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

Fabio Araujo† Antonio F. Galvao‡ João Victor Issler§

September 21, 2025

Abstract

This paper proposes a novel approach including identification and consistent estimation of valid stochastic discount factors (or pricing kernels). First, we investigate under which conditions we can establish identification for time-varying valid stochastic discount factors (SDFs) treated as a latent process. We show that a sufficient condition for identification is that the cross-sectional mean of pricing errors of a logarithm representation are zero. Second, based on the identification strategy, we propose a no-arbitrage estimator for a valid SDF. Third, we derive the asymptotic properties of this estimator, namely, its consistency and asymptotic normality, when the number of assets and time periods increase without bounds. Fourth, we derive a no-arbitrage one-factor model for the logarithm of asset returns, where the single factor is the logarithm of a valid SDF, containing all the common elements of (log) asset returns. An important by-product of this one-factor representation is that it implies a testable condition of our key identification assumption. Empirically, based on this novel setup, we first estimate a multi-effect linear regression model with intercept and slope heterogeneity allowing us to test our key identifying assumption, which was not rejected by the data. Second, we present a Fama-MacBeth experiment forecasting the cross-section of Fama-French portfolios in sample and out of sample. All in all, results are promising for this novel approach.

*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, and Walter Novaes. Special thanks are due to Caio Almeida, Robert Engle, Marcelo Fernandes, René Garcia, Lars Hansen, Tiago Martins, Leornardo Moog, Cristine Campos de Xavier Pinto, Raul Riva, Bernard Salanié, José Alexandre Scheinkman, and Farshid Vahid. We also thank José Gil Ferreira Vieira Filho, Rafael Burjack, Guilherme Kira, Luisa Curcio, and Pedro Guerra 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, FAPERJ, and the CAPES-PrInt program for financial support. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This paper supersedes all versions of Araujo, Issler and Fernandes (2005) and Araujo and Issler (2011), both cited in our references below.

†Programa Real Digital, Brazilian Central Bank. email: fabio.araujo@bcb.gov.br

‡Corresponding author. Department of Economics, Michigan State University, 486 W. Circle Drive 110 Marshall-Adams Hall East Lansing, MI 48824. email: agalvao@msu.edu

§Brazilian School of Economics and Finance FGV EPGE, Getulio Vargas Foundation. email: joao.issler@fgv.br

Keywords: Stochastic Discount Factor, No-Arbitrage, Panel Data, Common Features.
JEL Codes: C32, C33, E21, E44, G12.

1 Introduction

In this paper, we propose identification and consistent estimation of stochastic discount factors (or pricing kernels) by taking 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 valid 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, identification is established for time-varying valid stochastic discount factors $\{M_t\}_{t=1}^{\infty}$ treated as a latent factor. A sufficient condition for identification is that the cross-sectional mean of pricing errors of a logarithm representation are zero. In this context, based on economic theory – no-arbitrage – information contained mainly in the cross-section of priced returns is pivotal. Indeed, a valid SDF is identified as a simple non-linear function of the cross-section arithmetic and geometric means of priced returns. The core idea behind our identification strategy is explained by employing a simplified version of the asset-pricing equation as in Hansen and Singleton (1983): no-arbitrage, and a joint log-Normal specification for returns and a valid SDF. However, in the Appendix, we show that our approach can be replicated using a general Taylor Expansion of the Asset-Pricing Equation combined with joint weak-stationarity for returns and a valid SDF.

Second, based on the identification result, we propose a novel estimator for a valid SDF in a panel-data framework. The estimator entails sample counterparts of the population

moments that identify a valid SDF. We establish the asymptotic properties of the estimator – namely, its consistency and asymptotic normality when both the number of assets and of time periods increase without bounds.[1] 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 depending on the sample used 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.

Third, 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 common elements of (log) asset returns, allowing a common-idiosyncratic decomposition of them, where the intercept and slope parameters contain heterogeneous responses. This representation has an important empirical implication: it allows testing empirically the identification condition. We show that the cross-section average of the heterogeneous slopes must be equal to minus one under the identification condition, which allows for a simple Wald test based on estimated slopes to verify its validity.

