *quasi-structural system* (10), where we try to separately identify m_t and its respective factor loadings. Applying a projection argument to (10), consider the factor model relating $\widetilde{r_{i,t}}$ and $\widetilde{m_t}$, which are demeaned versions of $r_{i,t}$ and m_t respectively:

$$\widetilde{r_{i,t}} = -\beta_i \widetilde{m_t} + \eta_{i,t}, \tag{40}$$

Average (40) across i, taking the probability limit to obtain:

$$\underset{N \to \infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \widetilde{r_{i,t}} = -\left(\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \beta_i \right) \widetilde{m_t} = -\beta \cdot \widetilde{m_t}, \tag{41}$$

where the last equality defines notation. Equation (41) shows that we cannot separately identify β and $\widetilde{m_t}$. We have only one equation: the left-hand-side has observables, but the right-hand-side has two unknowns (β and $\widetilde{m_t}$). Therefore, we need an additional equation (restriction) to uniquely identify $\widetilde{m_t}$. As shown above, no-arbitrage offers $\beta = 1$. This happens either directly, by forming arbitrage portfolios and imposing no arbitrage, or indirectly, by consequence of differentiating the Pricing Equation with respect to m_t, recalling that no arbitrage implies the existence of the Pricing Equation. The unit elasticity is a natural consequence of the Asset Pricing Equation, since the product $M_t R_{i,t}$ must be unity, on average. Hence, increases in M_t must be offset by decreases in $R_{i,t}$ in the same magnitude, on average.

As is well known, an alternative route to separately identify factors and factor loadings is the application of large-sample principal-component and factor analyses; see, e.g., Stock and Watson (2002). However, there is an indeterminacy problem implicit in these methods; see Lawley and Maxwell (1971) for a classic discussion. Denote by $\Sigma_{\mathbf{r}} = \mathbb{E}\left(\widetilde{\mathbf{r}}_t \widetilde{\mathbf{r}}_t' \right)$ the variance-covariance matrix of logged returns, where $\widetilde{\mathbf{r}}_t$ stacks demeaned logged returns $\widetilde{r_{i,t}}$. The first principal component of $\widetilde{\mathbf{r}}_t$ is a linear combination $\alpha' \widetilde{\mathbf{r}}_t$ with maximal variance. As discussed in Dhrymes (1974), since its variance is $\alpha' \Sigma_{\mathbf{r}} \alpha$, the problem has no unique solution – we can make $\alpha' \Sigma_{\mathbf{r}} \alpha$ as large as we want by multiplying α by a constant $\kappa > 1$. Indeed, we are facing a *scale* problem, which is solved by imposing unit norm for α: in

a fixed N setting we have $\alpha'\alpha = 1$, and in a large-sample setting we have $\lim\limits_{N\to\infty} \alpha'\alpha = 1$. Alternatively, the no-arbitrage solution to the indeterminacy problem is to set the mean factor loading in (40) to unity: $\lim\limits_{N\to\infty} \frac{1}{N} \sum\limits_{i=1}^{N} \beta_i = \beta = 1$. Intuitively, this is equivalent to perform a reparameterization of the factor loadings from β_i to β_i/β.

2.3 Properties of the M_t Estimator

The first property of our estimator of M_t, labelled $\widehat{M_t}$, is that it is a function of asset-return data alone. No assumptions whatsoever about preferences have been made so far. Moreover, it is completely non-parametric.

Second, because $\widehat{M_t}$ is a consistent estimator, it is interesting to discuss to what it converges to. Of course, the SDF is a stochastic process: $\{M_t\}$. Since convergence in probability requires a limiting degenerate distribution, our estimator $\widehat{M_t}$ converges to the realization of M at time t. One important issue is that of identification: to what type of SDF $\widehat{M_t}$ converges to? Here, we must distinguish between complete and incomplete markets for securities. In the complete markets case, there is a unique positive SDF pricing all assets, which is identical to the mimicking portfolio M_t^*. Since our estimator is always positive, $\widehat{M_t}$ converges to this unique pricing kernel. Under incomplete markets, no-arbitrage implies that there exists at least one SDF M_t such that $M_t > 0$. There may be more than one. If there is only one positive SDF, then $\widehat{M_t}$ converges to it. If there are more than one, then $\widehat{M_t}$ converges to a convex combination of those positive SDFs. In any case, since all of them have identical pricing properties, the pricing properties of $\widehat{M_t}$ will approach those of all of these positive SDFs.

Third, from a different angle, it is straightforward to verify that our estimator was constructed to obey:

$$\operatorname*{plim}_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \widehat{M_t} R_{i,t} = 1, \tag{42}$$

which is a natural property arising from the moment restrictions entailed by the Asset-Pricing Equation (2), when populational means of the time-series and of the cross-sectional distributions are replaced by sample means. In finite samples, it does not price correctly

any specific asset, but it will price correctly all the assets used in computing it.

2.4   Comparisons with the Literature

As far as we are aware of, early studies in finance and macroeconomics dealing with the SDF did not try to obtain a direct estimate of it as we do: we treated $\{M_t\}$ as a stochastic process and constructed an estimate $\widehat{M_t}$, such that $\widehat{M_t} - M_t \xrightarrow{p} 0$. Conversely, most of the previous literature estimated the SDF indirectly as a function of consumption data from the National Income and Product Accounts (NIPA), using a parametric function to represent preferences; see Hansen and Singleton (1982, 1983, 1984), Brown and Gibbons (1985) and Epstein and Zin (1991). As noted by Rosenberg and Engle (2002), there are several sources of measurement error for NIPA consumption data that can pose a significant problem for this type of estimate. Even if this were not the case, there is always the risk that an incorrect choice of parametric function used to represent preferences will contaminate the final SDF estimate.

Hansen and Jagannathan (1991, 1997) point out that early studies imposed potentially stringent limits on the class of admissible asset-pricing models. They avoid dealing with a direct estimate of the SDF, but note that the SDF has its behavior (and, in particular, its variance) bounded by two restrictions. The first is Pricing Equation (2) and the second is $M_t > 0$. They exploit the fact that it is always possible to project M onto the space of payoffs, which makes it straightforward to express M^*, the mimicking portfolio, only as a function of observable returns:

$$M_{t+1}^* = \boldsymbol{\iota}_N' \left[\mathbb{E}_t \left(\mathbf{R}_{t+1} \mathbf{R}_{t+1}' \right) \right]^{-1} \mathbf{R}_{t+1}, \tag{43}$$

where $\boldsymbol{\iota}_N$ is a $N \times 1$ vector of ones, and $\mathbf{R}_{t+1}$ is a $N \times 1$ vector stacking all asset returns. Although they do not discuss it at any length in their paper, equation (43) shows that it is possible to identify M_{t+1}^* in the Hansen and Jagannathan framework. As in our case, (43) delivers an estimate of the SDF that is solely a function of asset returns and can therefore be used to verify whether preference-parameter values are admissible or not.

