A No-Arbitrage Approach to Asset Pricing using Panel Data Asymptotics*

Fabio Araujo
Brazilian Central Bank
email: fabio.araujo@bcb.gov.br

João Victor Issler
Brazilian School of Economics and Finance FGV EPGE
Getulio Vargas Foundation
email: joao.issler@fgv.br†

February, 2022

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

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

Abstract

We propose a no-arbitrage framework related to stochastic discount factors (or pricing kernels) that takes seriously the consequences of no-arbitrage in asset pricing. First, we derive a no-arbitrage one-factor model for the logarithm of asset returns, where the single factor is the logarithm of a valid stochastic discount factor, containing all the pervasive elements of (log) asset returns. Second, based on this one-factor model, we derive a consistent estimator of a valid SDF in a panel-data framework, when the number of assets and of time periods increase without bounds. Identification of a valid SDF is based on economic theory – no-arbitrage, the asset-pricing equation. The asymptotic character of this no-arbitrage SDF estimator is opposed to standard small-sample alternatives where it is hard to interpret empirical results since these often change when different groups of assets are used in estimation. In theory, asymptotic estimates are immune to this problem.

Based on a consistent estimator for a valid SDF, we first investigate which type of utility function best fits U.S. data among popular preference specifications in the literature: the constant-relative-risk-aversion (CRRA) coefficient utility function; the external habit utility function; and the Kreps-Porteus specification. Second, using estimation results, we present a no-arbitrage simulation study assessing how close our consistent SDF estimator is to actual SDF for medium-size panel-data samples. Finally, we estimate a multi-effect linear regression model that allows for parameter heterogeneity in the intercept and in its slope that is consistent with our derived one-factor model. Using regression results we assess how well this heterogeneous one-factor model fits the cross-section and time-series data of assets returns.

*This paper supersedes Araujo, Issler and Fernandes (2005) and Araujo and Issler (2011), both cited in our references below. We thank the comments and suggestions given by Jushan Bai, Marco Bonomo, Luis Braido, Xiaohong Chen, Valentina Corradi, Carlos E. Costa, Daniel Ferreira, Luiz Renato Lima, Oliver Linton, Humberto Moreira, Marcelo Moreira, Walter Novaes, and Farshid Vahid. Special thanks are due to Caio Almeida, Robert Engle, Marcelo Fernandes, René Garcia, Antonio Galvão, Lars Hansen, Cristine Campos de Xavier Pinto and José Alexandre Scheinkman. We also thank José Gil Ferreira Vieira Filho, Rafael Burjack, and Guilherme Kira for excellent research assistance. The usual disclaimer applies. Moreover, the views in this paper do not reflect those of the Banco Central do Brasil. Fabio Araujo and João Victor Issler gratefully acknowledge support given by CNPq-Brazil and Pronex. Issler also thanks INCT and FAPERJ for financial support. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

†Corresponding author: Graduate School of Economics, Getulio Vargas Foundation, Praia de Botafogo 190 s. 1100, Rio de Janeiro, RJ 22253-900, Brazil.

1 Introduction

In this paper, we propose a no-arbitrage framework related to stochastic discount factors (or pricing kernels) that takes seriously the consequences of the Asset-Pricing Equation established by Harrison and Kreps (1979), Hansen and Richard (1987), and Hansen and Jagannathan (1991, 1997). There, conditional on current information, asset prices today are a function of their expected future discounted payoffs using a stochastic discount factor (SDF).

If the Asset-Pricing Equation is valid for all assets at all times, it can serve as a basis to construct a whole new framework with the following ingredients. First, based on a no-arbitrage identification assumption, we derive a one-factor model for the logarithm of asset returns, where the single factor is the logarithm of a valid stochastic discount factor. Following the SDF literature, we interpret the factor as containing all the pervasive elements of (log) asset returns, allowing a pervasive-idiosyncratic decomposition of these returns, where the intercept and slope parameters contain heterogeneous responses (heterogeneous level and factor loadings). Second, based on this one-factor model, we derive a consistent estimator of a valid SDF in a panel-data framework, where the number of assets and of time periods increase without bounds. The asymptotic character of this SDF estimator differs from SDFs estimated using small samples, where it is hard to interpret empirical results since these often change when different groups of assets are employed in estimation. From a theoretical perspective, asymptotic estimates are immune to this problem, which increases their potential application in empirical studies. This is especially true in the *big data* era.

The core idea behind our consistent SDF estimator can be explained employing a simplified version of the asset-pricing equation as in Hansen and Singleton (1983): given no-arbitrage, and using a joint log-Normal specification for returns and a valid stochastic discount factor, we note that there exists a log-linear factor model for the (log of) returns, where the factor is a valid (log) SDF, which only varies across time, whereas returns and other elements in the factor model vary across time and assets. Then, we exploit the cross-sectional variation of returns to propose an asymptotically biased estimate of a valid SDF.

To remove the asymptotic bias, we rely on economic theory – once again employing the Asset-Pricing Equation – exploiting its time-series variation to construct a no-arbitrage estimator of a valid SDF. The latter constitutes our major identification assumption. Our consistent estimator of a valid SDF 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. On the Appendix, we show that our approach can be replicated using a general Taylor Expansion of the Asset-Pricing Equation, deriving a one-factor model for the *logarithm* of returns that does not impose conditional joint log-Normality for returns and a valid SDF.

A benefit of our approach is that we are able to study intertemporal asset pricing without the need to characterize preferences or to use consumption data, in a setting similar to that of 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 SDF estimation. 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 the equity-premium puzzle, excess sensitivity (excess smoothness) in consumption, etc.; see Flavin (1981), Hansen and Singleton (1982, 1983, 1984) Mehra and Prescott (1985), and Campbell and Deaton (1989).

Our approach is related to research done in different fields. From econometrics, it is related to the *common-feature* literature after Engle and Kozicki (1993), since the SDF can be viewed as a common feature (factor) of asset returns. 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), Hecq, Palm and Urbain (2006), Issler and Lima (2009), Athanasopoulos et al. (2011), and the financial econometrics literature related to the SDF approach, best represented by the work of Garcia, and Renault (2001), Rosenberg and Engle (2002), Bansal and Yaron (2004), Hansen and Scheinkman (2009), Pukthuanthong and Roll (2015), Christensen (2017), and Almeida, Ardison, and Garcia (2020). It is also related respectively to work on *common factors* in macroeconomics and in finance; see Stock and Watson (2002), Forni et al. (2000, 2005), Bai and Ng (2002, 2004), and Bai (2009), as examples of the former, and a large literature in finance based on the work of Fama and French (1992, 1993, 1996, 2015), Lettau and Ludvigson (2001), Sentana (2004), Sentana, Calzolari, and Fiorentini (2008), Harvey, Liu, and Zhu (2016), Campello, Galvão and Juhl (2019), Kozak, Nagel, and Santosh (2019) and Giglio, Kelly, and Xiu (2021) as examples of the latter. From macroeconomics, it is also related to panel-data studies testing optimal behavior in consumption, e.g., Runkle (1991), Attanasio and Browning (1995), Attanasio and Weber (1995), and Gomes and Issler (2017).

Empirically, based on a consistent estimator for a valid SDF, we first investigate which type of utility function best fits U.S. data among popular preference specifications used in the literature: the constant-relative-risk-aversion (CRRA) coefficient utility function; the external habit utility function; and the Kreps-Porteus specification proposed by Epstein and Zin (1991). Second, using estimation results, we present a no-arbitrage simulation study assessing how close our consistent SDF estimator is to actual SDF for medium-size panel-data samples. Finally, we estimate a multi-effect linear regression model that allows for parameter heterogeneity in the intercept and in its slope that is consistent with our derived one-factor model. Using regression results we assess how well this heterogeneous one-factor model fits the cross-section and time-series realizations of assets returns.

In our first application, with quarterly data, 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, with an equivalent average real annual discount factor of 0.96. When we examined the appropriateness of different functional forms to represent preferences (Power Utility, External Habit and Kreps-Porteus), we concluded that none of these standard preference representations are rejected by the data. However, since the External-Habit and the Kreps-Porteus specifications encompass the Power Utility specification, and we have not rejected a role for habit in the first, and for the return of the optimal portfolio in the second, we conclude that these two are our preferred specifications. It should be noted that these results are aligned with the current dominant view in the macro-finance literature: Campbell and Cochrane (1999) propose the external habit model to solve well-known puzzles in finance, whereas Epstein and Zin's model, separating risk aversion and intertemporal substitution, is a preferred specification in the asset-pricing literature following the work of Bansal and Yaron (2004).

In our second application, we generate data using a no-arbitrage dynamic (consumption) capital asset-pricing model and employ it to investigate how close our proposed estimate is to the actual SDF across simulations. The model is mid-size in terms of observations in the time-series (100) and the cross-sectional ($1,000$) dimensions. It entails heterogeneity in the first and second moments of asset returns, despite a predominant role for the common component represented by the SDF, which can generate sizable cross correlations among returns, depending on the value of the relative risk-aversion coefficient. On average, the proposed estimator is very close to the actual SDF, despite the fact that we employ a relatively small sample vis-a-vis the asymptotic framework. This result survives robustness-analysis exercises on several dimensions.

In our third application, we try to approximate the asymptotic environment with monthly U.S. time-series return data from 1980:1 through 2020:12 ($T = 492$ observations), collected for $N = 102,698$ assets, grouped in the following four categories: mutual funds ($68,085$), stocks ($29,627$), real-estate investment trusts REITs ($1,000$), and government bonds ($3,986$). Our estimate of M_t has an average of 0.9958 on a monthly basis, which amounts to 0.9504 on an equivalent yearly basis. We employed the *mixed-effect* panel-data model (also known as the *mixed linear* model) to assess the fit of our one-factor model to the data using our estimate of the SDF. This model takes into account individual heterogeneity in regression coefficients in estimation. Despite the fact that our sample includes the 1987 Black Monday episode, the burst of the Dotcom Bubble, the Great Recession, and the recent Covid-19 pandemic, the results show a good in-sample fit for our panel of returns, where fitted and actual values are aligned close to the 45 degree line, with the exception of a few outliers associated with those episodes.

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 macroeconomics and finance using our estimator: estimating preference parameters using the Consumption-based Capital Asset-Pricing Model (CCAPM), a Monte-Carlo simulation study comparing the actual SDF with our consistent estimator for mid-size panel samples, and finally and evaluation of how well our factor models fits in-sample a large panel data of asset returns. Section 4 concludes. A Technical Appendix contains our main results re-stated without the use of stringent assumptions used in the simplified version of the one-factor model.

2 Economic Theory and an SDF Estimator

2.1 The Core Idea

Harrison and Kreps (1979), Hansen and Richard (1987), and Hansen and Jagannathan (1991) describe a general framework to asset pricing, associated with the stochastic discount factor (SDF), which relies on the Asset-Pricing Equation[1]:

$$\mathbb{E}_{t-1}\left\{M_t x_{i,t}\right\} = p_{i,t-1}, \qquad i = 1, 2, \ldots, N, \text{ or}, \tag{1}$$

$$\mathbb{E}_{t-1}\left\{M_t R_{i,t}\right\} = 1, \qquad i = 1, 2, \ldots, N, \tag{2}$$

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

The existence of a SDF M_t 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, however, requires the existence of complete markets. If markets are

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

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 pricing all traded securities. Despite that, there will still exist a unique discount factor M_t^*, which is an element of the payoff space, pricing all traded securities. Moreover, any discount factor M_t can be decomposed as the sum of M_t^* and an error term orthogonal to payoffs, i.e., $M_t = M_t^* + \nu_t$, where $\mathbb{E}_t(\nu_t x_{i,t}) = 0$. The important fact here is that the pricing implications of any M_t are the same as those of M_t^*, also known as the mimicking portfolio.

We now state the basic assumptions needed to construct our approach and our consistent estimator of M_t:

Assumption 1 (Absence of Arbitrage Opportunities): We assume the absence of arbitrage opportunities in asset pricing, c.f., Ross (1976). This must hold for all $t = 1, 2, ..., T$.

Assumption 2 (Joint Weak Stationarity and Ergodicity): 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 scalar process $\{\ln(M_t)\}$ and the vector process $\{\ln(\mathbf{R}_t)\}$. We assume that $\{\ln(M_t)\}$ and $\{\ln(\mathbf{R}_t)\}$ are jointly covariance-stationary processes with finite first and second moments across assets (i). We also assume that $\{\ln(M_t)\}$, $\{\ln(\mathbf{R}_t)\}$, and $\{\ln(M_t\mathbf{R}_t)\}$ are ergodic processes.

Assumption 1 is a necessary and sufficient condition for the Pricing Equation (2) to hold. Equation (2) 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.

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 M_t is required here because we will take logs of M_t. The second implication 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).[2]

Assumption 2 controls the degree of time-series dependence in the data. Joint covariance-stationarity for $\{\ln(M_t)\}$ and $\{\ln(\mathbf{R}_t)\}$ implies that each of these processes is covariance stationary and linear combinations of the form $\{a\ln(M_t)\mathbf{I}_N + b\ln(\mathbf{R}_t)\}$ are also covariance stationary for any two finite arbitrary constants a and b. We are particularly interested in $a = b = 1$, which would yield covariance-stationarity for $\{\ln(M_t\mathbf{R}_t)\}$ as well. Assumption 2 disciplines the data being used to form SDF estimators and there is little one can do in a world where covariance-stationarity and ergodicity do not hold.