A benefit of our approach is that we are able to study important issues in finance and in macroeconomics without the need to characterize preferences, or to use consumption data, in a setting similar to that of Hansen and Jagannathan (1991, 1997). As is well-known, consumption-based estimates generated important puzzles, such as the equity-premium puzzle, excess sensitivity (excess smoothness) in consumption, among others; see Flavin (1981), Hansen and Singleton (1982, 1983, 1984) Mehra and Prescott (1985), and Campbell and Deaton (1989).

We perform three empirical experiments here. On our first application, we estimate our proposed factor model using a multi-effect linear regression setup. We try to approxi-

[1] Phillips and Moon (1999, 2000) discuss and develop a framework for analysis of panel data asymptotics with double indexed process for both sequential and joint limit theories.

mate 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 ($3,986$). The *mixed-effect* panel-data model takes into account individual heterogeneity in regression coefficients. When we tested whether the factor-model restriction is consistent with our identifying assumption using return data, Wald-test results did not reject the null at standard significance levels – a result that survives robustness tests, including that of a potential generated-regressor problem.

In our second application, extracted from Issler, Martins and Moog (2023) and from Issler and Moog (2025), we asked whether our consistent SDF estimator can price well the cross-sectional distribution of the 25 Fama-French portfolios using Fama-MacBeth regressions. Indeed, in-sample results for monthly data show that our one-factor model beats popular alternatives by a reasonable margin: it achieves a $\overline{R}^2 = 0.95$ pricing cross-sectional returns, while the two closest factor-model approaches achieve, respectively, a $\overline{R}^2 = 0.71$ and a $\overline{R}^2 = 0.60$. In-sample results for quarterly data are similar. We also employed a testing procedure where M_t estimates are asked to price well-known anomalies out of sample. Our proposed M_t estimate produces the highest out-of-sample $\overline{R}^2$ for 76% of the anomalies tested.

Our third application, included in the Appendix, is a mid-sized Monte-Carlo simulation of a no-arbitrage dynamic consumption CAPM with 100 time-series and $1,000$ cross-sectional observations. Results show that our consistent SDF estimator closely matches the actual SDF in small samples. Also, biases in the mean and variance estimates are negligible.

Our approach is related to research 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 Gaglianone and Issler (2023), 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), Christensen (2017), Almeida, Ardison, and Garcia (2020), Pukthuanthong, Roll, and Wang (2021), Sandulescu, Trojani, and Vedolin (2021), and Korsaye, Quaini, and Trojani (2021). It is also related respectively to work on *common factors* in macroeconomics; see Stock and Watson (2002), Forni et al. (2000, 2005), Bai and Ng (2002, 2004), and Bai (2009).

This paper is also connected to the common-factor literature modeling the cross-section of asset returns, based on the work of Fama and MacBeth (1973), Fama and French (1992, 1993, 1996, 2015), Lettau and Ludvigson (2001), Campello, Galvao and Juhl (2019), Feng, Giglio, and Xiu (2020), Kozak, Nagel, and Santosh (2020), Giglio, Kelly, and Xiu (2022), Kelly and Xiu (2023), and Chen, Pelger, and Zhu (2024). 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).

The remainder of this paper is organized as follows. The next section presents basic theoretical results, our estimation techniques, and a discussion of our main results. Section 3 shows results for several empirical tests in macroeconomics and finance using our estimator. Section 4 concludes. The Appendix contains the proofs of the main results, and a re-statement of them without the use of stringent assumptions used in the simplified version of the one-factor model.

A Word on Notation. Since we are dealing with panel data, we can compute sample means, variances, and covariances across time, across assets, or across both. Conditional means across time use the operator $\mathbb{E}_{t-1}[\cdot] = \mathbb{E}[\cdot|\mathcal{I}_{t-1}]$, where $\mathcal{I}_{t-1}$ represents the in-

formation set used by agents in forming expectations. Unconditional means across time and assets, or just across time, use the operator $\mathbb{E}[\cdot]$. Unconditional variances and covariances across time, or across time and assets, use the operator VAR$(\cdot)$ and COV$(\cdot)$, respectively. We specify which case we are referring to on an individual basis along the text. Unconditional means across assets alone use the operator $\mathbb{E}^i[\cdot]$. For stationary and ergodic series $X_{i,t}$, that converge in probability, we employ the following notation: $\frac{1}{N}\sum_{i=1}^{N} X_{i,t} \xrightarrow{p} \mathbb{E}^i[X_{i,t}]$, $\frac{1}{T}\sum_{i=1}^{N} X_{i,t} \xrightarrow{p} \mathbb{E}[X_{i,t}]$, and $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{i,t} \xrightarrow{p} \mathbb{E}[X_{i,t}]$. Conditions for convergence in probability are given in detail below. The number of assets is denoted by N and the number of time periods is denoted by $T = T_N$ and may depend on N. In what follows, for simplicity, we omit the subscript N of T_N. We also define $\mathbb{E}^i[X_{i,t}] \equiv \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} X_{i,t}\right]$.