If one regards (43) as a means to identify M^*, there are some limitations that must be pointed out. First, it is obvious from (43) that a conditional econometric model is needed to implement an estimate for M^*_{t+1}, since one has to compute the conditional moment $\mathbb{E}_t\left(\mathbf{R}_{t+1}\mathbf{R}'_{t+1}\right)$. To go around this problem, one may resort to the use of the unconditional expectation instead of conditional expectation, leading to $M^*_{t+1} = \boldsymbol{\iota}'_N\left[\mathbb{E}\left(\mathbf{R}_{t+1}\mathbf{R}'_{t+1}\right)\right]^{-1}\mathbf{R}_{t+1}$. Second, as the number of assets increases ($N \to \infty$), the use of (43) will suffer numerical problems in computing an estimate of $\left[\mathbb{E}_t\left(\mathbf{R}_{t+1}\mathbf{R}'_{t+1}\right)\right]^{-1}$. In the limit, the matrix $\mathbb{E}_t\left(\mathbf{R}_{t+1}\mathbf{R}'_{t+1}\right)$ will be of infinite order. Even for finite but large N there will be possible singularities in it, as the correlation between some assets may be very close to unity. Moreover, the number of time periods used in computing $\mathbb{E}_t\left(\mathbf{R}_{t+1}\mathbf{R}'_{t+1}\right)$ or $\mathbb{E}\left(\mathbf{R}_{t+1}\mathbf{R}'_{t+1}\right)$ must be at least as large as N, which is infeasible for most datasets of asset returns.

Our approach is related to the return to aggregate capital. For algebraic convenience, we use the log-utility assumption for preferences – where $M_{t+j} = \beta\frac{c_t}{c_{t+j}}$ – as well as the assumption of no production in the economy in illustrating their similarities. Under the Asset-Pricing Equation, since asset prices are the expected present value of the dividend flows, and since with no production dividends are equal to consumption in every period, the price of the portfolio representing aggregate capital $\bar{p}_t$ is:

$$\bar{p}_t = \mathbb{E}_t\left\{\sum_{i=1}^{\infty}\beta^i\frac{c_t}{c_{t+i}}c_{t+i}\right\} = \frac{\beta}{1-\beta}c_t.$$

Hence, the return on aggregate capital $\overline{R}_{t+1}$ is given by:

$$\overline{R}_{t+1} = \frac{\bar{p}_{t+1} + c_{t+1}}{\bar{p}_t} = \frac{\beta c_{t+1} + (1-\beta)c_{t+1}}{\beta c_t} = \frac{c_{t+1}}{\beta c_t} = \frac{1}{M_{t+1}}, \tag{44}$$

which is the reciprocal of the SDF.

Our approach is also related to several articles that have in common the fact that they reveal a trend in the SDF literature – proposing less restrictive estimates of the SDF compared to the early functions of consumption growth; see, among others, Chapman (1998), Aït-Sahalia and Lo (1998, 2000), Rosenberg and Engle (2002), Garcia, Luger,

and Renault (2003), Sentana (2004), Garcia, Renault, and Semenov (2006), and Sentana, Calzolari, and Fiorentini (2008). In some of these papers a parametric function is still used to represent the SDF, although the latter does not depend on consumption at all or only depends partially on consumption; see Rosenberg and Engle, who project the SDF onto the payoffs of a single traded asset; Aït-Sahalia and Lo (1998, 2000), who rely on equity-index option prices to nonparametrically estimate the projection of the average stochastic discount factor onto equity-return states; Sentana (2004), who uses factor analysis in large asset markets where the conditional mean and covariance matrix of returns are interdependently estimated using the kalman filter; Garcia, Renault and Semenov (2006), who introduce an exogenous reference level related to expected future consumption in addition to the standard consumption term; and Sentana, Calzolari, and Fiorentini (2008), who propose indirect estimators of common and idiosyncratic factors that depend on their past unobserved values in a constrained Kalman-filter setup. Sometimes non-parametric or semi-parametric methods are used, but the SDF is still a function of current or lagged values of consumption; see Chapman, among others, who approximates the pricing kernel using orthonormal Legendre polynomials in state variables that are functions of aggregate consumption.

Although our approach shares with these papers the construction of less stringent SDF estimators, we do not need to characterize preferences or to use consumption data. On the contrary, our approach is entirely based on prices of financial securities. Besides the regularity conditions we assume on the stochastic process of returns, we only assume the absence of arbitrage opportunities (the Asset-Pricing Equation). Compared with the group of papers cited above, this setup is a step forward in relaxing the assumptions needed to recover SDF estimates, while keeping a sensible balance with theory, since we are still using a *structural* basis for SDF estimation.

3 Empirical Applications in Macroeconomics and Finance

3.1 From Asset Prices to Preferences

An important question that can be addressed with our estimator of M_t is how to test and validate specific preference representations. Here we focus on three different preference specifications: the CRRA specification, which has a long tradition in the finance and macroeconomic literatures, the external-habit specification of Abel (1990), and the Kreps and Porteus (1978) specification used in Epstein and Zin (1991), which are respectively:

$$M_{t+1}^{CRRA} = \beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} \tag{45}$$

$$M_{t+1}^{EH} = \beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} \left(\frac{c_t}{c_{t-1}} \right)^{\kappa(\gamma-1)} \tag{46}$$

$$M_{t+1}^{KP} = \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} \right]^{\frac{1-\gamma}{\rho}} \left(\frac{1}{B_t} \right)^{1-\frac{1-\gamma}{\rho}}, \tag{47}$$

where c_t denotes consumption, B_t is the return on the optimal portfolio, β is the discount factor, γ is the relative risk-aversion coefficient, and κ is the time-separation parameter in the habit-formation specification. Notice that M_{t+1}^{EH} is a weighted average of M_{t+1}^{CRRA} and $\left(\frac{c_t}{c_{t-1}} \right)$. In the Kreps-Porteus specification the intertemporal elasticity of substitution in consumption is given by $1/(1-\rho)$ and $\alpha = 1 - \gamma$ determines the agent's behavior towards risk. If we denote $\theta = \frac{1-\gamma}{\rho}$, it is clear that M_{t+1}^{KP} is a weighted average of M_{t+1}^{CRRA} and $\left(\frac{1}{B_t} \right)$, with weights θ and $1 - \theta$, respectively.