Here, we seek a linear logarithmic representation for M_t and $\mathbf{R}_t$, in the simplest way possible in order to convey the core idea of this paper. To do so, we follow the basic assumptions in Hansen and Singleton (1983) to prove our main result. However, in the Appendix, we show that this result holds under a much less restrictive set of assumptions. Hansen and Singleton assume joint conditional log-Normality for $\ln(M_t)$ and $\ln(\mathbf{R}_t)$ and homoskedasticity (across time) for conditional variances. This allows to solve explicitly equation (2) to obtain:

$$r_{i,t} = -m_t - \frac{1}{2}\sigma_i^2 + \varepsilon_{i,t}, \qquad i = 1, 2, \cdots, N, \tag{3}$$

where $r_{i,t} = \ln(R_{i,t})$, $m_t = \ln(M_t)$, and $(m_t + r_{i,t}) - \mathbb{E}_{t-1}(m_t + r_{i,t}) = \varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_i^2)$. Although in the time dimension $\varepsilon_{i,t}$ is a martingale difference, i.e., $\mathbb{E}_{t-1}(\varepsilon_{i,t}) = 0$, it can have strong cross-sectional dependence. Of course, we could expand the formula for $\sigma_i^2 = \sigma_m^2 + \sigma_{r_i}^2 + 2\text{COV}(m_t, r_{i,t})$, where σ_m^2 is the variance of m_t and $\sigma_{r_i}^2$ is the variance of $r_{i,t}$, but at this point we wish to keep the notation simple enough to convey our core idea.

Since (3) holds for all i, but m_t does not vary across i, one is tempted to cross-sectionally aggregate it as follows:

$$\frac{1}{N}\sum_{i=1}^{N} r_{i,t} = -m_t - \frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 + \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}, \tag{4}$$

[2] In that sense, as the number of assets in $\mathbf{R}_t$ increases, we should avoid piling up assets related essentially to the same risk factors, such as a stock and several derivatives attached to it.

to obtain a consistent estimator of m_t, which will be based on $\frac{1}{N}\sum_{i=1}^{N} r_{i,t}$. However, it is unclear whether or not:

$$\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t} \xrightarrow{p} 0. \tag{5}$$

Recall that,

$$\varepsilon_{i,t} = (m_t + r_{i,t}) - \mathbb{E}_{t-1}(m_t + r_{i,t}) = (m_t - \mathbb{E}_{t-1}(m_t)) + (r_{i,t} - \mathbb{E}_{t-1}(r_{i,t})) = \varepsilon_t^m + \mu_{i,t}, \tag{6}$$

where ε_t^m and $\mu_{i,t}$ are innovations to m_t and $r_{i,t}$, respectively. There is an explicit part of $\varepsilon_{i,t}$ that depends on m_t alone, ε_t^m, which does not vary across i. The other part, $\mu_{i,t}$, also depends on ε_t^m. Consider now:

$$\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t} = \frac{1}{N}\sum_{i=1}^{N}\left(\varepsilon_t^m + \mu_{i,t}\right). \tag{7}$$

In order to have:

$$\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t} = 0, \text{ we need, } \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mu_{i,t} = -\varepsilon_t^m,$$

which seems at first as a knife-edge case.

Indeed, this is not true. And the reason is simple, m_t is a latent factor. As the pioneering work of Lawley and Maxwell (1971) forcefully shows, there is an inherent indeterminacy problem to identify factors and their respective loadings. To be able to identify m_t, we are allowed to choose the scale for m_t itself, and therefore to its innovation ε_t^m. Instead of dealing with the innovations ε_t^m and $\mu_{i,t}$, we can look directly at the role of $\varepsilon_{i,t}$ in (3). What may prevent the law-of-large numbers to hold in the cross-sectional dimension is the existence of a pervasive element in $\varepsilon_{i,t}$. Based on asset-pricing theory, all the pervasive factors for returns are in m_t. So, as long as we can make $\varepsilon_{i,t}$ devoid of m_t, we could apply the law-of-large numbers. Consider now an OLS-projection of $\varepsilon_{i,t}$ onto

$m_t - \mathbb{E}(m_t)$ to obtain:

$$\varepsilon_{i,t} = \delta_i (m_t - \mathbb{E}(m_t)) + \xi_{i,t}, \qquad \text{where } \delta_i = \frac{\text{COV}(\varepsilon_{i,t}, m_t)}{\text{VAR}(m_t)}, \tag{8}$$

which delivers a pervasive-idiosyncratic decomposition of $\varepsilon_{i,t}$, where $\xi_{i,t}$ is the idiosyncratic part and $\delta_i (m_t - \mathbb{E}(m_t))$ is the pervasive part associated with the stochastic discount factor.[3] If we combine (3) with (8), we obtain an affine beta-model for the log of asset returns, $r_{i,t}$, where there is a single factor – the log of a valid SDF, m_t:

$$\begin{aligned} r_{i,t} &= -\left(\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}(m_t)\right) + \beta_i m_t + \xi_{i,t}, \qquad i = 1, 2, \cdots, N, \text{ where,} \\ \beta_i &= -1 + \delta_i, \end{aligned} \tag{9}$$

and its error term $\xi_{i,t}$ is idiosyncratic by construction. Moreover, whereas (3) is not a proper *beta model*, since all assets have the same beta-coefficient for the factor m_t, in (9) the loadings for m_t vary with i – a standard feature of the whole literature in finance. So, this is a proper factor model with a single factor. However, it does not apply to returns, but to their logarithms.

A cross-sectional average of (9) yields:

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \xi_{i,t} = 0, \tag{10}$$

which gives:

$$\begin{aligned} \underset{N\to\infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} r_{i,t} &= \left(\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \beta_i\right) m_t - \lim_{N\to\infty} \frac{1}{2} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 - \left(\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_i\right) \mathbb{E}(m_t) \\ &= \overline{\beta} m_t - \overline{\sigma^2}, \end{aligned} \tag{11}$$

[3] Since $\varepsilon_{i,t}$ is a martingale-difference:

$$0 = \mathbb{E}_{t-1}(\varepsilon_{i,t}) = \delta_i \mathbb{E}_{t-1}(m_t - \mathbb{E}(m_t)) + \mathbb{E}_{t-1}(\xi_{i,t}),$$

which guarantees that $\xi_{i,t}$ has a zero unconditional mean, using the law of iterated expectations.

as long as $-\infty < \lim_{N\to\infty}\left(\frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 + \frac{1}{N}\sum_{i=1}^{N}\delta_i\mathbb{E}\left(m_t\right)\right) = \overline{\sigma^2} < \infty$ and $-\infty < \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\beta_i = \overline{\beta} < \infty$. Hence, $\frac{1}{N}\sum_{i=1}^{N} r_{i,t}$ will consistently estimate m_t apart from two bounded constant terms: $\overline{\sigma^2}$ and $\overline{\beta}$. In the log-linear beta model (9), the former is an intercept bias term and the latter is a slope bias term.

We now discuss identification of m_t and of M_t from equation (11). One key element of the identification strategy in this paper is that we will use economic theory to be able to identify M_t. Indeed, we want to keep our estimator as a *no-arbitrage* estimator of M_t. Then, the only possible tool at hand to recover any of these bias terms is to employ the asset-pricing equation itself, i.e., equation (2): we have just one restriction to identify two parameters and need an extra restriction, consistent with equation (2).

One critical issue for identification is that the asset-pricing equation applies to the level of a valid SDF, $M_t = \exp\left(m_t\right)$, but (11) offers a consistent estimator for an affine function of its logarithm, m_t. This is not a problem: the exponential function is continuous, so we can resort to the Continuous Mapping Theorem. Multiply (9) by minus one, take its cross-sectional average, and obtain:

$$-\frac{1}{N}\sum_{i=1}^{N} r_{i,t} = -\left(\frac{1}{N}\sum_{i=1}^{N}\beta_i\right)m_t + \frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 + \frac{1}{N}\sum_{i=1}^{N}\delta_i\mathbb{E}\left(m_t\right) - \frac{1}{N}\sum_{i=1}^{N}\xi_{i,t}. \qquad (12)$$

As $N \to \infty$, apply the Continuous Mapping Theorem to obtain:

$$\underset{N\to\infty}{\text{plim}}\prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}} = \exp\left(\overline{\sigma^2}\right) \times M_t^{-\overline{\beta}}. \qquad (13)$$

Trying to match the right-hand side of (13) to the asset-pricing equation (2), it is important to note that the asset-pricing equation only applies to M_t with unit power. So, a natural starting point to identify $\overline{\sigma^2}$ is to assume that:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\beta_i = \overline{\beta} = -1, \quad \text{implying,} \quad \underset{N\to\infty}{\text{plim}}\prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}} = \exp\left(\overline{\sigma^2}\right) \times M_t. \qquad (14)$$

Indeed (14) is the critical identification assumption used in this paper. It is consistent with the Asset-Pricing Equation since it allows us to further employ it to identify $\overline{\sigma^2}$ as our next step – given that now the right-hand side of the implied equation in (14) shows M_t with unit power. So, we label it a no-arbitrage identification assumption.

Assumption 3 (No-Arbitrage Identification): We assume that (14) holds, i.e., that $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\beta_i = \overline{\beta} = -1$.

Under Assumption 3, as $N \to \infty$, we obtain:

$$\overline{R}_t^G = \prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}} \xrightarrow{p} \exp\left(\overline{\sigma^2}\right) \times M_t, \tag{15}$$

where $\overline{R}_t^G$ is the geometric average of reciprocals of returns with equal weights $1/N$.

Note that Assumption 3 implies that $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\delta_i = 0$, therefore:

$$\lim_{N\to\infty}\left(\frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 + \frac{1}{N}\sum_{i=1}^{N}\delta_i\mathbb{E}\left(m_t\right)\right) = \lim_{N\to\infty}\frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 = \overline{\sigma^2}.$$

We are now left with the task of obtaining a consistent estimate for $\exp\left(\overline{\sigma^2}\right)$ to consistently estimate M_t. Multiply equation (2) by $\exp\left(\overline{\sigma^2}\right)$ to get:

$$\mathbb{E}_{t-1}\left\{\exp\left(\overline{\sigma^2}\right)M_t R_{i,t}\right\} = \exp\left(\overline{\sigma^2}\right), \qquad i = 1, 2, \ldots, N. \tag{16}$$

Take now the unconditional expectation of (16), use the law-of-iterated expectations, and average across $i = 1, 2, ..., N$, to get:

$$\exp\left(\overline{\sigma^2}\right) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left\{\exp\left(\overline{\sigma^2}\right)M_t R_{i,t}\right\} = \mathbb{E}\left\{\exp\left(\overline{\sigma^2}\right)M_t\frac{1}{N}\sum_{i=1}^{N}R_{i,t}\right\}. \tag{17}$$

Now, recall that (15) offers $\overline{R}_t^G$ as a consistent estimator for $\exp\left(\overline{\sigma^2}\right) \times M_t$. If we replace the latter with the former, as $T \to \infty$, we can then obtain a consistent estimator for $\exp\left(\overline{\sigma^2}\right)$ as follows:

$$\widehat{\exp\left(\overline{\sigma^2}\right)} = \frac{1}{T}\sum_{t=1}^{T}\left(\prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}} \frac{1}{N}\sum_{i=1}^{N} R_{i,t}\right) = \frac{1}{T}\sum_{t=1}^{T} \overline{R}_t^G \overline{R}_t^A, \tag{18}$$

where $\overline{R}_t^A = \frac{1}{N}\sum_{i=1}^{N} R_{i,t}$ is the arithmetic average of returns.

Note that the estimator of $\exp\left(\overline{\sigma^2}\right)$ is the sample counterpart of the right-hand-side of (17) in the time dimension, when we replaced $\exp\left(\overline{\sigma^2}\right) M_t$ with its consistent estimator $\overline{R}_t^G$. Finally, taking into account the results in (15) and in (18), we are able to propose a consistent estimator for M_t as the ratio:

$$\widehat{M_t} = \frac{\overline{R}_t^G}{\frac{1}{T}\sum_{j=1}^{T} \overline{R}_j^G \overline{R}_j^A}.$$

This is a simple function of the arithmetic average of returns, $\overline{R}_t^A$, and the geometric average of the reciprocal of returns, $\overline{R}_t^G$, both using equal weights $1/N$.

One important technical issue is that we need to let $N \to \infty$ first, before letting $T \to \infty$. This happens because we have substituted $\exp\left(\overline{\sigma^2}\right) M_t$ in (17) by its consistent estimator $\overline{R}_t^G$ before constructing (18) using the time dimension. We now state our main result.

Proposition 1 *Using the setup of Hansen and Singleton (1983) and Assumptions 1-3, the following holds:*

$$\underset{(N,T\to\infty)_{\text{seq}}.}{plim} \frac{\overline{R}_t^G}{\frac{1}{T}\sum_{j=1}^{T} \overline{R}_j^G \overline{R}_j^A} = M_t,$$

for all t, where $(N, T \to \infty)_{\text{seq}}$ denotes the sequential asymptotic approach proposed by Phillips and Moon (1999), when we let first $N \to \infty$, and then let $T \to \infty$.

It is important to note that $\widehat{M_t}$ is a function of N and T. The denominator explicitly shows that it depends on T and dependence on N is implicit since $\overline{R}_t^G$ and $\overline{R}_t^A$ are respectively geometric and arithmetic averages in the cross-sectional dimension. The only reason why we do not explicitly state its dependence on N and T is to avoid a cumbersome notation for $\widehat{M_t}$.

2.2 Discussing Identification

The Asset-Pricing Equation is a non-linear function of a valid SDF and of individual returns. But, we have shown above how to derive an exact log-linear relationship between returns and a valid SDF, which allows for a natural one-factor affine model linking $r_{i,t}$, $i = 1, 2, \cdots$ and m_t. Our SDF estimator is constructed using large-sample techniques under no-arbitrage, with $N \to \infty$ first, and then $T \to \infty$, since it relies on the existence of the Asset-Pricing Equation (2) to generate (9) and then imposes (14) to employ once again (2) to identify $\overline{\sigma^2}$ and them M_t. The latter is a no-arbitrage assumption because equation (2) only holds with a unit power for M_t. In our setup, the natural choice to identify and estimate of $\exp\left(\overline{\sigma^2}\right)$ is to use the additional assumption that $\overline{\beta} = -1$. Technically speaking, $\overline{\sigma^2}$ and $\overline{\beta}$ are not separately identifiable, setting $\overline{\beta} = -1$ identifies both under no-arbitrage.