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, 1997) present a general asset pricing framework based on stochastic discount factors (SDF), characterized by the following Asset-Pricing Equation, which is our starting point:[2]

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

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

where M_t is a valid SDF, $p_{i,t}$ is the price of asset i at time t, $x_{i,t}$ its payoff, $R_{i,t} = x_{i,t}/p_{i,t-1}$ its gross return, and N the number of assets. We now state our key assumptions:

Assumption 1 (No Arbitrage): Asset prices reflect no arbitrage opportunities (Ross, 1976) for all $t = 1, 2, ..., T$.

[2]See also Rubinstein (1976) and Ross (1976).

Assumption 2 (Stationarity and Ergodicity): Let $\{R_{i,t}\}$ be a cross-sectionally independent triangular array and $\mathbf{R}_t = (R_{1,t}, R_{2,t}, ..., R_{N,t})'$. Consider the scalar process $\{\ln M_t\}$ and vector process $\{\ln \mathbf{R}_t\}$. We assume $\{\ln(M_t\mathbf{R}_t)\}$ is stationary, and all three processes are ergodic with finite second moments: $\sup_t \mathbb{E}[\ln R_{i,t}]^2 < \infty$ and $\sup_t \mathbb{E}[\ln M_t]^2 < \infty$.

Assumption 1 (no arbitrage) is necessary and sufficient for existence of a strictly positive SDF that makes the asset-pricing equation hold for all assets. Positivity is needed for logarithmic transformations.[3] This assumption underlies arbitrage pricing theory (APT), essential for estimating a valid SDF from return data. APT requires that arbitrage portfolios be well-diversified (Ross, 1976), implying that asset selection in $\mathbf{R}_t$ must avoid redundancies to prevent singularity in their variance-covariance matrix.

Assumption 2 limits temporal dependence and ensures moment conditions for applying limit theorems for triangular arrays. In particular, a weak law of large numbers (WLLN) applies to cross-sectional averages of $R_{i,t}$ and of its suitable transformations (see, e.g., Durrett (2019) for WLLN and CLT for triangular arryas.).

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. We follow the basic assumptions in Hansen and Singleton (1983) to prove our main results. However, in the Appendix, we show that these results hold under a much less restrictive set of assumptions.

Hansen and Singleton assume conditional log-Normality for $M_t\mathbf{R}_t$ and homoskedasticity (across time) for respective conditional variances. Using the properties of multivariate Normal distributions, these assumptions allow us to rewrite the equations in (2) as a log-linear system:

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

[3]Under weak conditions (Hansen and Jagannathan, 1991, 1997) a valid M_t exists without requiring complete markets. However, uniqueness of M_t does require completeness. When markets are incomplete, there may exist infinitely many SDFs consistent with observed prices, but there is a unique M_t^* (mimicking portfolio) in the payoff space pricing all traded securities. Any SDF can be expressed as $M_t = M_t^* + \nu_t$, with $\mathbb{E}_t[\nu_t x_{i,t}] = 0$. Thus, M_t and M_t^* have identical pricing properties.

where $r_{i,t} = \ln(R_{i,t})$ and $m_t = \ln(M_t)$ and the term σ_i^2 is time invariant.[4] Note that, in our context, the pricing errors in (3), $\varepsilon_{i,t} = (m_t + r_{i,t}) - \mathbb{E}_{t-1}[m_t + r_{i,t}]$, $i = 1, 2, ..., N$, have a joint conditional multivariate *Normal* distribution. So, all marginal conditional distributions are *Normal*, i.e., $\varepsilon_{i,t} | \mathcal{I}_{t-1} \sim \mathcal{N}(0, \sigma_i^2)$ for all i.