For consistent estimates, we can always write:

$$m_{t+1} = \widehat{m_{t+1}} + \eta_{t+1}, \tag{48}$$

where η_{t+1} is the approximation error between m_{t+1} and its estimate $\widehat{m_{t+1}}$.

The properties of η_{t+1} will depend on the properties of M_{t+1} and $R_{i,t+1}$, and, in general, it will be serially dependent and heterogeneous. Using (48) and the expressions in (45), (46) and (47), we arrive at:

$$\widehat{m_{t+1}} = \ln \beta - \gamma \Delta \ln c_{t+1} - \eta_{t+1}^{CRRA}, \tag{49}$$

$$\widehat{m_{t+1}} = \ln \beta - \gamma \Delta \ln c_{t+1} + \kappa \left(\gamma - 1 \right) \Delta \ln c_t - \eta_{t+1}^{EH}, \tag{50}$$

$$\widehat{m_{t+1}} = \theta \ln \beta - \theta \gamma \Delta \ln c_{t+1} - (1-\theta) \ln B_{t+1} - \eta_{t+1}^{KP}, \tag{51}$$

Perhaps the most appealing way of estimating (49), (50) and (51), simultaneously testing for over-identifying restrictions, is to use the generalized method of moments (GMM) proposed by Hansen (1982). Lagged values of returns, consumption and income growth, and also of the logged consumption-to-income ratio can be used as instruments in this case. Since (49) is nested into (50), we can also perform a redundancy test for $\Delta \ln c_t$ in (49). The same applies regarding (49) and (51), since the latter collapses to the former when $\ln B_{t+1}$ is redundant.

In our first empirical exercise, we apply our techniques to returns available to the average U.S. investor, who has increasingly become more interested in global assets over time. Real returns were computed using the consumer price index in the U.S. Our data base covers U.S.$ real returns on G7-country stock indices and short-term government bonds, where exchange-rate data were used to transform returns denominated in foreign currency into U.S.$. In addition to G7 returns on stocks and bonds, we also use U.S.$ real returns on gold, U.S. real estate, bonds on AAA U.S. corporations, and on the SP 500. The U.S. government bond is chosen to be the 90-day T-Bill, considered by many to be a "riskless asset." All data were extracted from the DRI database, with the exception of real returns on real-estate trusts, which are computed by the National Association of Real-Estate Investment Trusts in the U.S.[9] Our sample period starts in 1972:1 and ends in 2000:4. Overall, we averaged the real U.S.$ returns on these 18 portfolios or assets[10],

---

[9]Data on the return on real estate are measured using the return of all publicly traded REITs – Real-Estate Investment Trusts.

[10]The complete list of the 18 portfolio- or asset-returns, all measured in U.S.$ real terms, is: returns on the NYSE, Canadian Stock market, French Stock market, West Germany Stock market, Italian Stock

which are, in turn, a function of thousands of assets. These are predominantly U.S. based, but we also cover a wide spectrum of investment opportunities across the globe. This is an important element of our choice of assets, since diversification allows reducing the degree of correlation of returns across assets, whereas too much correlation may generate no convergence in probability for sample means.

In estimating equations (49) and (50), we must use additional series. Real per-capita consumption growth was computed using private consumption of non-durable goods and services in constant U.S.$. We also used real per-capita GNP as a measure of income – an instrument in running some of these regressions. Consumption and income series were seasonally adjusted.

Figure 1 below shows our estimator of the SDF – $\widehat{M_t}$ – for the period 1972:1 to 2000:4. It is close to unity most of the time and bounded by the interval [0.85, 1.15]. The sample mean of $\widehat{M_t}$ is 0.9927, implying an annual discount factor of 0.9711, or an annual discount rate of 2.97%, a very reasonable estimate.
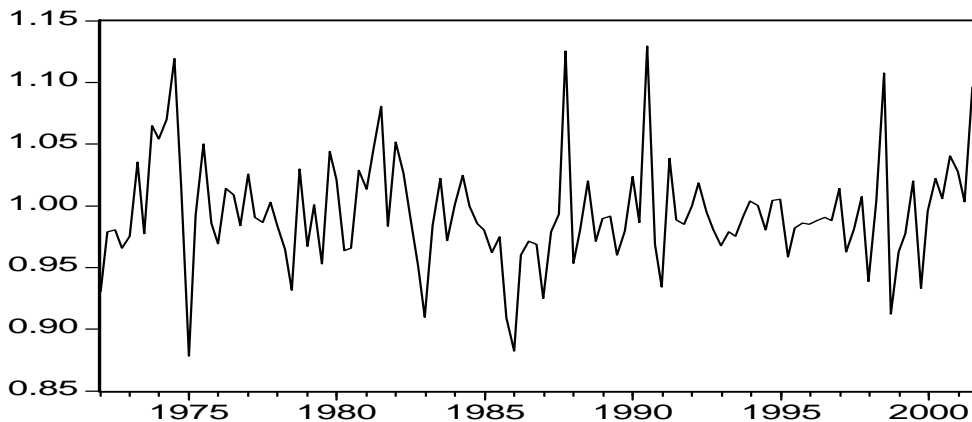

Figure 1: Stochastic Discount Factor

Tables 1, 2, and 3 present GMM estimation of equations (49), (50) and (51), re-

market, Japanese Stock market, U.K. Stock market, 90-day T-Bill, Short-Term Canadian Government Bond, Short-Term French Government Bond, Short-Term West Germany Government Bond, Short-Term Italian Government Bond, Short-Term Japanese Government Bond, Short-Term U.K. Government Bond. As well as on the return of all publicly traded REITs – Real-Estate Investment Trusts in the U.S., on Bonds of AAA U.S. Corporations, Gold, and on the SP 500.

spectively. We used as a basic instrument list two lags of all real returns employed in computing $\widehat{M_t}$, two lags of $\ln\left(\frac{c_t}{c_{t-1}}\right)$, two lags of $\ln\left(\frac{y_t}{y_{t-1}}\right)$, and one lag of $\ln\left(\frac{c_t}{y_t}\right)$. This basic list was altered in order to verify the robustness of empirical results. We also include OLS estimates to serve as benchmarks in all three tables.