Recall that our proposed affine beta-model is as follows:

$$r_{i,t} = -\left(\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}\left(m_t\right)\right) + \beta_i m_t + \xi_{i,t}, \qquad i = 1, 2, \cdots, N, \text{ where,} \quad \beta_i = -1 + \delta_i. \tag{19}$$

In the context of equation (19), from the seminal work of Lawley and Maxwell (1971), the indeterminacy problem for factor models can be explained by the following equation:

$$\widetilde{\beta}_i \widetilde{m}_t = \left(\widetilde{\beta}_i \kappa^{-1}\right)\left(\kappa \widetilde{m}_t\right) = \beta_i m_t, \tag{20}$$

i.e., a factor and its respective loadings are indentifiable only up to multiplication by non-zero finite scalars κ^{-1} and κ, which make equation (20) hold. Clearly, κ determines the *scale* of m_t and also the mean of β_i. Suppose, for example, that:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \widetilde{\beta}_i = \lambda \neq -1,$$

then, we can set $\kappa = -\lambda$, which will make $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \beta_i = \lambda(-\lambda)^{-1} = -1$, allowing (19) to hold for $\beta_i m_t$ with $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \beta_i = -1$.

A solution for this same problem is also present in principal-component analysis – a traditional method for identifying factors and factor loadings; see, e.g., Stock and Watson (2002). Denote by $\Sigma_{\mathbf{r}} = \mathbb{E}\left\{[\mathbf{r}_t - \mathbb{E}(\mathbf{r}_t)][\mathbf{r}_t - \mathbb{E}(\mathbf{r}_t)]'\right\}$ the variance-covariance matrix of logged returns, where $\mathbf{r}_t$ is an $N \times 1$ vector of returns. The first principal component is identified as the first factor. It is chosen to be a linear combination $\alpha'(\mathbf{r}_t - \mathbb{E}(\mathbf{r}_t))$ with maximal variance $\alpha'\Sigma_{\mathbf{r}}\alpha$. As is well-known, the problem has no unique solution, since we can make $\alpha'\Sigma_{\mathbf{r}}\alpha$ as large as we want by multiplying α by a constant *scale* $c > 1$. Indeed, we are facing the same *scale* problem present in equation (20). In a fixed N setting, the solution is to constrain the variance of the first principal component by imposing that α has unit norm, i.e., that $\alpha'\alpha = 1$.

Our identification strategy $\left(\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \beta_i = -1\right)$ imposes:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_i = 0, \tag{21}$$

i.e., in the limit, the projection coefficients of $\varepsilon_{i,t}$ onto $m_t - \mathbb{E}(m_t)$ *must* cancel out exactly.

From a different angle, compute the innovations $r_{i,t} - \mathbb{E}_{t-1}(r_{i,t})$ in the beta-model (9) to obtain:

$$\mu_{i,t} = \beta_i \varepsilon_t^m + \xi_{i,t},$$

which makes:

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mu_{i,t} = \overline{\beta} \varepsilon_t^m. \tag{22}$$

Using our decomposition in (6), we can also perform an OLS projection of $\mu_{i,t}$ onto ε_t^m, as follows:

$$\mu_{i,t} = \tau_i \varepsilon_t^m + \zeta_{i,t}, \qquad \text{where } \tau_i = \frac{\text{COV}\left(\mu_{i,t}, \varepsilon_t^m\right)}{\text{VAR}\left(\varepsilon_t^m\right)},$$

where the $\tau_i \varepsilon_t^m$ term captures the pervasive portion of $\mu_{i,t}$, and $\zeta_{i,t}$ captures the idiosyncratic portion. This implies that:

$$\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mu_{i,t}=\tau\varepsilon_t^m, \tag{23}$$

where $-\infty \leq \lim\limits_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\tau_i=\tau\leq\infty$, and $\operatorname*{plim}\limits_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\zeta_{i,t}=0$.

It is instructive to compare (22) and (23): our identification assumption imposes $\overline{\beta}=\tau=-1$ to deliver the exact *scale* for ε_t^m – and thus the scale for m_t, and the power for M_t – to be able to apply the Asset-Pricing Equation to identify $\exp\left(\overline{\sigma^2}\right)$ and then M_t. As argued above, solving a *scale* problem is intrinsic to factor-model analysis – a choice *must* be made. We chose to use economic theory (no-arbitrage, the asset-pricing equation) to identify $\overline{\beta}$, $\exp\left(\overline{\sigma^2}\right)$ and then M_t.

Under our identification assumption, we obtain the following:

$$m_t=-\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}r_{i,t}-\overline{\sigma^2},$$

where the following variance restriction holds in the time dimension:

$$\operatorname*{plim}_{N\to\infty}\mathrm{VAR}\left(\frac{1}{N}\sum_{i=1}^{N}r_{i,t}\right)=\mathrm{VAR}\left(m_t\right).$$

These results map m_t into the negative of the return of an equally weighted *market-portfolio*, which validates a limiting beta of -1.

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 in the no-arbitrage class of SDF estimators, since it is based on the asset-pricing equation and uses it to identify M_t. Moreover, it is completely non-parametric and a function of asset-return data alone. Whatsoever, no assumptions about preferences have been made in identification.

Second, because $\widehat{M_t}$ is a consistent estimator, it is interesting to discuss to what it 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 M_t as well. 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 valid SDFs.

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

$$\begin{aligned}
\underset{(N,T\to\infty)_{\text{seq.}}}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\widehat{M_t}R_{i,t} &= \underset{(N,T\to\infty)_{\text{seq.}}}{\text{plim}} \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\frac{\overline{R}_t^G}{\frac{1}{T}\sum_{j=1}^{T}\overline{R}_j^G\overline{R}_j^A}R_{i,t} \\
&= \underset{(N,T\to\infty)_{\text{seq.}}}{\text{plim}} \frac{1}{T}\sum_{t=1}^{T}\frac{\overline{R}_t^G\overline{R}_t^A}{\frac{1}{T}\sum_{j=1}^{T}\overline{R}_j^G\overline{R}_j^A} \\
&= \underset{T\to\infty}{\text{plim}}\frac{1}{T}\sum_{t=1}^{T}\frac{R_t^G R_t^A}{\frac{1}{T}\sum_{j=1}^{T}R_j^G R_j^A} = 1, \qquad (24)
\end{aligned}$$

where R_t^A and R_t^G are the probability limits of $\overline{R}_t^G$ and $\overline{R}_t^A$, respectively, when we let $N \to \infty$.

Note that the result in (24) is a natural property arising from the moment restrictions entailed by the Asset-Pricing Equation (2), when populational means are replaced by sample means. In finite samples, it does not price correctly any specific asset. However, on average, it will price correctly all the assets in the economy. Moreover, in the limit, the average pricing errors will be nil due to (24). The latter is a very attractive property.

We now check whether or not our proposed factor model (9) prices securities correctly. Our key assumption for the simplified pricing model is that $\ln(M_t R_{i,t})|\mathcal{I}_{t-1} \sim \mathcal{N}(\mathbb{E}_{t-1}(r_{i,t}+m_t);\sigma_i^2)$, where $\mathcal{I}_{t-1}$ is the information set used by the agent based on information up to period $t-1$. From the properties of the log-Normal distribution, we obtain:

$$1=\mathbb{E}_{t-1}\left(M_tR_{i,t}\right)=\exp\left\{\mathbb{E}_{t-1}\left(r_{i,t}+m_t\right)+\frac{1}{2}\sigma_i^2\right\}. \tag{25}$$

From (9), recalling that $\beta_i=-1+\delta_i$, we get:

$$r_{i,t}+m_t=-\frac{1}{2}\sigma_i^2+\delta_i\left(m_t-\mathbb{E}\left(m_t\right)\right)+\xi_{i,t}.$$

Therefore:

$$\mathbb{E}_{t-1}\left(r_{i,t}+m_t\right)=-\frac{1}{2}\sigma_i^2+\delta_i\mathbb{E}_{t-1}\left(m_t-\mathbb{E}\left(m_t\right)\right)+\mathbb{E}_{t-1}\left(\xi_{i,t}\right). \tag{26}$$

Recall now that $\varepsilon_{i,t}$ is a Martingale Difference, i.e., that $\mathbb{E}_{t-1}\left(\varepsilon_{i,t}\right)=0$. Since we have decomposed it as:

$$\varepsilon_{i,t}=\delta_i\left(m_t-\mathbb{E}\left(m_t\right)\right)+\xi_{i,t},$$

we arrive at:

$$0=\mathbb{E}_{t-1}\left(\varepsilon_{i,t}\right)=\delta_i\mathbb{E}_{t-1}\left(m_t-\mathbb{E}\left(m_t\right)\right)+\mathbb{E}_{t-1}\left(\xi_{i,t}\right). \tag{27}$$

Combining (26) and (27), we get:

$$\mathbb{E}_{t-1}\left(r_{i,t}+m_t\right)=-\frac{1}{2}\sigma_i^2,$$

but from (25),

$$1=\mathbb{E}_{t-1}\left(M_tR_{i,t}\right)=\exp\left\{-\frac{1}{2}\sigma_i^2+\frac{1}{2}\sigma_i^2\right\}=\exp\left\{0\right\}. \tag{28}$$

So, indeed, (9) is a proper asset-pricing model.

2.4 Comparisons with the Literature

Several early studies in the 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. By construction, our approach avoid these problems since it is not based on a choice for preferences.

This paper is related to the work of Hansen and Jagannathan (1991, 1997), who 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 a valid SDF, but note that an admissible SDF has its behavior (and, in particular, its variance) bounded by two restrictions. The first is the Pricing Equation (2) and the second is the fact that $M_t > 0$. They exploit the fact that it is always possible to project M onto the space of returns, which makes it straightforward to express M^*, the mimicking portfolio, only as a function of observable returns and the price of a risk-free asset.

Start with:

$$\mathbb{E}\left(\mathbf{R}_t M_t\right) = \iota, \tag{29}$$

where $\mathbf{R}_t = (R_{1,t}, R_{2,t}, \ldots\ R_{N,t})'$ is an $N \times 1$ vector, with $\mathbb{E}\left(\mathbf{R}_t\right) = \mu$, and $\mathbb{E}\left[\left(\mathbf{R}_t - \mu\right)\left(\mathbf{R}_t - \mu\right)'\right] = \Sigma_{\mathbf{R}}$, and ι is an $N \times 1$ vector stacking ones. Although we do not observe M_t, Hansen and Jagannathan propose a least-squares projection of M_t onto $(\mathbf{R}_t - \mu)$ and a constant as follows:

$$M_t = \alpha + \left(\mathbf{R}_t - \mu\right)' \beta + \upsilon_t, \tag{30}$$

where $\mathbb{E}\left(M_t\right) = \alpha$. Pre-multiplying (30) by $(\mathbf{R}_t - \mu)$ and taking expectations allows solving for β:

$$\beta = \Sigma_{\mathbf{R}}^{-1}\left(\iota - \mu\alpha\right),$$

which identifies β as long as we identify α.

If one observes the price of a risk-free asset, q_t, then one can identify α as:

$$\mathbb{E}\left(M_t\right) = \mathbb{E}\left(q_t\right) = \alpha,$$

and construct a minimum-variance estimator M_t^*, as:

$$M_t^* = \alpha + (\mathbf{R}_t - \mu)' \beta,$$

since it is the least-squares projection of M_t onto $(\mathbf{R}_t - \mu)$ – a linear function of asset returns alone – something that has inspired our approach.

Of course, for any sample of returns, we will estimate a different M_t^*, but, in principle, we can let $N \to \infty$ to obtain a single asymptotic estimator as we propose in this paper. If the number of assets is *large*, since $\Sigma_{\mathbf{R}}$ is an $N \times N$ matrix, its inversion will be problematic for large samples – making it hard to estimate β in this context.

We now ask whether we could use the same principles used to construct our SDF estimator to circumvent the problem of dealing with a large number of assets in the approach of Hansen and Jagannathan. Recall that, in constructing our estimator M_t we have first taken averages in the cross-sectional dimension. Here, this could be applied as well to circumvent the large-sample inversion problem for $\Sigma_{\mathbf{R}}$. Hence, our solution is to reduce the dimensionality of $\mathbf{R}_t$ by recognizing that the Asset-Pricing Equation implies a single factor for for $\mathbf{R}_t$ for large N. Pre-multiply (29) by $\frac{1}{N}\iota'$, to obtain:

$$\mathbb{E}\left(\overline{R}_t^A M_t\right) = 1,$$

which prices the equally-weighted portfolio of all assets – a scalar. Project M_t onto $\left(\overline{R}_t^A - \mathbb{E}\left(\overline{R}_t^A\right)\right)$ and a constant as follows:

$$M_t = \alpha + \left(\overline{R}_t^A - \mathbb{E}\left(\overline{R}_t^A\right)\right)\beta_A + \upsilon_t^A.$$

As before, we can identify α as $\mathbb{E}(q_t) = \alpha$, and β_A as:

$$\beta_A = \frac{1 - \mathbb{E}\left(\overline{R}_t^A\right)\alpha}{\text{VAR}\left(\overline{R}_t^A\right)}.$$

By using the returns of the equally-weighted portfolio, $\overline{R}_t^A$, we avoid the invertibility problem alluded above, since, as $N \to \infty$, $\overline{R}_t^A$ will still have time-series variation and $\mathrm{VAR}\left(\overline{R}_t^A\right)$ will be well defined. Hence, β_A is identified and we can construct an asymptotic version of M_t^*, as follows:

$$M_t^{**} = \alpha + \left(\overline{R}_t^A - \mathbb{E}\left(\overline{R}_t^A\right)\right)\beta_A. \tag{31}$$

This illustrates the potential gains in constructing an asymptotic estimator for M_t by using the cross-sectional average of returns, something that can only be achieved in a panel-data framework.