Across time, for every asset, $\varepsilon_{i,t}$ is a martingale difference. Hence, $\mathbb{E}_{t-1}[\varepsilon_{i,t}] = 0$ and $\mathbb{E}[\mathbb{E}_{t-1}[\varepsilon_{i,t}]] = \mathbb{E}[\varepsilon_{i,t}] = 0$. Across individual assets, for every time period, the mean of $\varepsilon_{i,t}$ is also nil. Therefore, as in Hansen and Singleton (1983, p. 253, eq.(9)), we have here $\mathbb{E}^i[\varepsilon_{i,t}] = 0$. We will use this convenient property of $\varepsilon_{i,t}$ as an identifying assumption in this section, further assuming that $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 = \overline{\sigma}^2 < \infty$ for completeness.

Cross-Sectional Restrictions for M_t Identification

Here, we follow Hansen and Singleton by assuming that $\mathbb{E}^i[\varepsilon_{i,t}] = 0$ for all t, making it a suitable tool to identify the latent stochastic process $\{M_t\}$ of a valid SDF, which solves the well-known Euler equation of the consumption-saving decision of an agent embedding the purchase of a set of assets to transfer wealth across time.

In using $\mathbb{E}^i[\varepsilon_{i,t}] = 0$, we build a bridge between two important strands of the literature on asset pricing. The first strand searches for estimates of a valid SDF (a single factor) that are a function of return data; see Hansen and Jagannathan (1991, 1997) on the minimum-variance SDF and more recently Korsaye, Quaini and Trojani (2019) and Sandulescu, Trojani, and Vedolin (2021) on versions based on other optimizing criteria. The second strand, after the work of Fama and MacBeth (1973) and Fama and French (1992), shows that cross-sectional means of return data can be explained by multiple relevant factors that price the cross-sectional dimension of asset returns. So, it calls for zero mean pricing errors in the cross-sectional dimension after controlling for these factors.

[4] By construction, the conditional variance of $\varepsilon_{i,t}$ reads as:

$$\sigma_i^2 = \sigma^2_{\mu^r_{i,t}} + \sigma^2_{\mu^m_t} + 2\text{COV}\left(\mu^r_{i,t}, \mu^m_t\right), \qquad i = 1, 2, \cdots, N, \tag{4}$$

where $\mu^m_t := m_t - \mathbb{E}_{t-1}[m_t]$, and $\mu^r_{i,t} := r_{i,t} - \mathbb{E}_{t-1}[r_{i,t}]$, $\sigma^2_{\mu^r_{i,t}} = \text{VAR}\left(\mu^r_{i,t}\right)$, $\sigma^2_{\mu^m_t} = \text{VAR}(\mu^m_t)$, and $\text{COV}\left(\mu^r_{i,t}, \mu^m_t\right)$ is the covariance between $\mu^r_{i,t}$ and μ^m_t.

A key advantage of focusing on a single-factor approach is that it explicitly rules out factor proliferation. Instead of focusing on an increasing number of factors to price assets, our approach increases indefinitely the number of diversifiable asset returns, placing them into a valid SDF that can be identified by using (2) and the restriction that pricing errors $\varepsilon_{i,t}$ will have a zero cross-sectional mean.

To implement identification, take the average across i of equation (3):

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

From a WLLN for triangular arrays, $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it} \xrightarrow{p} \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}\right] \equiv \mathbb{E}^i\left[\varepsilon_{i,t}\right]$. Hence, using $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 = \overline{\sigma}^2 < \infty$, and imposing $\mathbb{E}^i\left(\varepsilon_{i,t}\right) = 0$ as N diverges, we obtain:

$$\mathbb{E}^i\left[r_{i,t}\right] = -m_t - \frac{1}{2}\overline{\sigma}^2. \tag{5}$$

By rewriting (5) and exponentiating, we obtain:

$$\exp\left(\frac{1}{2}\overline{\sigma}^2\right) M_t = \exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right). \tag{6}$$