Table 1 Power-Utility Function Estimates $\widehat{m_t} = \ln\beta - \gamma\Delta\ln c_t - \eta_t^{CRRA}$			
Instrument Set	β (SE)	γ (SE)	OIR Test (P-Value)
OLS Estimate	1.002 (0.006)	1.979 (0.884)	–
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N.$	0.999 (0.003)	1.125 (0.517)	(0.9953)
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N,$ $\Delta\ln c_{t-1}, \Delta\ln c_{t-2}.$	1.001 (0.003)	1.370 (0.511)	(0.9964)
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N,$ $\Delta\ln y_{t-1}, \Delta\ln y_{t-2}.$	1.000 (0.003)	1.189 (0.523)	(0.9958)
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N, \Delta\ln c_{t-1},$ $\Delta\ln c_{t-2}, \Delta\ln y_{t-1}, \Delta\ln y_{t-2}, \ln\frac{c_{t-1}}{y_{t-1}}.$	0.999 (0.003)	1.204 (0.514)	(0.9985)

Notes: (1) Except when noted, all estimates are obtained using the generalized method of moments (GMM) of Hansen (1982), with robust Newey and West (1987) estimates for the variance-covariance matrix of estimated parameters. (2) OIR Test denotes the over-identifying restrictions test discussed in Hansen (1982). (3) A constant is included as instrument in GMM estimation.

Table 1 reports results obtained using a power-utility specification for preferences. The first thing to notice is that there is no evidence of rejection in over-identifying restrictions tests in any GMM regression we have run. Moreover, all of them showed sensible estimates for the discount factor and the risk-aversion coefficient: $\widehat{\beta} \in [0.999, 1.001]$, where in all cases the discount factor is not statistically different from unity and $\widehat{\gamma} \in [1.125, 1.370]$, where in all cases the relative risk-aversion coefficient is likewise not statistically different from unity. Our preferred regression is the last one in Table 1, where all instruments are used in estimation. There, $\widehat{\beta} = 0.999$ and $\widehat{\gamma} = 1.204$. These numbers are close to what could be expected *a priori* when power utility is considered; see the discussion in Mehra and Prescott (1985). They are in line with several panel-data estimates of the relative risk-aversion coefficient, such as Runkle (1991), Attanasio and Weber (1985) and

Blundell, Browning and Meghir (1994).

Our estimates $\widehat{\beta}$ and $\widehat{\gamma}$ in Table 1 are somewhat different from early estimates of Hansen and Singleton (1982, 1984). As is well known, the equity-premium puzzle emerged as a result of rejecting the over-identifying restrictions implied by the complete system involving real returns on equity and on the T-Bill: Hansen and Singleton's estimates of γ are between 0.09 and 0.16, with a median of 0.14, all statistically insignificant in testing. All of our estimates are statistically significant, and their median estimate is 1.20 – almost ten times higher.

Table 2				
External-Habit Utility-Function Estimates				
$\widehat{m_t} = \ln\beta - \gamma\Delta\ln c_t + \kappa(\gamma-1)\Delta\ln c_{t-1} - \eta_t^{EH}$				
Instrument Set	β (SE)	γ (SE)	κ (SE)	OIR Test (P-Value)
OLS Estimate	1.002 (0.006)	1.975 (0.972)	-0.008 (0.997)	–
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N.$	1.005 (0.003)	1.263 (0.618)	-2.847 (8.333)	(0.9911)
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N,$ $\Delta\ln c_{t-1}, \Delta\ln c_{t-2}.$	0.9954 (0.003)	1.308 (0.562)	1.997 (3.272)	(0.9954)
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N,$ $\Delta\ln y_{t-1}, \Delta\ln y_{t-2}.$	0.987 (0.003)	1.592 (0.688)	3.588 (3.742)	(0.9951)
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N, \Delta\ln c_{t-1},$ $\Delta\ln c_{t-2}, \Delta\ln y_{t-1}, \Delta\ln y_{t-2}, \ln\frac{c_{t-1}}{y_{t-1}}.$	0.987 (0.002)	1.161 (0.621)	8.834 (32.769)	(0.9980)

Notes: Same as Table 1.

Table 2 reports results obtained when (external) habit formation is considered in preferences. Results are very similar to those obtained with power utility. A slight difference is the fact that, with one exception, all estimates of the discount factor are smaller than unity. We cannot reject time-separation for all regressions we have run – κ is statistically zero in testing everywhere. In this case, the external-habit specification collapses to that of power-utility, which should be preferred as a more parsimonious model.

Table 3				
Kreps–Porteus Utility-Function Estimates				
$\widehat{m_t} = \theta \ln \beta - \theta\gamma\Delta \ln c_t - (1 - \theta)\ln B_t - \eta_t^{KP}$				
Instrument Set	β	γ	θ	OIR Test
	(SE)	(SE)	(SE)	(P-Value)
OLS Estimate	1.007	3.141	0.831	–
	(0.006)	(0.886)	(0.022)	
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N.$	1.001	1.343	0.933	(0.9963)
	(0.004)	(0.723)	(0.014)	
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N,$	1.003	1.360	0.922	(0.9980)
$\Delta \ln c_{t-1}, \Delta \ln c_{t-2}.$	(0.004)	(0.768)	(0.012)	
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N,$	1.000	0.926	0.927	(0.9969)
$\Delta \ln y_{t-1}, \Delta \ln y_{t-2}.$	(0.004)	(0.756)	(0.013)	
$r_{i,t-1}, r_{i,t-2}, \forall i = 1, 2, \cdots N, \Delta \ln c_{t-1},$	0.997	0.362	0.901	(0.9996)
$\Delta \ln c_{t-2}, \Delta \ln y_{t-1}, \Delta \ln y_{t-2}, \ln \frac{c_{t-1}}{y_{t-1}}.$	(0.004)	(0.761)	(0.012)	

Notes: Same as Table 1.

Results using the Kreps-Porteus specification are reported in Table 3. To implement its estimation a first step is to find a proxy to the optimal portfolio. We followed Epstein and Zin (1991) in choosing the NYSE for that role, although we are aware of the limitations they raise for this choice. With that caveat, we find that the optimal portfolio term has a coefficient that is close to zero in value (θ close to unity), although $(1 - \theta)$ is not statically zero in any regressions we have run. If it were, then the Kreps-Porteus would collapse to the power-utility specification. The estimates of the relative risk-aversion coefficient are not very similar across regressions, ranging from 0.362 to 1.360. Moreover, they are not statistically different from zero at the 5% significance level, which differs from previous estimates in Tables 1 and 2. There is no evidence of rejection in over-identifying restrictions tests in any GMM regression we have run, which is in sharp contrast to the early results of Epstein and Zin using this same specification.