One may ask what is lost when we cross-sectionally aggregate returns. To answer this question, consider the well-known factor model proposed by Stock and Watson (2002) applied to $\mathbf{R}_t$ in a context of large N, T:

$$\mathbf{R}_t = \mu + \Lambda \mathbf{F}_t + e_t, \tag{32}$$

where the N assets of the economy are governed by $r \ll N$ common factors, $\mathbb{E}(\mathbf{R}_t) = \mu$, the factor loading matrix $\Lambda = (\lambda_{ij})$ is of dimension $N \times r$, the vector of zero-mean factors $\mathbf{F}_t$ is of dimension $r \times 1$, and the error term e_t is cross-sectionally independent and temporally *i.i.d.* Pre-multiply (32) by $\frac{1}{N}\iota'$ to obtain:

$$\overline{R}_t^A - \mathbb{E}\left(\overline{R}_t^A\right) = \sum_{j=1}^{r}\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_{ij}\right)F_{jt} + \frac{1}{N}\sum_{i=1}^{N}e_{it},$$

We can now examine what happens when we let $N \to \infty$. Given the properties of e_t in the cross-sectional dimension, $\frac{1}{N}\sum_{i=1}^{N} e_{it} \xrightarrow{p} 0$. Stock and Watson also assume that $\frac{1}{N}\Lambda'\Lambda \to I_r$, so the limit of $\frac{1}{N}\sum_{i=1}^{N}\lambda_{ij}$, call it λ_j, is well defined as $N \to \infty$. Therefore,

$$\operatorname*{plim}_{N\to\infty}\left(\overline{R}_t^A - \mathbb{E}\left(\overline{R}_t^A\right)\right) = R_t^A - \mathbb{E}\left(R_t^A\right) = \sum_{j=1}^{r}\lambda_j F_{jt}$$

contains a *single factor* which is a linear combination of the r original factors in (32). However, we have eliminated the idiosyncratic components of returns by cross-sectionally aggregating them for large N. In the end, our regressor for large N will be a linear function of the factors in the beta model for returns $\mathbf{R}_t$, but this is exactly the proper characterization of SDFs in a factor-model setup.

Under some conditions, our estimator of M_t 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. Since asset prices are the expected present discounted value of the dividend flow and, under 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 to aggregate capital $\overline{R}_{t+1}$ is given by:

$$\overline{R}_{t+1} = \frac{\bar{p}_{t+1} + c_{t+1}}{\bar{p}_t} = \frac{\beta c_{t+1} + (1-\beta) c_{t+1}}{\beta c_t} = \frac{c_{t+1}}{\beta c_t} = \frac{1}{M_{t+1}}, \tag{33}$$

which is the reciprocal of a valid SDF.

Indeed, Gomes and Issler (2017) exploit this property discussing the optimality of aggregate consumption, showing no signs of asset-pricing puzzles when pricing $\overline{R}_t$. This reinforces the role that the an aggregate portfolio can have in asset pricing, something that ought to deserve more attention of the finance literature.

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 for SDFs: the CRRA specification (Power Utility), 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,

$$M_t^{CRRA} = \beta \left(\frac{c_t}{c_{t-1}} \right)^{-\gamma} \tag{34}$$

$$M_t^{EH} = \beta \left(\frac{c_t}{c_{t-1}} \right)^{-\gamma} \left(\frac{c_{t-1}}{c_{t-2}} \right)^{\kappa(\gamma-1)} \tag{35}$$

$$M_t^{KP} = \left[\beta \left(\frac{c_t}{c_{t-1}} \right)^{-\gamma} \right]^{\frac{1-\gamma}{\rho}} \left(\frac{1}{B_t} \right)^{1-\frac{1-\gamma}{\rho}}, \tag{36}$$

where c_t denotes consumption, B_t is the return on the optimal portfolio, β is the discount factor[4], γ is the relative risk-aversion coefficient, and κ is the time-separation parameter in the habit-formation specification.

As is well-known, M_t^{EH} is a geometric weighted average of M_t^{CRRA} and $\left(\frac{c_{t-1}}{c_{t-2}} \right)$. If $\kappa = 0$, we are back to the CRRA specification. In the Kreps-Porteus specification the elasticity of intertemporal 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^{KP} is a geometric weighted average of M_t^{CRRA} and $\left(\frac{1}{B_t} \right)$, with weights θ and $1-\theta$, respectively. If $\theta = 1$, we are also back to the CRRA specification.

For consistent estimates, we can always write:

$$m_t = \widehat{m_t} - \eta_t, \tag{37}$$

where η_t is the approximation error between m_t and its consistent estimate $\widehat{m_t}$.

The properties of η_t will depend on the properties of M_t and $R_{i,t}$. In general, it will be serially dependent and heterogeneous. Using (37) and the expressions in (34), (35) and

[4] Not to be confused with the beta coefficients of our factor model (9), β_i,

(36), we arrive at:

$$\widehat{m_t} = \ln \beta - \gamma \Delta \ln c_t + \eta_t^{CRRA}, \tag{38}$$

$$\widehat{m_t} = \ln \beta - \gamma \Delta \ln c_t + \kappa (\gamma - 1) \Delta \ln c_{t-1} + \eta_t^{EH}, \tag{39}$$

$$\widehat{m_t} = \theta \ln \beta - \theta \gamma \Delta \ln c_t - (1 - \theta) \ln B_t + \eta_t^{KP}, \tag{40}$$

The most appealing way of estimating (38), (39) and (40), simultaneously testing for over-identifying restrictions, is to use the generalized method of moments (GMM). 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 (38) is nested into (39), we can also perform a redundancy test for $\Delta \ln c_{t-1}$ in (38). The same applies regarding (38) and (40), since the latter collapses to the former when $\ln B_t$ 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 currencies 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.” The data were extracted from different sources: NIPA data for the U.S. (Personal Consumption Expenditures of Nondurable and Services and GNP), with seasonal adjustment, were extracted from the FRED database, kept by St. Louis FED. We also extracted from FRED the Moody’s Aaa Corporate Bond Yield. All G7 return data for stocks, as well as exchange rates for all G7 countries, were extracted from the OECD database. The U.S. Treasury Bill Rate, as well as the Treasury-bill rates for G7 countries, were extracted from the International Financial Statistics (IFS) data base of the IMF. From Bloomberg, we extracted the price of Gold, the NYSE Index, and the SP500 Index. finally, the FTSE Nareit U.S. Real Estate Index was extracted from the homepage of the National Association of Real Estate Investment Trusts (Nareit).

Our sample period starts in 1977:Q2 and ends in 2019:Q2, at the quarterly frequency, comprising $T = 169$ time periods. Overall, we averaged the real U.S.$ returns on these 18 portfolios or assets[5], which are, in turn, each one 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.

In estimating equations (38) and (39) by GMM, 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.$. Since there is no constructed deflator for non-durables and services, we constructed a measure of Fisher's Ideal Price Index to deflate nominal consumption. This deflator was used throughout to deflate returns as well. We also used real per-capita GNP (y_t) as a measure of income – an instrument in running some of these regressions.

Figure 1 below shows our estimator of the SDF – $\widehat{M_t}$ – for the period 1977:Q2 to 2019:Q2. It is close to unity most of the time and bounded by the interval [0.85, 1.25]. The sample mean of $\widehat{M_t}$ is 0.99 on a quarterly basis, implying an annual discount factor of 0.96, a very reasonable estimate, especially considering that our sample includes the great recession.

[5] 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 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.

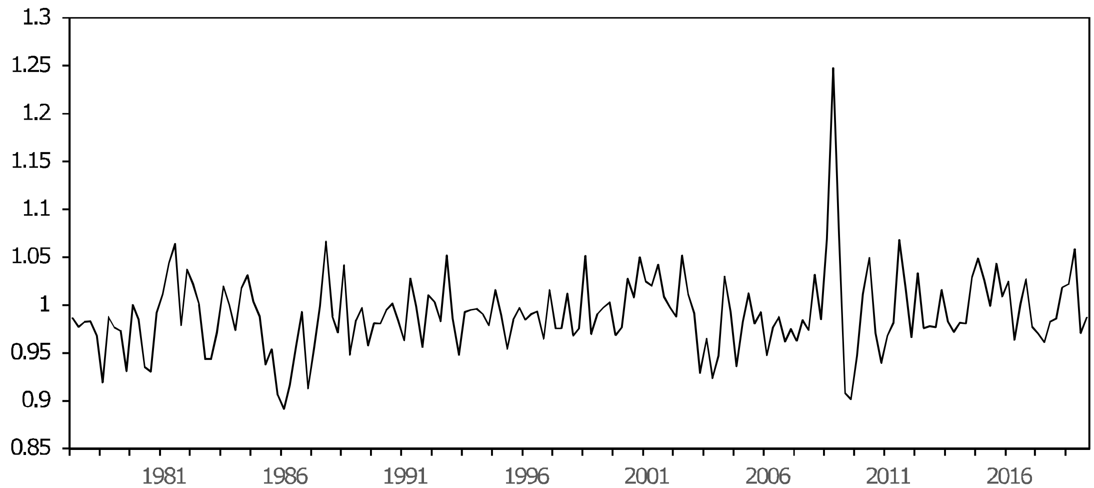

Figure 1: Estimate of the Stochastic Discount Factor for G7 Countries.

Tables 1, 2, and 3 present GMM estimation of equations (38), (39) and (40), respectively. 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 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: $\widehat{\beta}$ is very close to 0.997 for all obtained estimates and estimates for γ are all in the interval $[1.15,\ 1.40]$ – all very significant at usual levels. Our preferred regression is the last one in Table 1, where all instruments are used in estimation. There, $\widehat{\beta} = 0.997$ and $\widehat{\gamma} = 1.360$, both significant at usual levels.

Table 2 reports results obtained when (external) habit formation is considered in pref-

Power-Utility Function Estimates $m_t = \ln\beta - \gamma\Delta\ln c_t - \eta_t^{CRRA}$			
Instrument Set	β (SE)	γ (SE)	OIR Test (P-Value)
OLS Estimate	1.000388 (0,004184)	1.811034 (0.486580)	-
$r_{i,t-1}, r_{i,t-2},\ \forall i = 1, 2, ...N,$	0.997380 (0.002501)	1.403094 (0.256690)	(0.993129)
$r_{i,t-1}, r_{i,t-2},\ \forall i = 1, 2, ...N,$ $\Delta\ln c_{t-1}, \Delta\ln c_{t-2}$	0.995964 (0.002154)	1.153144 (0.198670)	(0.996510)
$r_{i,t-1}, r_{i,t-2},\ \forall i = 1, 2, ...N,$ $\Delta\ln y_{t-1}, \Delta\ln y_{t-2}$	0.996595 (0.002418)	1.241693 (0.242484)	(0.995987)
$r_{i,t-1}, r_{i,t-2},\ \forall i = 1, 2, ...N,\ \Delta\ln c_{t-1},$ $\Delta\ln c_{t-2}, \Delta\ln y_{t-1}, \Delta\ln y_{t-2}$	0.997099 (0.002206)	1.360465 (0.204433)	(0.998300)

Table 1: Power Utility Function – GMM Estimates

erences. Results are very different to those obtained with power utility. First, estimates of γ jump from the interval [1.15, 1.40] to the interval [3.59, 4.25], all statistically significant. Second, the estimates of κ are all in the interval [0.889, 1.000] and all statistically significant as well. In this case, the external-habit specification should be preferred vis-a-vis the Power Utility specification. As before, we find no evidence of rejection in over-identifying restrictions tests in any regression we have run.

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 index for that role. Estimates for the discount factor β are all higher than unity, and statistically different than unity on almost all counts. The optimal portfolio term coefficient, θ, has an estimate in the interval [0.514, 0.582], but it is neither statistically equal to zero nor statistically equal to unity at usual significance levels. Despite the odd result regarding β estimates, the Kreps-Porteus specification should be preferred vis-a-vis the Power Utility specification. Again, there is no evidence of rejection in over-identifying restrictions tests in any GMM regression we have run.

External-Habit Utility-Function Estimates $m_t = \ln\beta - \gamma\Delta\ln c_t + \kappa(\gamma-1)\Delta\ln c_{t-1} - \eta_t^{EH}$				
Instrument Set	β (SE)	γ (SE)	κ (SE)	OIR Test (P-Value)
OLS Estimate	0.997260 (0.004241)	3.307682 (0.710761)	0.858723 (0.210985)	-
$r_{i,t-1}, r_{i,t-2}$, $\forall i = 1, 2, ...N$,	0.996853 (0.002517)	3.737319 (0.685271)	0.888717 (0.089653)	(0.990324)
$r_{i,t-1}, r_{i,t-2}$, $\forall i = 1, 2, ...N$, $\Delta\ln c_{t-1}, \Delta\ln c_{t-2}$	0.994999 (0.002367)	3.588043 (0.677137)	0.999759 (0.092894)	(0.992727)
$r_{i,t-1}, r_{i,t-2}$, $\forall i = 1, 2, ...N$, $\Delta\ln y_{t-1}, \Delta\ln y_{t-2}$	0.996288 (0.002446)	3.707508 (0.661173)	0.932906 (0.088334)	(0.994561)
$r_{i,t-1}, r_{i,t-2}$, $\forall i = 1, 2, ...N$, $\Delta\ln c_{t-1}$, $\Delta\ln c_{t-2}, \Delta\ln y_{t-1}, \Delta\ln y_{t-2}$	0.997110 (0.002346)	4.246917 (0.659643)	0.895911 (0.068874)	(0.996886)

Table 2: External Habit Utility Function – GMM Estimates

Since the External-Habit and the Kreps-Porteus specifications encompass the power utility specification, and we have rejected $H_0 : \kappa = 0$ for the former, and have rejected $H_0 : \theta = 1$ for the latter, we conclude that these two are our preferred specifications. It should be noted that Campbell and Cochrane (1999) propose the external habit model to solve well-known puzzles in finance, whereas the Epstein and Zin's specification separating risk aversion and intertemporal substitution is currently a preferred specification in the finance literature. In that sense, our empirical results so far are aligned with the dominant view in the macro-finance literature.