Since the $r_{i,t}$ on the right-hand-side of (6) are data alone, it identifies the product:

$$\exp\left(\frac{1}{2}\overline{\sigma}^2\right) M_t.$$

Note that, if we are able to identify $\exp\left(\frac{1}{2}\overline{\sigma}^2\right)$ – a constant term – then M_t is ultimately identified. Multiply equation (2) by $\exp\left(\frac{1}{2}\overline{\sigma}^2\right)$, to obtain:

$$\mathbb{E}_{t-1}\left[\exp\left(\frac{1}{2}\overline{\sigma}^2\right) M_t R_{i,t}\right] = \exp\left(\frac{1}{2}\overline{\sigma}^2\right), \qquad i = 1, 2, \ldots, N. \tag{7}$$

Take the unconditional mean of (7) across *time* and *assets*[5], and use the law of iterated

[5]It is critical to take the unconditional mean of (7) across *time* and *assets* here, at the cost of not

expectations to obtain:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\overline{\sigma}^2\right) M_t R_{i,t}\right] = \exp\left(\frac{1}{2}\overline{\sigma}^2\right). \tag{8}$$

Recall that $\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right) = \exp\left(\frac{1}{2}\overline{\sigma}^2\right) M_t$ from (6). Using this last result, we can identify $\exp\left(\frac{1}{2}\overline{\sigma}^2\right)$ as:

$$\mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right) R_{i,t}\right] = \exp\left(\frac{1}{2}\overline{\sigma}^2\right). \tag{9}$$

Therefore, from equation (6), M_t is identified as the ratio between $\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)$ and $\mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right) R_{i,t}\right]$.[6]

An additional implication of the assumption that $\mathbb{E}^i\left[\varepsilon_{i,t}\right] = 0$ is the following. From equation (5) we can compute two important moments across *time* alone:

$$\begin{aligned} \mathrm{COV}\left(\mathbb{E}^i\left[r_{i,t}\right], m_t\right) &= -\sigma_m^2 = \mathrm{COV}\left(-m_t, m_t\right), \text{ and,} & (10)\\ \mathrm{VAR}\left(\mathbb{E}^i\left[r_{i,t}\right]\right) &= \sigma_m^2, & (11) \end{aligned}$$

which implies that:

$$\rho_{\mathbb{E}^i[r_{i,t}],m_t} = \frac{\mathrm{COV}\left(\mathbb{E}^i\left[r_{i,t}\right], m_t\right)}{\sigma_m \times \sigma_m} = \frac{\mathrm{COV}\left(-m_t, m_t\right)}{\sigma_m^2} = \frac{-\sigma_m^2}{\sigma_m^2} = -1, \tag{12}$$

where $\rho_{\mathbb{E}^i[r_{i,t}],m_t}$ is the correlation coefficient between $\mathbb{E}^i\left[r_{i,t}\right]$ and m_t.

Structural-Form (Factor-Model) Identification of M_t

The assumption that $\mathbb{E}^i\left[\varepsilon_{i,t}\right] = 0$ for all t also delivers "structural-form identification" of M_t, in the sense that it implies a single-factor model in the finance tradition, where the common factor m_t has: (i) heterogeneous *betas* for all assets, (ii) the factor-model error term is orthogonal to m_t – the single factor in the system (3), and (iii) the error term of

having a proper SDF pricing all assets. We thank an anonymous referee who pointed that out to us.

[6]Note that the strategy implicit in equation (9) is to replace higher order terms related to the variance with lower other terms which are a function of R_{it} and r_{it}.

the factor model has a zero cross-sectional and time-series mean.

To obtain a proper factor model, for each i, first use the following OLS projection across *time* to eliminate the dependence of $\varepsilon_{i,t}$ on m_t, and rewrite $\varepsilon_{i,t}$ as:[7]

$$\varepsilon_{i,t} = \delta_i \left(m_t - \mathbb{E}\left[m_t\right]\right) + \xi_{i,t}, \text{ where } \delta_i = \frac{\text{COV}\left(\varepsilon_{i,t}, \left(m_t - \mathbb{E}\left[m_t\right]\right)\right)}{\text{VAR}\left(m_t\right)}. \tag{13}$$

Combining (3) and (13), we obtain a proper single-factor model with heterogeneous intercepts and slopes:

$$r_{i,t} = -\left(\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}\left[m_t\right]\right) + \beta_i m_t + \xi_{i,t}, \text{ where } \beta_i = -1 + \delta_i. \tag{14}$$