Since the Kreps-Porteus encompasses the power utility specification, the former should be preferred to the latter in principle because $(1 - \theta)$ is not statistically zero. A reason against it is the limitation in choosing a proxy for the optimal portfolio. Therefore, the picture that emerges from the analysis of Tables 1, 2 and 3 is that both the power-utility and the Kreps-Porteus specifications fit the CCAPM reasonably well when our estimator of the SDF is employed in estimation. Since κ is statistically zero, we find little evidence

in favor of external habit formation using our data.

3.2 Out-of-Sample Asset-Pricing Forecasting Exercise

Next, we present the results of an asset-pricing out-of-sample forecasting exercise in the panel-data dimension. In constructing our estimator of the SDF, we try to approximate the asymptotic environment with monthly U.S. time-series return data from 1980:1 through 2007:12 ($T = 336$ observations), collected for $N = 16,193$ assets, grouped in the following four categories: mutual funds ($7,932$), stocks ($6,009$), real estate (383), and government bonds ($1,869$). After computing $\widehat{M_t}$, we price individual return data not used in constructing it, measuring the distance between forecast prices and 1 using the δ pricing-error measure proposed in Hansen and Jagannathan (1997).

All return data used in this exercise come from CRSP. Mutual-Fund return data comes from the CRSP Mutual Fund Database, which reports open-ended mutual-fund returns using survivor-bias-free data. Bias can arise, for example, when a older fund splits into other share classes, each new share class being permitted to inherit the entire return/performance history of the older fund. Stock return data comes from the CRSP U.S. Stock and CRSP U.S. Indices, which collects returns from NYSE, AMEX, NASDAQ, and, more recently, NYSE Arca. Real-Estate return data comes from the CRSP/Ziman Real Estate Data Series. It collects return data on real-estate investment trusts (REITs) that have traded on the NYSE, AMEX and NASDAQ exchanges. Finally, government-bond return data comes from CRSP Monthly Treasury U.S. Database, which collects monthly returns of U.S. Treasury bonds with different maturities.

The first step to perform our exercise is computing $\widehat{M_t}$. Since we do not have a random sample of returns, we decided to work with each of the four categories above, weighting them by their respective importance in the median U.S. household portfolio. For each of the four asset categories (mutual funds, stocks, real estate, and government bonds) we computed the geometric average of the reciprocal of all asset returns and the arithmetic average of all asset returns. Based on the "Wealth and Asset Ownership" tables of 2004, provided by the U.S. Census Bureau, we decided to weight the returns in each of the

four categories as follows: Mutual Funds (10%), Stocks (10%), Real Estate (60%), and Government Bonds (20%)[11]. They are a close approximation of the median (and also the mean) value of assets owned by U.S. households in these four categories. Local changes in these weights (from 5 up to 20 percentage points for individual categories) produce no virtual change on the results of our exercise. Our final estimate $\widehat{M_t}$ results from weighting geometric and arithmetic averages of returns in each of these four categories.

Once we obtain $\widehat{M_t}$, we forecast a group of returns not included in computing it for all the 336 observations in the time-series dimension, comparing our results with unity. Under the law of one price this exercise is similar in spirit to the one in Hansen and Jagannathan (1997). Our forecasting exercise is performed using nominal returns either in constructing the SDF or in out-of-sample evaluation of returns. Obviously, the product $M_t R_{i,t}$ is invariant to price inflation as long as the same price index is used in deflating M_t and $R_{i,t}$.

Our estimate of M_t has a nominal mean of 0.9922 in a monthly basis, which amounts to 0.9106 in a yearly basis. In comparison, average yearly CPI inflation for the same period is 3.85%. The plot of $\widehat{M_t}$ follows below in Figure 2.

---

[11]These tables can be downloaded from http://www.census.gov/hhes/www/wealth/2004_tables.html. These weights we propose using come from Table 1, which has the "Median Value of Assets for Households, by Type of Asset Owned and Selected Characteristics."
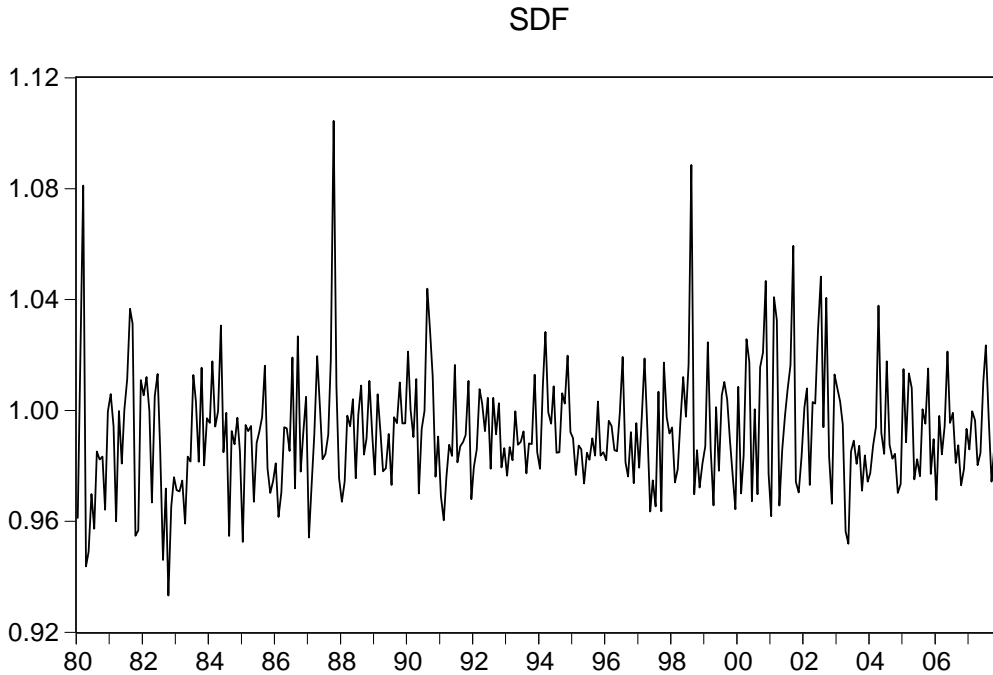
Figure 2: Stochastic Discount Factor