3.2 Simulation Study

Pukthuanthong and Roll (2015) propose an interesting *simulation* study, based on what they call an *agnostic* view of stochastic discount factors (SDFs), consistent with our approach, since it is based on the Asset-Pricing Equation. First, they generate a gross riskless rate, R_t^f, and a SDF at time $t = 1, 2, \cdots, T$, as:

$$M_t = \frac{1}{R_t^f} \exp\left(\mu_t - \frac{\sigma_\mu^2}{2}\right),$$

Kreps-Porteus Utility-Function Estimates $m_t = \theta \ln \beta - \theta\gamma\Delta \ln c_t - (1-\theta) \ln B_t - \eta_t^{KP}$				
Instrument Set	β (SE)	γ (SE)	θ (SE)	OIR Test (P-Value)
OLS Estimate	1.003671 (0.004609)	2.212549 (0.532732)	0.643915 (0.027097)	
$r_{i,t-1}, r_{i,t-2}$, $\forall i = 1, 2, ...N$,	1.008285 (0.002686)	2.743907 (0.363743)	0.514460 (0.030306)	(0.983547)
$r_{i,t-1}, r_{i,t-2}$, $\forall i = 1, 2, ...N$, $\Delta \ln c_{t-1}, \Delta \ln c_{t-2}$	1.004015 (0.002152)	2.195702 (0.278868)	0.581821 (0.022761)	(0.994278)
$r_{i,t-1}, r_{i,t-2}$, $\forall i = 1, 2, ...N$, $\Delta \ln y_{t-1}, \Delta \ln y_{t-2}$	1.004394 (0.002213)	2.210713 (0.265318)	0.578335 (0.025315)	(0.994313)
$r_{i,t-1}, r_{i,t-2}$, $\forall i = 1, 2, ...N$, $\Delta \ln c_{t-1}$, $\Delta \ln c_{t-2}, \Delta \ln y_{t-1}, \Delta \ln y_{t-2}$	1.005907 (0.002263)	2.521649 (0.286836)	0.565332 (0.020057)	(0.997809)

Table 3: Kreps-Porteus Utility Function – GMM Estimates

where $\mu_t \sim i.i.d.\mathcal{N}\left(0, \sigma_\mu^2\right)$. Second, they generate gross *un-scaled* returns using:

$$\widetilde{R}_{i,t} = \phi \exp\left(\zeta_{i,t} - \frac{\sigma_\zeta^2}{2}\right),$$

where ϕ is the expected gross return for asset i, assumed the same across assets, and $\zeta_{i,t} \sim i.i.d.\mathcal{N}\left(0, \sigma_\zeta^2\right)$, where σ_ζ^2 does not vary across i as well. Final *scaled* returns are computed as:

$$R_{i,t} = \frac{\widetilde{R}_{i,t}}{\frac{1}{T}\sum_{t=1}^{T} M_t \widetilde{R}_{i,t}} \exp\left(\upsilon_{i,t} - \frac{\sigma_\upsilon^2}{2}\right),$$

where $\upsilon_{i,t} \sim i.i.d.\mathcal{N}\left(0, \sigma_\upsilon^2\right)$, which implies that the Asset-Pricing Equation holds unconditionally in the time dimension for every asset i, i.e.,

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

Here, we implement a broader simulation study, which includes the joint dynamics of gross returns and of the SDF. In searching for a dynamic model which would be tractable, we decided to employ the (consumption) capital-asset-pricing model (CCAPM) setup of Hansen and Singleton (1983), where they assume that $\ln\left(M_t \mathbf{R}_t\right)$ has a conditional multivariate *Normal* distribution with homoskedastic variance in the time dimension, i.e., $\ln(M_t \mathbf{R}_{i,t})|\mathcal{I}_{t-1} \sim \mathcal{N}(\mathbb{E}_{t-1}(r_{i,t}+m_t);\sigma_i^2)$, where we can further decompose $\sigma_i^2 = \sigma_m^2 + \sigma_{r_i}^2 + 2\text{COV}(m_t, r_{i,t})$, where σ_m^2 is the variance of m_t and $\sigma_{r_i}^2$ is the variance of $r_{i,t}$. They also assume the *Power-Utility* specification for the utility function, $u(c_t)$, which yields:

$$M_t = \beta \left(\frac{c_t}{c_{t-1}} \right)^{-\gamma},$$

where γ is the relative risk-aversion coefficient and β is the discount factor. As discussed above, these assumptions give us a log-linear system on $r_{i,t}$ and $\Delta \ln(c_t)$, $i = 1, \ldots, N$, as follows:

$$r_{i,t} - \gamma \Delta \ln(c_t) = -\ln(\beta) - \frac{\sigma_i^2}{2} + \varepsilon_{i,t}, \qquad i = 1, \ldots, N. \tag{41}$$

Hansen and Singleton show that the dynamic representation of the data follows a vector autoregression $VAR(p)$ process with contemporaneous relationships and the same reduced-rank constraints exploited by Engle and Kozicki (1993), Vahid and Engle (1993), and Engle and Issler (1995), in the common-feature literature. Indeed, if we denote by $\mathbf{X}_t = (r_{1,t}, r_{2,t}, \cdots, r_{N,t}, \Delta \ln(c_t))'$ an $(N+1)\times 1$ vector containing all logged gross returns and the instantaneous growth rate of consumption, it is straightforward to show that $\mathbf{X}_t$ obeys a *structural* $VAR(p)$ model with Gaussian errors, as follows:

$$\mathbf{A}_0 \mathbf{X}_t = \mathbf{c}_0 + \begin{pmatrix} \mathbf{0} \\ a_1 \end{pmatrix} \mathbf{X}_{t-1} + \begin{pmatrix} \mathbf{0} \\ a_2 \end{pmatrix} \mathbf{X}_{t-2} + \cdots + \begin{pmatrix} \mathbf{0} \\ a_p \end{pmatrix} \mathbf{X}_{t-p} + \mathbf{v}_t,$$

where $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_v})$, and:

$$\mathbf{A}_0 = \begin{pmatrix} 1 & ... & ... & 0 & -\gamma \\ 0 & 1 & ... & 0 & -\gamma \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 1 & -\gamma \\ 0 & ... & ... & 0 & 1 \end{pmatrix}_{(N+1)\times(N+1)}, \quad \mathbf{c}_0 = \begin{pmatrix} -\ln(\beta) - \frac{\sigma_1^2}{2} \\ -\ln(\beta) - \frac{\sigma_2^2}{2} \\ \vdots \\ -\ln(\beta) - \frac{\sigma_N^2}{2} \\ \mu_{\Delta c} - \frac{\sigma_{\Delta c}^2}{2} \end{pmatrix},$$

$$\mathbf{v}_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{N,t} \\ \varepsilon_{\Delta c,t} \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} \mathbf{0} \\ a_j \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \\ a_1^j & a_2^j & \cdots & a_N^j & a_{N+1}^j \end{pmatrix}_{(N+1)\times(N+1)} \quad j = 1, \ldots, p,$$

where $\mu_{\Delta c}$ is the constant in the last equation of the VAR, i.e., the equation for $\Delta \ln(c_t)$, and $\sigma_{\Delta c}^2$ its conditional variance.

Here, $\Delta \ln(c_t)$ represents the common feature (factor), which obeys its own dynamics, transmitted to the rest of the system, $(r_{1,t}, r_{2,t}, \cdots, r_{N,t})'$, by the reduced-rank structure in $\begin{pmatrix} \mathbf{0} \\ a_j \end{pmatrix}$, $j = 1, \ldots, p$, and the contemporaneous relationships $r_{i,t} - \gamma \Delta \ln(c_t)$, embedded in $\mathbf{A}_0$. Given the properties of logged gross returns and the instantaneous growth rate of consumption, since $\mathbf{A}_0$ is a non-singular matrix, a standard reduced-form stationary $VAR(p)$ model for $\mathbf{X}_t$ is given by:

$$\begin{aligned} \mathbf{X}_t &= \mathbf{A}_0^{-1}\mathbf{c}_0 + \mathbf{A}_0^{-1} \begin{pmatrix} \mathbf{0} \\ a_1 \end{pmatrix} \mathbf{X}_{t-1} + \mathbf{A}_0^{-1} \begin{pmatrix} \mathbf{0} \\ a_2 \end{pmatrix} \mathbf{X}_{t-2} + \cdots \\ & \quad + \mathbf{A}_0^{-1} \begin{pmatrix} \mathbf{0} \\ a_p \end{pmatrix} \mathbf{X}_{t-p} + \mathbf{A}_0^{-1}\mathbf{v}_t, \qquad \text{or,} \end{aligned} \tag{42}$$

$$\mathbf{X}_t = \mathbf{c} + \mathbf{B}_1\mathbf{X}_{t-1} + \mathbf{B}_2\mathbf{X}_{t-2} + \cdots + \mathbf{B}_p\mathbf{X}_{t-p} + \boldsymbol{\eta}_t. \tag{43}$$

where $\mathbf{c} = \mathbf{A}_0^{-1}\mathbf{c}_0$, $\mathbf{B}_i = \mathbf{A}_0^{-1}\begin{pmatrix} \mathbf{0} \\ a_i \end{pmatrix}$, $i = 1, \ldots, p$, and $\boldsymbol{\eta}_t = \mathbf{A}_0^{-1}\mathbf{v}_t$. It is important to note that $\boldsymbol{\Sigma}_{\mathbf{v}}$ allows for heterogeneity in the variances of the terms in $\mathbf{v}_t$, which is explicitly taken into account in the definition of $\mathbf{c}_0$. Moreover, even if $\boldsymbol{\Sigma}_{\mathbf{v}}$ is a diagonal matrix, the reduced form variance-covariance matrix will not be diagonal, since $\mathbb{E}\left[\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right] = \left(\mathbf{A}_0^{-1}\right)\boldsymbol{\Sigma}_{\mathbf{v}}\left(\mathbf{A}_0^{-1}\right)'$. This can imply considerable cross-correlation for asset returns, stemming from the presence of the common component $\gamma\Delta\ln(c_t)$ in the returns of different assets.

Given an initial condition $\mathbf{X}_0$ for $\mathbf{X}_t$ in (43), set as $\mathbb{E}[\mathbf{X}_t]$, and a sufficient burning period for the dynamic system to operate, it is straightforward to simulate data for $\mathbf{X}_t$ that will satisfy the Asset-Pricing Equation (2) by construction. As a consequence, it will also obey the log-linear one-factor model derived above, i.e., equation (9).

To simplify matters in our simulation, we employed a $VAR(p)$ of order one, i.e., $p = 1$, using parameters consistent with those obtained in GMM estimation in Table 1 – that uses data at the quarterly frequency. The estimate for β was converted to the monthly frequency as $\beta = 0.999$, whereas $\gamma = 1.30$ was kept the same. We have set $\mathbb{E}[\Delta\ln(c_t)] = 0.0021$ at the monthly frequency, which yields an annual growth rate of 2.55%. We chose $\text{VAR}(\Delta\ln(c_t)) = 1.78E-5$.

After a burning period of 30 months, we simulate the system with $N = 1,000$ assets, and $T = 100$ monthly observations. In our preferred setting, the autoregressive coefficient for consumption growth is set to 0.3, i.e., $a_{N+1}^1 = 0.3$, whereas the other coefficients in the last line of $\begin{pmatrix} \mathbf{0} \\ a_1 \end{pmatrix}$ are all relatively small, selected to be close to 1^{-7}. Our preferred average (annual) volatility of the $1,000$ assets was set to 20%. We simulated the system 100 times.

In order to perform a robustness check to our results, we have also varied a_{N+1}^1 to be $a_{N+1}^1 = 0.1$ and $a_{N+1}^1 = 0.5$ in different sets of simulations. Although our preferred average (annual) volatility of the $1,000$ assets was set to 20%, we have also changed that for different simulations sets to be 15%, and 30%, where the latter allows for a lot of heterogeneity in (log) returns.

Figure 2 presents the average across simulations of the actual SDF in the CCAPM, $M_t = \beta\left(\frac{c_t}{c_{t-1}}\right)^{-\gamma}$, and our proposed estimate in this paper, $\widehat{M}_t = \frac{\overline{R}_t^G}{\frac{1}{T}\sum_{j=1}^T \overline{R}_j^G \overline{R}_j^A}$, for which we have used samples of $N = 1,000$ and $T = 100$ for each simulation. Average simulated and estimated values are extremely close, especially taking into account the tight SDF scale in the vertical axis for all the plots presented next in Figures 2, 3 and 4. The first shows our benchmark case, with $a_{N+1}^1 = 0.3$.