A zero time series mean for $\xi_{i,t}$ comes directly from equation (13)), recalling that $\mathbb{E}\left[\varepsilon_{i,t}\right] = 0$. We are left to show that $\mathbb{E}^i\left[\varepsilon_{i,t}\right] = 0$ also delivers $\mathbb{E}^i\left[\xi_{i,t}\right] = 0$. As shown above, equation (10) is a consequence of $\mathbb{E}^i\left[\varepsilon_{i,t}\right] = 0$. Note that it leads to $\mathbb{E}^i\left[\delta_i\right] = 0$:

$$\begin{aligned}
\text{COV}\left(\mathbb{E}^i\left[r_{i,t}\right], m_t\right) + \sigma_m^2 &= 0, \text{ or,} \\
\mathbb{E}^i\text{COV}\left(r_{i,t}, m_t\right) + \mathbb{E}^i\text{COV}\left(m_t, m_t\right) &= 0, \text{ or,} \\
\mathbb{E}^i\text{COV}\left(r_{i,t} + m_t + \frac{1}{2}\sigma_i^2, m_t\right) &= 0, \text{ or,} \\
\mathbb{E}^i\left[\frac{\text{COV}\left(\varepsilon_{i,t}, \left(m_t - \mathbb{E}\left[m_t\right]\right)\right)}{\text{VAR}\left(m_t\right)}\right] &= \mathbb{E}^i\left[\delta_i\right] = 0.
\end{aligned} \tag{15}$$

In turn, from (13), $\mathbb{E}^i\left[\delta_i\right] = 0$ coupled with $\mathbb{E}^i\left[\varepsilon_{i,t}\right] = 0$, implies that $\mathbb{E}^i\left[\xi_{i,t}\right] = 0$.

Next, we state the key formal sufficient condition assumption for identification of M_t and the existence of a proper single-factor model representation for $r_{i,t}$ in the finance tradition.

Assumption 3 (Identifying Assumption): We assume the following: $\mathbb{E}^i\left[\varepsilon_{i,t}\right] = 0$, $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \sigma_i^2 = \overline{\sigma}^2 < \infty$, $\mathbb{E}^i|r_{i,t}| < \infty$, and $\mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right) R_{i,t}\right] \neq 0$.

[7] Given the assumption of Normality for $\varepsilon_{i,t}$, orthogonality implies independence.

Lemma 1 *Using the setup of Hansen and Singleton (1983), and Assumptions 1-3, we can identify M_t as:*

$$M_t = \frac{\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)}{\mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right) R_{i,t}\right]}. \tag{16}$$

Moreover, in a panel-data context, Assumptions 1-3 allow us to re-write the Asset-Pricing Equation as the single-factor logarithmic model:

$$r_{i,t} = -\left(\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}\left[m_t\right]\right) + \beta_i m_t + \xi_{it}, \text{ where } \beta_i = -1 + \delta_i, \tag{17}$$

with $\mathbb{E}\left[\xi_{i,t}\right] = \mathbb{E}^i\left[\xi_{i,t}\right] = 0$; $\xi_{i,t}$ orthogonal to the single factor m_t; and $\mathbb{E}^i\left[\delta_i\right] = 0$, or $\mathbb{E}^i\left[\beta_i\right] = -1$.

Lemma 1 provides an identification by construction (see, e.g., Section 3.4 of Lewbel (2019)).[8] In Appendix A.2, Lemma 5 generalizes the above result dispensing with the Log-Normality assumption.

An interesting result in Lemma 1 is that the single-factor representation provides an empirically testable implication of the identification condition. The model in (17) can be estimated using panel data methods and the identifying condition $\mathbb{E}^i\left[\beta_i\right] = -1$ can be empirically tested by employing a Wald test for the mean slope.

A Consistent Estimator of M_t and its Asymptotic Properties

Identification shows that $M_t = \frac{\exp(-\mathbb{E}_i[r_{i,t}])}{\mathbb{E}[\exp(-\mathbb{E}_i[r_{i,t}])R_{i,t}]}$. We use a simple sample analogue version of it to estimate M_t consistently. $\mathbb{E}^i\left[r_{i,t}\right]$ is estimated as:

$$\frac{1}{N}\sum_{i=1}^{N} r_{i,t}. \tag{18}$$

[8]Identification by construction means that one is able to write a closed-form expression for the parameter of interest as a function of available information (or data). An example is to prove that the parameter of interest equals the probability limit of its corresponding estimator. A caution here is not to assume the result.