We want our forecasting exercise to be out of sample. In choosing the group of assets which will have their returns priced, we require that they have not been included in computing $\widehat{M_t}$. To cover a wide spectrum of assets to be priced, we chose to work with stocks, divided in 10 categories of capitalization, according to the CRSP Stock File Capitalization Decile Indices. Their returns are calculated for each of the Stock File Indices market groups. All securities, excluding ADRs on a given exchange or combination of exchanges, are ranked according to capitalization and then divided into ten equal parts, each rebalancing every year using the security market capitalization at the end of the previous year to rank securities. The largest securities are placed in portfolio 10 and the smallest in portfolio 1. Value-Weighted Index Returns including all dividends are calculated on each of the ten portfolios. Because of the value-weighted character of these portfolios, and the fact that they are rebalanced every year, their returns cannot be written as a fixed-weight linear combination of the returns used in computing $\widehat{M_t}$ – therefore do not lie in the space

of returns used in computing $\widehat{M_t}$. This makes our forecasting exercise out-of-sample in the panel-data dimension.

We evaluate our estimator $\widehat{M_t}$ in terms of its ability to price the returns of these ten portfolios divided into capitalization categories. We use the distance measure δ proposed in Hansen and Jagannathan, which represents the smallest adjustment required in our estimator to bring it to an admissible SDF. Results are presented in Table 4.

Table 4 Out-of-Sample Asset-Pricing Forecast Evaluation SDF Proxy: $\widehat{M_t}$		
Returns of	Capitalization Portfolios 1-10	Capitalization Portfolios 1-5	Capitalization Portfolios 6-10
Distance Measure	$\widehat{\delta}$	$\widehat{\delta}$	$\widehat{\delta}$
	(Robust SE)	(Robust SE)	(Robust SE)
	0.1493	0.0912	0.0677
	(0.0483)	(0.0442)	(0.0589)

Notes: The capitalization portfolios tested in the first three columns come form the CRSP Stock File Capitalization Decile Indices. These are divided into 10 capitalization groups, by decile of capitalization. The largest securities are placed in portfolio 10 and the smallest in portfolio 1. Estimates of the Hansen and Jagannathan distance δ and its respective robust standard error are computed using the MATLAB code made available by Mike Cliff[12]. Robust SE are computed using the procedure proposed by Newey and West (1987).

When pricing all 10 capitalization portfolios, the performance of our estimator comes short of expected. The distance δ is significant at the usual levels of significance. In trying to understand the reasons for rejecting admissibility, we divided the 10 portfolios into two groups: "smaller caps," with deciles of capitalization from 1 to 5, and "larger caps," with deciles of capitalization from 6 to 10. In pricing the smaller caps portfolios, δ is still significant at the usual levels, although only marginally so. However, when the larger caps are priced, our estimator of the SDF is admissible and δ is far from significant; see also the cross-plot of the required adjustment vs. the SDF value depicted in Figure 3.

Finally, the evidence in Table 4 leads to the conclusion that our initial rejection was due to misspricing smaller-cap stocks. We do not see this result as a serious drawback for our estimator. As is well known, there is a much greater volatility in terms of entry and

---

[12]The code can now be downloaded from:
http://sites.google.com/site/mcliffweb/programs

exit of smaller firms into the marketplace, whose historical positive returns are always recorded, but some negative results are not recorded due to bankruptcy. Hence, one may expect some bias in using smaller-cap firms historical returns in asset-pricing tests, which may be the case here when using capitalization deciles 1 to 5.
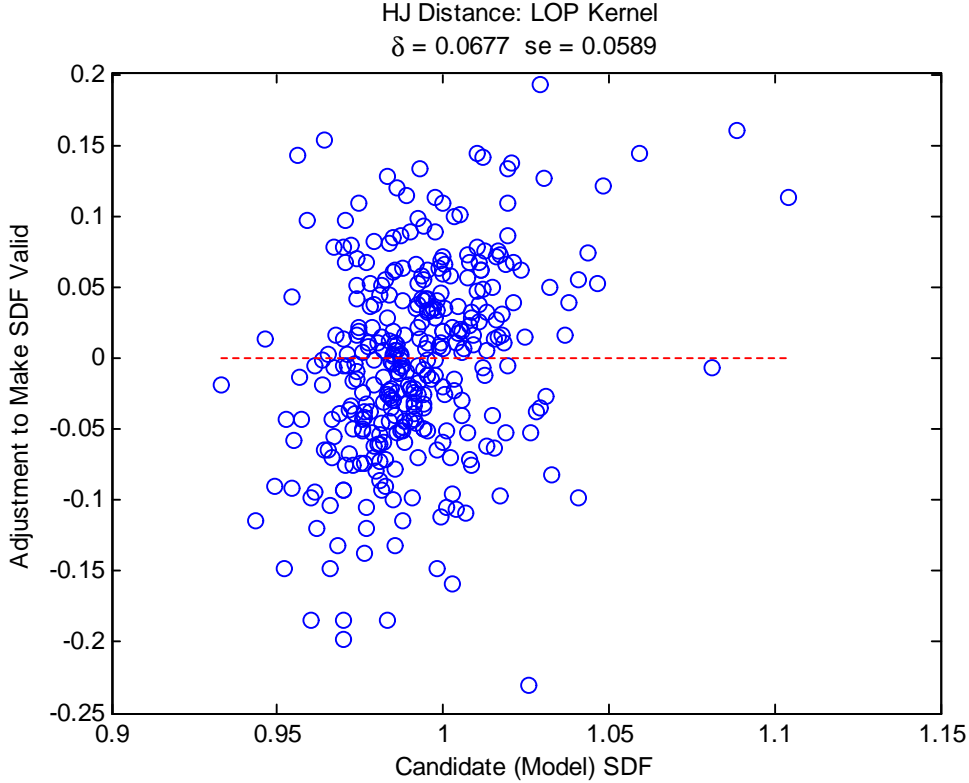

Figure 3: Admissibility Adjustment vs. SDF Value

4   Conclusions

In this paper, we propose a novel consistent estimator for the stochastic discount factor (SDF), or pricing kernel, that exploits both the time-series and the cross-sectional dimensions of asset prices. We treat the SDF as a random process that can be estimated consistently as the number of time periods and assets in the economy grow without bounds. To construct our estimator, we basically rely on standard regularity conditions on the stochastic processes of asset returns and on the absence of arbitrage opportunities in

asset pricing. Our SDF estimator depends exclusively on appropriate averages of asset returns, which makes its computation a simple and direct exercise. Because it does not depend on any assumptions on preferences, or on consumption data, we are able to use our SDF estimator to test directly different preference specifications which are commonly used in finance and in macroeconomics. We also use it in an out-of-sample asset-pricing forecasting exercise.