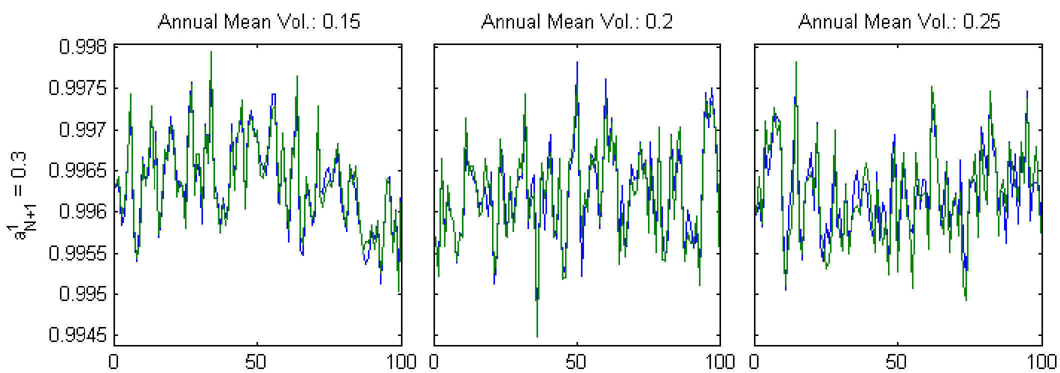

Figure 2: Average Simulated Stochastic Discount Factor of CCAPM (Blue) and Average Estimate in This Paper (Green). Number of Simulations = 100.

Figures 3 and 4 below also show simulation results for alternative settings, where the only change was on the autoregressive coefficient for consumption growth, from 0.3 to either 0.1 or 0.5. In them, we see very little difference between the average actual SDF, $M_t = \beta\left(\frac{c_t}{c_{t-1}}\right)^{-\gamma}$, and the average across simulations of $\widehat{M}_t = \frac{\overline{R}_t^G}{\frac{1}{T}\sum_{j=1}^T \overline{R}_j^G \overline{R}_j^A}$. The only striking difference across these figures is the degree of persistence observed in both actual and estimated SDF.

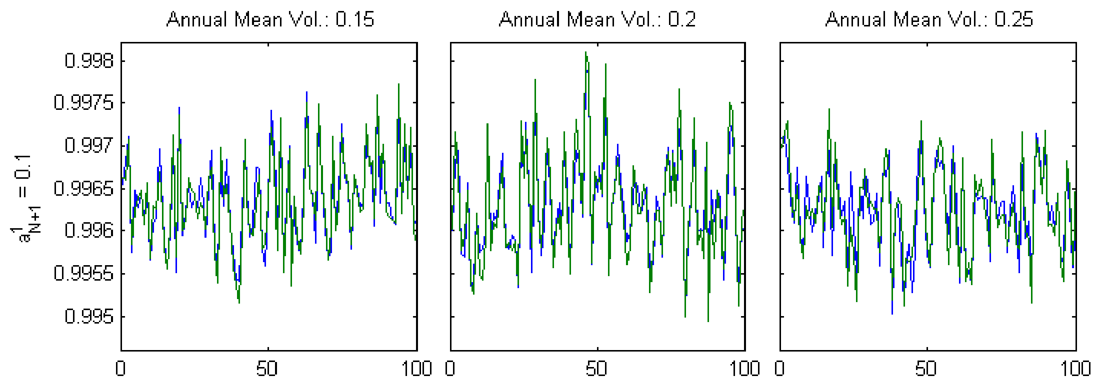

Figure 3: Average Simulated Stochastic Discount Factor of CCAPM (Blue) and Average Estimate in This Paper (Green). Number of Simulations = 100.

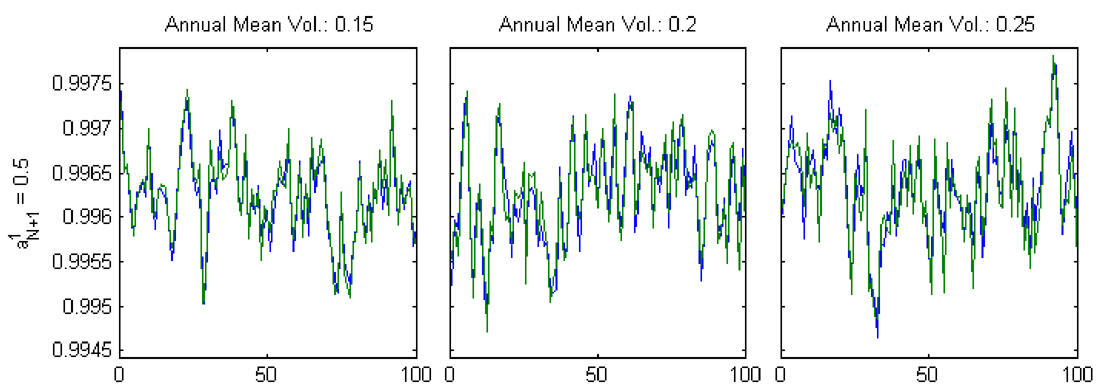

Figure 4: Average Simulated Stochastic Discount Factor of CCAPM (Blue) and Average Estimate in This Paper (Green). Number of Simulations = 100.

3.3 Factor Model Evaluation: A Panel-Data Analysis

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 2020:12 (T = 492 observations), collected for $N = 102,698$ assets, grouped in the following four categories: mutual funds $(68,085)$, stocks $(29,627)$, real-estate REITs $(1,000)$, and government bonds $(3,986)$.

All return data used in this exercise came from either from CRSP or Bloomberg. 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 collect 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, and also from Bloomberg.

The first step to perform our exercise is to compute $\widehat{M_t}$. It becomes clear immediately that we do not have a random sample of returns in the cross sectional dimension, since from the total number of $N = 102,698$ assets, 66% came from the Mutual-Fund category, 29% came from the Stocks category, 4% came from the Treasury-Bond category and only 1% came from the Real-Estate REITs category. However, based on the "Wealth and Asset Ownership" tables of 2004, provided by the U.S. Census Bureau, the approximate weights that each of these four categories should receive are as follows: Mutual Funds (10%), Stocks (10%), Government Bonds (20%), and Real Estate (60%)[6]. Therefore, we treated our sample as a stratified sample, weighting their respective returns using these approximate weights for each of the four asset categories (mutual funds, stocks, real estate, and government bonds). Hence, for each category, we first computed a version of $\widehat{M_t} = \frac{\overline{R}_t^G}{\frac{1}{T}\sum_{j=1}^{T}\overline{R}_j^G\overline{R}_j^A}$ for each category, later aggregating results using category weights.

[6] 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."

As a robustness check, we changed these weights (from 5 up to 20 percentage points for individual categories), computing again $\widehat{M_t}$. However, this produced virtually no change on reported results.

Using the baseline weights, our estimate of M_t has a mean of 0.9958 on a monthly basis, or 0.9504 on a yearly basis. The plot of $\widehat{M_t}$ follows below in Figure 5.

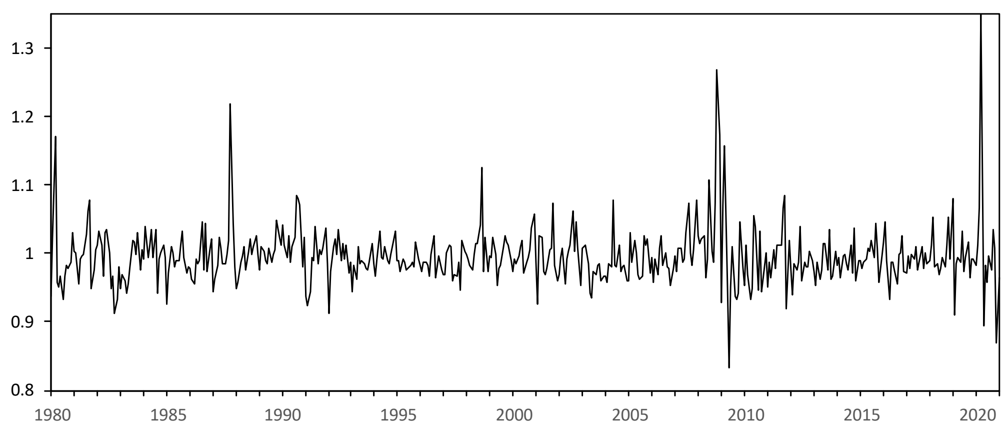

Figure 5: Stochastic Discount Factor Panel-Data Estimate

The affine beta-model for the log of asset returns, $r_{i,t}$, with the single factor m_t, and idiosyncratic error term $\xi_{i,t}$, is as follows:

$$r_{i,t} = -\left(\frac{1}{2}\sigma_i^2 - \delta_i \mathbb{E}\left(m_t\right)\right) + \beta_i m_t + \xi_{i,t}, \qquad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T. \tag{44}$$

Since the total number of time observations is $T = 492$, and the total number of assets is $N = 102,698$, we have too many parameters to estimate if we want to account for cross-sectional heterogeneity in the intercept and the slope of the affine model. Searching for a tractable model to estimate, we considered the *mixed-effect* panel-data model (also known as the *mixed linear* model) that takes into account individual heterogeneity in regression coefficients; see Cameron and Trivedi (2010) for a basic introduction in a panel-data context:

$$
\begin{aligned}
r_{i,t} &= -\left(\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}(m_t)\right) + \beta_i m_t + \xi_{i,t}, \qquad \text{where} \qquad (45)\\
\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}(m_t) &= \frac{1}{2}\sigma_0^2 + \delta_0 \mathbb{E}(m_t) + \mu_i, \qquad \mu_i \sim \mathcal{N}\left(0, \sigma_\mu^2\right),\\
\beta_i &= \beta_0 + \nu_i, \qquad \nu_i \sim \mathcal{N}\left(0, \sigma_\nu^2\right).
\end{aligned}
$$

Following the *mixed-effect* panel-data literature, we estimated (45) by maximum likelihood assuming a (zero-mean) Gaussian *random-effect* specification for μ_i and ν_i. Results are presented in Table 4, below.

$r_{it} = -\left(\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}(m_t)\right) + \beta_i m_t + \xi_{i,t}$ $\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}(m_t) = \frac{1}{2}\sigma_0^2 + \delta_0 \mathbb{E}(m_t) + \mu_i, \quad \mu_i \sim \mathcal{N}\left(0, \sigma_\mu^2\right)$ $\beta_i = \beta_0 + \nu_i, \quad \nu_i \sim \mathcal{N}\left(0, \sigma_\nu^2\right)$					
Weights on (T,S,RE,MF)	β_0 (SE)	$\frac{1}{2}\sigma_0^2 + \delta_0 \mathbb{E}(m_t)$ (SE)	σ_μ^2 (SE)	σ_ν^2 (SE)	$H_0 : \beta_0 = -1$ (P-Value)
(0.20, 0.15, 0.60, 0.05)	-1.010296 (0.0178636)	0.0028611 (0.0000582)	0.0020201 (0.0001047)	0.8895272 (0.020098)	0.33218 (0.56438)

Note: Robust standard errors in parentheses.

Table 4: *Mixed-effect* Model Estimation

In Table 4, β_0 – the mean of β_i – is -1.01 and it is statistically equal to -1 in testing. This suggests that $\delta_0 \mathbb{E}(m_t) \simeq 0$, and therefore that $\frac{1}{2}\sigma_0^2 + \delta_0 \mathbb{E}(m_t) \simeq \frac{1}{2}\sigma_0^2$. Recall that we have used $\overline{\beta} = \beta_0 = -1$ as our main identification assumption. Indeed, if our panel had a balanced sample, we should have obtained exactly -1.[7] The mean of $\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}(m_t)$ is also estimated with high precision and the same is true for the estimates of σ_μ^2 and of σ_ν^2.

We perform a sensitivity analysis of the results in Table 4 by varying the weights given to different classes of assets in calculating the SDF consistent estimate. These changes in weights range from 5 up to 20 percentage points for individual categories. Results are presented in Table 5. There is very little change in results – either quantitatively or qualitatively, which is reassuring.

[7] This is true if we observe m_t, which is not the case, since it is a latent variable. Indeed, we have a consistent estimator of it using a large sample. So, we are facing a well-known *generated-regressor problem*, which is mitigated since the sample we employ in estimation is *large*. Indeed, in this case, we can think of the test of $\beta_0 = -1$ as a specification test.

Next, we investigate whether or not the one-factor SDF affine model fits well the panel-data distribution of asset returns. For all the assets in the four categories, Mutual Funds, Stocks, Government Bonds, and Real Estate, we recover their individual intercepts and slopes and then forecast in sample the set of observed returns for all time periods. Indeed, we have used the best-linear-unbiased predictor (BLUP) for $r_{i,t}$ under the assumptions of the mixed-effect model (45).

Results are presented in Figure 6, below. Despite the fact that our sample includes the 1987 Black Monday episode, the burst of the Dotcom Bubble, the Great Recession, and the Covid-19 pandemic, the results show a good in-sample fit for our panel of returns, where fitted and actual values are aligned close to the 45 degree line, with the exception of a few outliers associated with those episodes.