A key feature of our approach is that it combines a general Taylor Expansion of the Pricing Equation with standard panel-data asymptotic theory to derive a novel consistent estimator for the SDF. In this context, we show that the econometric identification of the SDF only requires using the "common-feature property" of the logarithm of the SDF. We have followed two literature trends here: first, in financial econometrics, recent work avoids imposing stringent functional-form restrictions on preferences prior to estimation of the SDF; see Chapman (1998), Aït-Sahalia and Lo (1998, 2000), Rosenberg and Engle (2002), Garcia, Luger, and Renault (2003), Sentana (2004), Garcia, Renault, and Semenov (2006), and Sentana, Calzolari, and Fiorentini (2008); second, in macroeconomics, early rejections of the optimal behavior for consumption using time-series data found by Hall(1978), Flavin(1981, 1993), Hansen and Singleton(1982, 1983, 1984), Mehra and Prescott(1985), Campbell (1987), Campbell and Deaton(1989), and Epstein and Zin(1991) were overruled by subsequent results using panel data by Runkle (1991), Blundell, Browning, and Meghir (1994), Attanasio and Browning (1995), and Attanasio and Weber (1995), among others.

The techniques discussed in this paper were applied to two relevant issues in macro-economics and finance: the first asks what type of parametric preference-representation could be valid using our SDF estimator, and the second asks whether or not our SDF estimator can price returns in an out-of-sample forecasting exercise. In the first application, we used quarterly data of U.S.\$ real returns from 1972:1 to 2000:4 representing investment opportunities available to the average U.S. investor. They cover thousands of assets worldwide, but are predominantly U.S.-based. Our SDF estimator – $\widehat{M_t}$ – is close to unity most of the time and bounded by the interval $[0.85, 1.15]$, with an equivalent average annual discount factor of 0.9711, or an annual discount rate of 2.97%. When we examined

the appropriateness of different functional forms to represent preferences, we concluded that standard preference representations used in finance and in macroeconomics cannot be rejected by the data. Moreover, estimates of the relative risk-aversion coefficient are close to what can be expected *a priori* – between 1 and 2, statistically significant and not different from unity in statistical tests. In the second application, we tried to approximate the asymptotic environment by working with monthly U.S. time-series return data from 1980:1 through 2007:12 ($T = 336$ observations), which were collected for a total of $N = 16,193$ assets. We showed that our SDF proxy can price reasonably well the returns of stocks with a higher capitalization level, whereas it shows some difficulty in pricing stocks with a lower level of capitalization. Because there is more volatility in terms of entry and exit of smaller firms into the marketplace, which may generate a bias in historical returns for "lower cap" returns, rejection in this case may not be too problematic.

References

[1] Abel, A., 1990, "Asset Prices under Habit Formation and Catching Up with the Joneses,"*American Economic Review Papers and Proceedings,* 80, 38-42.

[2] Athanasopoulos, G., Guillen, O.T.C., Issler, J.V., and Vahid, F. (2011), "Model Selection, Estimation and Forecasting in VAR Models with Short-run and Long-run Restrictions," forthcoming, *Journal of Econometrics.*

[3] Attanasio, O., and M. Browning, 1995, "Consumption over the Life Cycle and over the Business Cycle," *American Economic Review*, vol. 85(5), pp. 1118-1137.

[4] Attanasio, O., and G. Weber, 1995, "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey," *Journal of Political Economy*, vol. 103(6), pp. 1121-1157.

[5] Bai, J. and Serena Ng, 2004, "Evaluating Latent and Observed Factors in Macroeconomics and Finance," Working Paper: University of Michigan.

[6] Blundell, Richard, Martin Browning, and Costas Meghir, 1994, "Consumer Demand and the Life-Cycle Allocation of Household Expenditures," *Review of Economic Studies,* Vol. 61, No. 1, pp. 57-80 .

[7] Bollerslev, T., Engle, R.F. and Wooldridge, J., 1988, "A Capital Asset Pricing Model with Time Varying Covariances", *Journal of Political Economy*, vol. 96, pp.116-131.

[8] Brown, D.P. and Gibbons, M.R. (1985). "A Simple Econometric Approach for Utility-based Asset Pricing Models," *Journal of Finance*, 40(2): 359-81.

[9] Campbell, J.Y. (1987), "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis," *Econometrica*, vol. 55(6), pp. 1249-73.

[10] Campbell, J.Y. and Deaton, A. (1989), "Why is Consumption so Smooth?". *Review of Economic Studies*, 56:357-374.

[11] Campbell, J.Y. and Mankiw, G. (1989), "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence", *NBER Macroeconomics Annual*, 4:185-216.

[12] Chapman, David A. (1997), "Approximating the asset pricing kernel", *Journal of Finance*, Vol. 52, pp. 1383-1410.

[13] Cochrane, J.H. (2002). *Asset Pricing,* Princeton and Oxford: Princeton University Press. Revised Edition.

[14] Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, pp. 987-1006.

[15] Engle, R. F., Issler, J. V., 1995, "Estimating common sectoral cycles," *Journal of Monetary Economics*, vol. 35, 83–113.

[16] Engle, R.F. and Kozicki, S. (1993). "Testing for Common Features", *Journal of Business and Economic Statistics*, 11(4): 369-80.

[17] Engle, R. and Marcucci, J. (2006), "A long-run Pure Variance Common Features model for the common volatilities of the Dow Jones," *Journal of Econometrics*, Volume 132(1), pp. 7-42.

[18] Epstein, L. and S. Zin, 1991, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Investigation," *Journal of Political Economy*, 99, 263-286.

[19] Fama, E.F. and French, K.R. (1992)." The Cross-Section of Expected Stock Returns". *Journal of Finance*, 47(2): 427-65.

[20] Fama, E.F. and French, K.R. (1993). "Common Risk Factors in the Returns on Stock and Bonds". *Journal of Financial Economics*, 33(1): 3-56.

[21] Flavin, M. (1981), "The Adjustment of Consumption to Changing Expectations About Future Income". *Journal of Political Economy*, 89(5):974-1009.