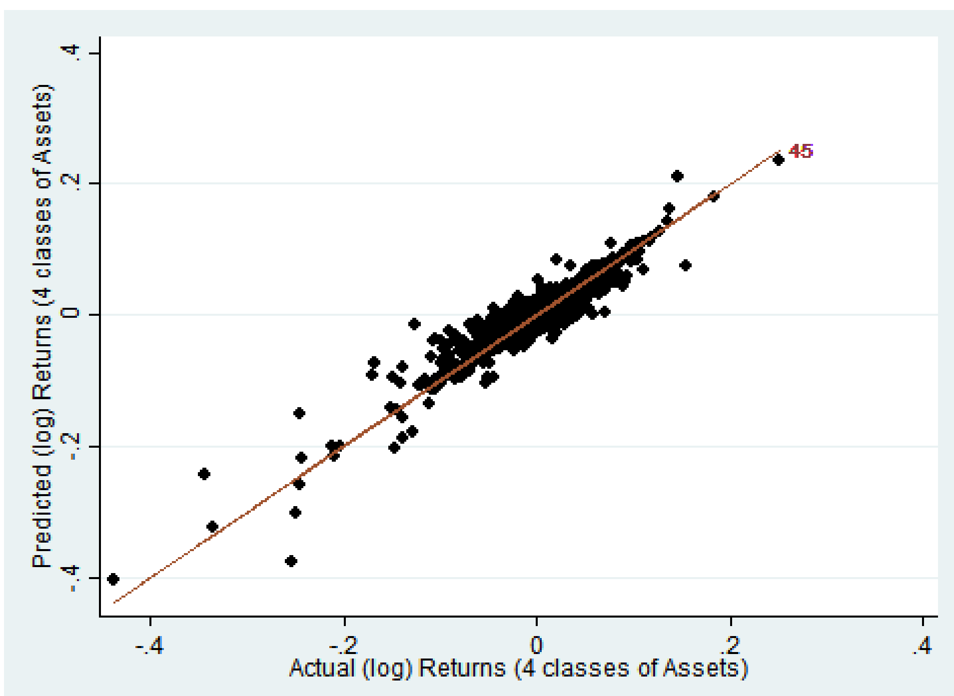

Figure 6: Predicted versus Actual Returns for all Asset Categories and Time Periods

$$r_{it} = -\left(\tfrac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}(m_t)\right) + \beta_i m_t + \xi_{i,t}$$
$$\tfrac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}(m_t) = \tfrac{1}{2}\sigma_0^2 + \delta_0 \mathbb{E}(m_t) + \mu_i, \quad \mu_i \sim \mathcal{N}\left(0, \sigma_\mu^2\right)$$
$$\beta_i = \beta_0 + \nu_i, \quad \nu_i \sim \mathcal{N}\left(0, \sigma_\nu^2\right)$$

Weights on				β_0	$\frac{1}{2}\sigma_0^2 + \delta_0 \mathbb{E}(m_t)$	$H_0 : \beta_0 = -1$
Treasuries	Stocks	REITS	Mutual Funds	(SE)	(SE)	(P-Value)
0.1	0.15	0.6	0.15	-1.09208 (0.23942)	0.005573 (0.00147)	0.14791 (0.70054)
0.1	0.15	0.7	0.05	-1.16708 (0.21726)	0.006201 (0.000903)	0.65127 (0.41966)
0.1	0.2	0.6	0.1	-1.12406 (0.21726)	0.006099 (0.001317)	0.32606 (0.56799)
0.15	0.1	0.6	0.15	-1.02994 (0.29832)	0.005175 (0.00147)	0.01007 (0.92005)
0.15	0.1	0.7	0.05	-1.10497 (0.28106)	0.005811 (0.000959)	0.13949 (0.70879)
0.15	0.2	0.5	0.15	-1.01882 (0.27748)	0.00558 (0.001678)	0.0046 (0.94594)
0.15	0.2	0.6	0.05	-1.09386 (0.26644)	0.006214 (0.00122)	0.12409 (0.72464)
0.2	0.1	0.6	0.1	-0.99979 (0.34218)	0.005301 (0.001318)	0.00000 (0.9995)
0.2	0.15	0.5	0.15	-0.95667 (0.34218)	0.005176 (0.001612)	0.01736 (0.89518)
0.2	0.15	0.6	0.05	-1.03172 (0.33046)	0.005815 (0.001219)	0.00922 (0.92352)
0.2	0.2	0.5	0.1	-0.98866 (0.32025)	0.005706 (0.001546)	0.00125 (0.97176)
0.25	0.1	0.5	0.15	-0.89456 (0.37637)	0.004777 (0.001579)	0.07849 (0.77935)
0.25	0.1	0.6	0.05	-0.96962 (0.38898)	0.005421 (0.001267)	0.0061 (0.93775)
0.25	0.15	0.5	0.1	-0.92653 (0.37051)	0.005302 (0.001503)	0.03932 (0.84281)
0.25	0.2	0.5	0.05	-0.9585 (0.36587)	0.005826 (0.001519)	0.01287 (0.90969)

Note: Robust standard errors in parentheses. Variance estimates omitted for the sake of space.

Table 5: *Mixed-effect* Model Estimation – Sensitivity Analysis

4 Conclusion

In this paper, we propose a no-arbitrage approach to asset pricing with the following ingredients. First, based on no-arbitrage, we derive a one-factor model for the logarithm of asset returns, where the single factor is the logarithm of a valid stochastic discount factor. We interpret the factor as containing all the pervasive elements of (log) asset returns, allowing a pervasive-idiosyncratic decomposition of these returns. Second, based on this one-factor model, we derive a consistent estimator of a valid SDF in a panel-data framework, when the number of assets and of time periods increase without bounds. The asymptotic character of this SDF estimator is opposed to standard small-sample alternatives where it is hard to interpret empirical results since these often change when different groups of assets are used in estimation. From a theoretical perspective, asymptotic estimates are immune to this problem. This increases its potential application in empirical studies, especially in the big data era.

Our consistent estimator of a valid stochastic discount factor (SDF) exploits the cross-sectional variation of returns to propose an asymptotically biased estimate of a valid SDF. Our key identifications assumption is used to eliminate the bias and guarantees that the proposed estimator is within the no-arbitrage class. Our SDF estimator depends exclusively on appropriate averages of asset returns (geometric and arithmetic), 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.

The techniques discussed in this paper were applied to three issues in macroeconomics and finance. In the first application, we used quarterly data of U.S.$ real returns from 1977:Q2 to 2019:Q2, comprising $T = 169$ time periods and thousands of assets worldwide, to examine whether popular preference specifications used in macroeconomics and finance fit the data. Our SDF estimator – $\widehat{M_t}$ – is close to unity most of the time and bounded by the interval $[0.85, 1.25]$, with an equivalent average quarterly discount factor of 0.99. When we examined the appropriateness of different functional forms to represent preferences (Power Utility, External Habit and Kreps-Porteus), we concluded that none of these standard preference representations are rejected by the data. However, after testing for exclusion restrictions, we conclude that the External-Habit and the Kreps-Porteus specifications are our preferred specifications, which concurs with the current dominant view in the macro-finance literature.

In our second application, a mid-size simulation exercise is implemented with a no-arbitrage dynamic (consumption) capital asset-pricing model using 100 time-series observations and and $1,000$ cross-sectional observations. It entails heterogeneity in the first and second moments of asset returns. Results show that, on average, our consistent SDF estimator is very close to the actual SDF, despite the fact that we employ a small sample relative to the asymptotic framework.

In our third application, we try to approximate the asymptotic environment with monthly U.S. time-series return data from 1980:1 through 2020:12 ($T = 492$ observations), collected for $N = 102,698$ assets. Our estimate of M_t has an average of 0.99 on a monthly basis. We employed the *mixed-effect* panel-data model to assess the fit of our one-factor model to the data. Despite the fact that our sample includes the 1987 Black Monday episode, the burst of the Dotcom Bubble, the Great Recession, and the recent Covid-19 pandemic, the empirical results show a good in-sample fit for our panel of returns, where fitted and actual values are aligned close to the 45 degree line.

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] Almeida, C., Ardison, K. and Garcia. R. (2020), "Nonparametric assessment of hedge fund performance," *Journal of Econometrics*, vol. 214(2), pp. 349-378.

[3] Araujo, F., J.V. Issler and M. Fernandes (2005): "Estimating the Stochastic Discount Factor without a Utility Function," Working Paper: Brazilian School of Economics and Finance FGV EPGE, Getulio Vargas Foundation.

[4] Araujo, F., and J.V. Issler (2011): "A Stochastic Discount Factor Approach to Asset Pricing using Panel Data Asymptotics," Working Paper: Brazilian School of Economics and Finance FGV EPGE, Getulio Vargas Foundation.

[5] 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," *Journal of Econometrics*, vol. 164 (1), pp. 116-129.

[6] 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.

[7] 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.

[8] Bai, J., 2009, "Panel data models with interactive fixed effects", *Econometrica*, Vol. 77, pp. 1229–1279.

[9] Bai, J. and Serena Ng, 2002, "Determining the number of factors in approximate factor models", *Econometrica*, vol. 70, pp. 191–221.

[10] Bai, J. and Serena Ng, 2004, "A PANIC attack on unit roots and cointegration," *Econometrica*, Vol. 72, pp. 1127–1177.

[11] Bansal, Ravi, and Amir Yaron, 2004, "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *The Journal of Finance*, Vol. 59, pp. 1481-1509.

[12] 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.

[13] Cameron, A.C., and Trivedi, P.K. (2010), "*Microeconometrics Using Stata,*" Revised Edition, Stata Press.

[14] Campbell, J.Y. and John H. Cochrane, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, Vol. 107, pp. 205-251.

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

[16] Campello, M., Galvão, A.F., and Juhl, T. (2019), "Testing for slope heterogeneity bias in panel data models," *Journal of Business and Economic Statistics*, 37 (4), 749-760.

[17] Christensen, T.M. (2017), "Nonparametric Stochastic Discount Factor Decomposition," *Econometrica*, Vol. 85(5), pp. 1501-1536.

[18] Dhrymes, P.J. (1974), "*Econometrics: Statistical Foundations and Applications,*" 2nd. Edition, New York: Springer-Verlag.

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

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

[21] 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.

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

[23] 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.

[24] Fama, E.F. and French, K.R. (1996), "Multifactor Explanations of Asset Pricing Anomalies," *The Journal of Finance*, 51, 55-84.

[25] Fama, E.F. and French, K.R. (2015), "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*, 116, 1-22.

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

[27] 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.

[28] Forni, M., Hallin, M, Lippi, M. and Reichlin, L., 2005, “The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting”, *Journal of the American Statistical Association*, Vol. 100 pp.830-840.

[29] 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.

[30] Giglio, S., Kelly, B., and Xiu, D. (2021), “Factor Models, Machine Learning, and Asset Pricing,” Working Paper: Yale University.

[31] Gomes, F.A.R., and Issler, J.V. (2017), “Testing Consumption Optimality using Aggregate Data,” *Macroeconomic Dynamics*, Vol. 21(5), pp. 1119-1140.

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

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

[34] 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.

[35] Hansen, L.P. and J.A. Scheinkman (2009), “Long-Term Risk: An Operator Approach,” *Econometrica*, 77(1), pp. 177 - 234.

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

[37] 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.

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

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

[40] Harvey, C.R., Liu, Y., and Zhu, H. (2016), "...and the Cross-Section of Expected Returns". *Review of Financial Studies*, vol. 29(1), pp. 5-68.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] Kozak, S., Nagel, S., and Santosh, S. (2019), "Shrinking the Cross-Section," Working Paper: University of Chicago.

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

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

[48] 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.

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

[50] Phillips, P.C.B., Moon, H.R., 1999. "Linear Regression Limit Theory for Nonstationary Panel Data." *Econometrica* 67, 1057-1111.

[51] Pukthuanthong, K., and R. Roll (2015): "Agnostic Tests of Stochastic Discount Factor Theory," Working Paper, California Institute of Technology, # 1405, pp. 1-56.

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

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

[54] Rubinstein, M. (1976), "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics*, vol. 7(2), pp. 407-425.

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

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

[57] 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.

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

[59] 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.

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

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

[62] 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.

A Appendix: A General Consistency Proof for our SDF Estimator

To construct a consistent estimator for M_t under general assumptions, we consider a second-order Taylor Expansion of the exponential function around x, with increment h, as follows:

$$e^{x+h} = e^x + he^x + \frac{h^2 e^{x+\lambda(h)\cdot h}}{2}, \tag{46}$$

$$\text{where } \lambda(h) : \mathbb{R} \to (0,1) \,. \tag{47}$$

It is important to stress that (46) is an exact relationship and not an approximation. This is due to the nature of the function $\lambda(h) : \mathbb{R} \to (0, 1)$, which maps into the *open* unit interval. Thus, the last term is evaluated between x and $x+h$, making (46) to hold *exactly.*

For the expansion of a generic function, $\lambda(\cdot)$ would depend on x and h. However, the exponential function has a special property: dividing (46) by e^x yields:

$$e^h = 1 + h + \frac{h^2 e^{\lambda(h)\cdot h}}{2}, \tag{48}$$

showing that (48) 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 (48) with the Pricing Equation (2), we let $h = \ln(M_t R_{i,t})$ in (48) to obtain:

$$M_t R_{i,t} = 1 + \ln(M_t R_{i,t}) + \frac{\left[\ln(M_t R_{i,t})\right]^2 e^{\lambda(\ln(M_t R_{i,t}))\cdot\ln(M_t R_{i,t})}}{2}, \tag{49}$$

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 (49):

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

Taking the conditional expectation of both sides of (49) gives:

$$\begin{aligned} \mathbb{E}_{t-1}\left(M_t R_{i,t}\right) &= 1 + \mathbb{E}_{t-1}\left(\ln(M_t R_{i,t})\right) + \mathbb{E}_{t-1}\left(z_{i,t}\right), \text{ or,} & (50) \\ 0 &= \mathbb{E}_{t-1}\left(M_t R_{i,t}\right) - 1 = \mathbb{E}_{t-1}\left(\ln(M_t R_{i,t})\right) + \mathbb{E}_{t-1}\left(z_{i,t}\right), & (51) \end{aligned}$$

where (51) is a direct consequence of the Asset-Pricing Equation (2), since its left-hand side cancels out yielding:

$$\mathbb{E}_{t-1}\left(z_{i,t}\right) = -\mathbb{E}_{t-1}\left\{\ln(M_t R_{i,t})\right\}. \tag{52}$$

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

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

$$\ln(M_t\mathbf{R}_t) = \mathbb{E}_{t-1}\{\ln(M_t\mathbf{R}_t)\} + \boldsymbol{\varepsilon}_t$$
$$= -\boldsymbol{\gamma}^2_{t|t-1} + \boldsymbol{\varepsilon}_t. \tag{53}$$

Denoting by $\mathbf{r}_t = \ln(\mathbf{R}_t)$, which elements are denoted by $r_{i,t} = \ln(R_{i,t})$, and by $m_t = \ln(M_t)$, (53) yields the following system of equations:

$$r_{i,t} = -m_t - \gamma^2_{i,t|t-1} + \varepsilon_{i,t}, \qquad i = 1, 2, \ldots, N. \tag{54}$$

The system in equation (54) generalizes the result in equation (3), where Gaussianity was imposed for the sake of simplicity. The first main difference here is that the term $\gamma^2_{i,t|t-1}$ is not a constant and it has serial correlation. The second main difference is that $\varepsilon_{i,t}$ is not Gaussian. The system (54) shows that the (log of the) SDF is a common feature, in the sense of Engle and Kozicki (1993), of all (logged) asset returns[8].

The sources of cross-sectional variation in every equation of the system (54) are $\varepsilon_{i,t}$ and $\gamma^2_{i,t|t-1}$. However, as we show next, the terms $\gamma^2_{i,t|t-1}$ are a linear function of the lagged $\varepsilon_{i,t}$, tying the cross-sectional variation in (54) ultimately to $\varepsilon_{i,t}$.

From Assumption 2, $\{\ln(M_t)\}$ and $\{\ln(\mathbf{R}_t)\}$ are joint covariance-stationary processes with finite first and second moments. This yields covariance-stationarity for $\{\ln(M_t\mathbf{R}_t)\}$ as well. We then apply Wold's Theorem to write the individual Wold representations as:

$$\ln(M_t R_{i,t}) = m_t + r_{i,t} = \mu_i + \sum_{j=0}^{\infty} b_{i,j}\varepsilon_{i,t-j}, \qquad i = 1, 2, \ldots, N, \tag{55}$$

where, for all i, $b_{i,0} = 1$, $|\mu_i| < \infty$, $\sum_{j=0}^{\infty} b^2_{i,j} < \infty$, and $\varepsilon_{i,t}$ is a white-noise process. Using (52), in light of (55), leads to:

$$\gamma^2_i \equiv \mathbb{E}(z_{i,t}) = -\mathbb{E}\{\ln(M_t R_{i,t})\} = -\mu_i, \tag{56}$$

which is well defined and time-invariant under Assumption 2. Taking conditional expectations $\mathbb{E}_{t-1}(\cdot)$ of (55), allows computing:

$$\gamma^2_{i,t|t-1} = \mathbb{E}_{t-1}(z_{i,t}) = -\mathbb{E}_{t-1}\{\ln(M_t R_{i,t})\} = -\gamma^2_i - \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}, \qquad i = 1, 2, \ldots, N,$$

leading to the following system, once we consider (54):

$$r_{i,t} = -m_t - \gamma^2_i + \varepsilon_{i,t} - \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}, \qquad i = 1, 2, \ldots, N. \tag{57}$$

[8] 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 in the context of a VAR model. The concept of common features was proposed 10 years later by Engle and Kozicki (1993).

Equation (57) generalizes equation (3), where now the error term, $\varepsilon_{i,t}-\sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}$, has serial correlation, capturing the dynamic nature of $\gamma^2_{i,t|t-1}$. As before, we will proceed performing an OLS projection of:

$$\varepsilon_{i,t}-\sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}, \qquad i=1,2,\ldots,N, \tag{58}$$

onto $m_t-\mathbb{E}(m_t)$ in order to obtain a diversifiable error term in the cross-sectional dimension. As shown below, despite the dynamic nature of (58), the OLS projection will make it an idiosyncratic error term. To advance, we need to introduce some notation to be able to deal with multivariate wold representations in the context of joint weak-stationarity.

From joint weak-stationarity of $X_t=(m_t, r_{1,t}, r_{2,t}, \cdots, r_{N,t})'$, we get a vector wold representation as follows:

$$\begin{aligned}\begin{pmatrix} m_t \\ r_{1,t} \\ r_{2,t} \\ \vdots \\ r_{N,t}\end{pmatrix} &= \begin{pmatrix} \mathbb{E}(m_t) \\ \mathbb{E}(r_{1,t}) \\ \mathbb{E}(r_{2,t}) \\ \vdots \\ \mathbb{E}(r_{N,t})\end{pmatrix} + \sum_{j=0}^{\infty} \mathbf{\Psi}_j \begin{pmatrix} \varepsilon^m_{t-j} \\ \mu_{1,t-j} \\ \mu_{2,t-j} \\ \vdots \\ \mu_{N,t-j}\end{pmatrix}, \qquad (59)\\ \text{with } \mathbf{\Psi}_0 &= \mathbf{I}_{N+1}, \text{ and,}\end{aligned}$$

$$\mathbb{E}\left[\begin{pmatrix} \varepsilon^m_{t} \\ \mu_{1,t} \\ \mu_{2,t} \\ \vdots \\ \mu_{N,t}\end{pmatrix}\begin{pmatrix} \varepsilon^m_{t-j} \\ \mu_{1,t-j} \\ \mu_{2,t-j} \\ \vdots \\ \mu_{N,t-j}\end{pmatrix}'\right] = \begin{cases} \Omega, & \text{for } j=0 \\ \mathbf{0}, & \text{for } j\neq 0\end{cases}, \tag{60}$$

which we can rewrite conveniently as a system of individual wold representations:

$$m_t = \mathbb{E}(m_t)+\sum_{j=0}^{\infty} c_j \varepsilon^m_{t-j}, \text{ with } c_0=1, \text{ and,} \tag{61}$$

$$r_{i,t} = \mathbb{E}(r_{i,t})+\sum_{j=0}^{\infty} d_{i,j}\mu_{i,t-j}, \text{ with } d_{i,0}=1, \quad i=1,2,\cdots,N. \tag{62}$$

To be able to arrive at the individual wold representations in (61) and (62) from (59), we have to re-scale the lagged terms on the right-hand side of (59). Indeed, for every equation in (59), these terms are linear combinations of *lagged* $\left(\varepsilon^m_t, \mu_{1,t}, \mu_{2,t}, \cdots, \mu_{N,t}\right)'$. To obtain the convenient form of (61) and (62) in terms of their respective one-step innovations $\left(\varepsilon^m_t, \mu_{1,t}, \mu_{2,t}, \cdots, \mu_{N,t}\right)'$, one has to re-scale error terms as follows: the $\mu_{i,t}$s are re-scaled in terms of ε^m_t as $\varepsilon^m_t=\frac{\sigma_{\varepsilon^m}}{\sigma_{\mu_i}}\mu_{i,t}$, and the $\mu_{k,t}$s are re-scaled in terms of $\mu_{i,t}$ as $\mu_{i,t}=\frac{\sigma_{\mu_i}}{\sigma_{\mu k}}\mu_{k,t}$, where σ_{μ_i}, σ_{μ_k} and σ_{ε^m} are the standard deviations of $\mu_{i,t}$, $\mu_{k,t}$ and ε^m_t, respectively. To maintain the equality in (59), we have to adjust the moving-average coefficients in the matrices $\mathbf{\Psi}_l$, $l = 1, 2, \cdots$, using, respectively, the reciprocals of the scales used above.

It is important to stress that, due to the nature of (60), the following covariance structure for the errors ε_t^m, $\mu_{i,t-j}$, and $\mu_{k,t-j}$ applies:

$$\mathbb{E}\left[\varepsilon_t^m \mu_{i,t-j}\right] = \begin{cases} \gamma_{\mu_i \varepsilon^m}, & \text{for } j = 0 \\ 0, & \text{for } j \neq 0 \end{cases} \quad \text{and } \mathbb{E}\left[\mu_{i,t}\mu_{k,t-j}\right] = \begin{cases} \sigma^2_{\mu_i}, & \text{for } k = i \text{ and } j = 0 \\ \gamma_{\mu_i\mu_k}, & \text{for } k \neq i \text{ and } j = 0 \\ 0, & \text{for } j \neq 0 \end{cases}, \tag{63}$$

where $\gamma_{\mu_i \varepsilon^m}$ and $\gamma_{\mu_i\mu_k}$ are cross covariances between ε_t^m and $\mu_{i,t}$, and $\mu_{i,t}$ and $\mu_{k,t}$, respectively, while $\sigma^2_{\mu_i}$ is the variance of $\mu_{i,t}$.

With the setup in (61), (62), and (63), given joint stationarity of $X_t = (m_t, r_{1,t}, r_{2,t}, \cdots, r_{N,t})'$, we can rewrite (58) for every $i = 1, 2, \cdots, N$, as follows:

$$\begin{aligned} \varepsilon_{i,t} - \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j} &= \left(\mu_{i,t} + \varepsilon_t^m\right) - \sum_{j=1}^{\infty} c_j \varepsilon_{t-j}^m - \sum_{j=1}^{\infty} d_{i,j}\mu_{i,t-j} \\ &= \left(\mu_{i,t} + \varepsilon_t^m\right) - \sum_{j=1}^{\infty} c_j \varepsilon_{t-j}^m - \sum_{j=1}^{\infty} d_{i,j}\mu_{i,t-j} - \left(\mu_{i,t} + \varepsilon_t^m\right) + \left(\mu_{i,t} + \varepsilon_t^m\right) \\ &= 2\left(\mu_{i,t} + \varepsilon_t^m\right) - \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}^m - \sum_{j=0}^{\infty} d_{i,j}\mu_{i,t-j} \\ &= 2\left(\mu_{i,t} + \varepsilon_t^m\right) - \left(m_t - \mathbb{E}(m_t)\right) - \left(r_{i,t} - \mathbb{E}(r_{i,t})\right). \end{aligned}$$

It is important to emphasize that, although $\varepsilon_{i,t} - \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}$ has an infinite lag structure, perhaps suggesting that we need to control for an infinite number of factors (or of lags of m_t), we were able to rewrite it as function of m_t alone: both $(m_t - \mathbb{E}(m_t))$ and $(r_{i,t} - \mathbb{E}(r_{i,t}))$ are functions of m_t alone, which is also the case of $\left(\mu_{i,t} + \varepsilon_t^m\right)$, which is certainly a function of m_t, but is orthogonal to its lags. Hence controlling for pervasive factors only requires to compute the OLS projection of $\varepsilon_{i,t} - \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}$ onto $m_t - \mathbb{E}(m_t)$. This is very important, since it fits exactly the same procedure done in Section 2.1 – equation (8).

Using (63), first, we compute:

$$\begin{aligned} \text{COV}\left[m_t, \left(\varepsilon_{i,t} - \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}\right)\right] &= \mathbb{E}\left\{\left(m_t - \mathbb{E}(m_t)\right)\left[\begin{array}{c} 2\left(\mu_{i,t} + \varepsilon_t^m\right) - \\ \left(m_t - \mathbb{E}(m_t)\right) - \left(r_{i,t} - \mathbb{E}(r_{i,t})\right) \end{array}\right]\right\} \\ &= 2\left(\gamma_{\mu_i \varepsilon^m} + \text{VAR}\left(\varepsilon_t^m\right)\right) - \text{VAR}\left(m_t\right) - \text{COV}\left(m_t, r_{i,t}\right). \end{aligned}$$

Based on this result, we can compute the OLS projection coefficient of $\varepsilon_{i,t} - \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}$

onto $m_t - \mathbb{E}(m_t)$ as:

$$\begin{aligned}\delta_{i,m} &= \frac{\text{COV}\left(m_t, \varepsilon_{i,t} - \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j}\right)}{\text{VAR}(m_t)} \\ &= \frac{2\left(\gamma_{\mu_i\varepsilon^m} + \text{VAR}(\varepsilon_t^m)\right) - \text{VAR}(m_t) - \text{COV}(m_t, r_{i,t})}{\text{VAR}(m_t)},\end{aligned}$$

and finally write the whole projection for every $i = 1, 2, \cdots, N$, as:

$$\varepsilon_{i,t} - \sum_{j=1}^{\infty} b_{i,j}\varepsilon_{i,t-j} = \delta_{i,m}(m_t - \mathbb{E}(m_t)) + \xi_{i,t},$$

where now $\xi_{i,t}$ has serial correlation but is devoid of any pervasive effects since it is orthogonal to $m_t - \mathbb{E}(m_t)$. It also has a zero unconditional mean.

As before, we can now rewrite (57) as a factor model:

$$\begin{aligned}r_{i,t} &= -\left(\gamma_i^2 + \delta_{i,m}\mathbb{E}(m_t)\right) + \beta_i m_t + \xi_{i,t}, \qquad i = 1, 2, \ldots, N, \text{ where,} \\ \beta_i &= \delta_{i,m} - 1,\end{aligned} \tag{64}$$

for which:

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \xi_{i,t} = 0.$$

Since the setup in (64) is identical to that of (9), we can repeat all the steps in Section 2.1 leading to *Proposition 1*. Hence, we can re-state our main result here.[9]

Theorem 2 *Under Assumptions 1-3, as $N, T \to \infty$, with N diverging first and T diverging later, the realization of the SDF at time t, denoted by M_t, can be consistently estimated using:*

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

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.

It is important to note that $\widehat{M_t}$ is a function of N and T. The denominator explicitly shows that it depends on T and dependence on N is implicit since $\overline{R}_t^G$ and $\overline{R}_t^A$ are respectively geometric and arithmetic averages in the cross-sectional dimension. The only reason why we do not explicitly state its dependence on N and T is to avoid a cumbersome notation for $\widehat{M_t}$.

[9] Note, however, that the role previsously played by $\sigma_i^2/2$ is now played by γ_i^2, but otherwise all else is the same.