[22] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), "The Generalized Dynamic Factor Model: Identification and Estimation", *Review of Economics and Statistics*, 2000, vol. 82, issue 4, pp. 540-554.

[23] Garcia, René, and Eric Renault, 2001, "Latent Variables Models for Stochastic Discount Factors", in *Handbook of Mathematical Finance*, E. Jouini, J. Cvitanic and M. Musiela (Eds.), Cambridge University Press, pp. 159-184.

[24] Garcia, René, Richard Luger, and Eric Renault, 2003, "Empirical Assessment of an Intertemporal Option Pricing Model with Latent Variables", *Journal of Econometrics*, 116, pp. 49-83.

[25] Garcia, René, Eric Renault, and Andrei Semenov, 2006, "Disentangling Risk Aversion and Intertemporal Substitution Through a Reference Utility Level." *Finance Research Letters* 3:181, 193.

[26] Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," Chapter 19 in D.J. Aigner and A.S. Goldberger (eds.) *Latent Variables in Socio-Economic Models*, Amsterdam: North Holland.

[27] Hall, R. (1978), "Stochastic Implications of the Life Cycle Permanent Income Hypothesis: Theory and Evidence". *Journal of Political Economy*, 86(6), 971-987.

[28] Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[29] Hansen, L. P. and Jagannathan, R. (1991), "Implications of Security Market Data for Models of Dynamic Economies", *Journal of Political Economy*, 99(2):225-262.

[30] Hansen, L. P. and Jagannathan, R. (1997), "Assessing Specification Errors in Stochastic Discount Factor Models." *Journal of Finance*, 1997, 52(2), pp. 557-590.

[31] Hansen, Lars Peter and Eric Renault (2009), "Pricing Kernels and Stochastic Discount Factors," forthcoming in the *Encyclopedia of Quantitative Finance*.

[32] Hansen, L.P. and Scott F. Richard, 1987, "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models," *Econometrica*, 55:587-613.

[33] Hansen, Lars Peter and José A. Scheinkman (2009), "Long-Term Risk: An Operator Approach," *Econometrica*, 77(1), pp. 177 - 234.

[34] Hansen, L. and K. Singleton, 1982,"Generalized Instrumental Variables Estimation of Nonlinear Expectations Models," *Econometrica*, 50(5), pp. 1269-1286.

[35] Hansen, L. and Singleton, K. (1983), "Stochastic Consumption, Risk Aversion and the Temporal Behavior of Asset Returns". *Journal of Political Economy*, 91(2):249-265.

[36] Hansen, L. and K. Singleton, 1984, Erratum of the article "Generalized Instrumental Variables Estimation of Nonlinear Expectations Models," *Econometrica*, 52(1), pp. 267-268.

[37] Harrison, J.M. and Kreps, D.M. (1979), "Martingales and arbitrage in multiperiod securities markets". *Journal of Economic Theory*, 20, 381–405.

[38] Hecq, A., Palm, F.C., Urbain, J.-P., 2006, "Testing for common cyclical features in VAR models with cointegration," *Journal of Econometrics*, Volume 132(1), pp. 117-141.

[39] Issler, J. V., Vahid, F., 2001, "Common cycles and the importance of transitory shocks to macroeconomic aggregates," *Journal of Monetary Economics*, vol. 47, 449–475.

[40] Issler, J.V. and Vahid, F. (2006), "The Missing Link: Using the NBER Recession Indicator to Construct Coincident and Leading Indices of Economic Activity," *Journal of Econometrics*, vol. 132(1), pp. 281-303.

[41] Issler, J.V. and Lima, L.R. (2009), "A Panel-Data Approach to Economic Forecasting: The Bias-Corrected Average Forecast," *Journal of Econometrics*, Vol. 152(2), pp. 153-164.

[42] Kreps, D. and E. Porteus, 1978, "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica,* 46, 185-200.

[43] Lawley, D. N. and Maxwell, A. E. (1971), "*Factor Analysis as a Statistical Method.*" Second edition. London: Butterworths.

[44] Lettau, M. and Ludvigson, S. (2001). "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109(6): 1238-87.

[45] Mehra, R., and Prescott, E., 1985, "The Equity Premium: A Puzzle," *Journal of Monetary Economics,* 15, 145-161.

[46] Mulligan, C. (2002), "Capital, Interest, and Aggregate Intertemporal Substitution," Working Paper # w9373: National Bureau of Economic Research.

[47] Newey, W. K., West, K. D., 1987, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, vol. 55, 703–708.

[48] Rosenberg, J. and Engle, R.F. (2002), "Empirical Pricing Kernels", *Journal of Financial Economics*, Vol. 64, 3, pp. 341-372.

[49] Ross, S.A. (1976), "The arbitrage pricing of capital asset pricing", *Journal of Economic Theory*, 13, pp. 341-360.

[50] Runkle, D. (1991), "Liquidity Constraints and the Permanent Income Hypothesis: Evidence from Panel Data". *Journal of Monetary Economics*, 27(1):73-98.

[51] Sentana, Enrique, 2004, "Factor representing portfolios in large asset markets," *Journal of Econometrics,* Volume 119, Issue 2, pp. 257-289.

[52] Sentana, Enrique, Giorgio Calzolari, Gabriele Fiorentini, 2008, "Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks," *Journal of Econometrics*, Volume 146, Issue 1, pp. 10-25.

[53] Sentana, Enrique, Gabriele Fiorentini, 2001, Identification, estimation and testing of conditionally heteroskedastic factor models, *Journal of Econometrics*, Volume 102, Issue 2, pp. 143-164.

[54] Stock, J. and Watson, M. (1989) "New Indexes of Leading and Coincident Economic Indicators", *NBER Macroeconomics Annual*, pp. 351-95.

[55] Stock, J. and Watson, M. (1993), Eds., "*New Research on Business Cycles, Indicators and Forecasting*" Chicago: University of Chicago Press, for NBER.

[56] Stock, J. and Watson, M. (2002) "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, Vol. 97, pp. 1167-1179.

[57] Vahid, F. and Engle, R. F., 1993, "Common trends and common cycles," *Journal of Applied Econometrics*, vol. 8, 341–360.

[58] Vahid, F., Engle, R. F., 1997, "Codependent cycles," *Journal of Econometrics*, vol. 80, 199–221.

[59] Vahid, F., Issler, J. V., 2002, "The importance of common cyclical features in VAR analysis: A Monte Carlo study," *Journal of Econometrics*, 109, 341–363.