Multiply (18) by minus one, and use the Continuous Mapping Theorem to estimate $\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)$ by:

$$\prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}} := \overline{R}_t^G, \tag{19}$$

where $\overline{R}_t^G$ is the geometric average of reciprocals of returns.

For the denominator of (16), recall from equation (9) that $\exp\left(\frac{1}{2}\overline{\sigma}^2\right)$ is identified as:

$$\mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right) R_{i,t}\right] = \exp\left(\frac{1}{2}\overline{\sigma}^2\right). \tag{20}$$

It is straightforward to propose a consistent estimator for $\exp\left(\frac{1}{2}\overline{\sigma}^2\right)$. Recall that $\overline{R}_t^G$ is a consistent estimator for $\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)$. If we replace the latter with its consistent estimator, and use sample averages to replace the unconditional expectations (across time and assets), we obtain an estimator for $\exp\left(\frac{1}{2}\overline{\sigma}^2\right)$ as follows:

$$\begin{aligned}
\widehat{\exp\left(\frac{1}{2}\overline{\sigma}^2\right)} &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[\left(\prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}}\right) R_{i,t}\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[\left(\prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}}\right)\frac{1}{N}\sum_{i=1}^{N} R_{i,t}\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\overline{R}_t^G\overline{R}_t^A,
\end{aligned} \tag{21}$$

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

Finally, taking into account results for (19) and for (21), we are able to propose the following consistent estimator for M_t as the ratio:

$$\widehat{M_t} = \frac{\overline{R}_t^G}{\frac{1}{T}\sum_{j=1}^{T}\overline{R}_j^G\overline{R}_j^A}. \tag{22}$$

The estimator for a valid SDF in equation (22) is very simple, intuitive, and easy to implement in practice. It 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 $1/N$. The estimator employs the sample analogue of the population moments that identify the SDF, and does not depend on any parametric function used to characterize preferences.

The next proposition states our main consistency result.

Proposition 2 *Using the setup of Hansen and Singleton (1983) and Assumptions 1-3, the following holds, as* $(N,T) \to \infty$*:*

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

for all t*.*

Now we discuss the limiting distribution of the SDF estimator.

Proposition 3 *Using the setup of Hansen and Singleton (1983), Assumptions 1-3, as* $(N,T) \to \infty$*, and* $T/N \to 0$ *the following holds:*

$$\sqrt{T}\left(\widehat{M}_t - M_t\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, M_t^2 V),$$

where $V = \frac{V^{GA}}{\left(\mathbb{E}\left[R_t^G R_t^A\right]\right)^2}$ *with* $V^{GA} = \text{VAR}\left(R_t^G R_t^A\right)$*,* $R_t^G := \exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)$*, and* $R_t^A := \mathbb{E}^i\left[R_{i,t}\right]$*.*

A consistent estimator of the denominator of V is $\left(\frac{1}{T}\sum_{t=1}^{T} \overline{R}_t^G \overline{R}_t^A\right)^2$. Moreover, one is able to use a Newey and West (1987) type of estimator to estimate V^{GA} as:

$$\widehat{V}^{GA} = \frac{1}{T}\sum_{t=1}^{T}(\overline{R}_t^G \overline{R}_t^A - \overline{\overline{R}_t^G \overline{R}_t^A})^2 + \sum_{j=-S_T}^{S_T} k_T(j) \frac{1}{T}\sum_{t=1}^{T-j}(\overline{R}_t^G \overline{R}_t^A - \overline{\overline{R}_t^G \overline{R}_t^A})(\overline{R}_{t+j}^G \overline{R}_{t+j}^A - \overline{\overline{R}_{t+j}^G \overline{R}_{t+j}^A}), \tag{23}$$

where $\overline{\overline{R}_t^G \overline{R}_t^A} = \frac{1}{T}\sum_{t=1}^{T} \overline{R}_t^G \overline{R}_t^A$, $k_T(\cdot)$ is a real-valued kernel with $k_T(0) = 0$ – e.g., Bartlett kernel $k_T(j) = k(j/S_T)$, where $k(x) = 1 - |x|$ for $x \in [0,1]$, and 0 otherwise –, and S_T is a bandwidth parameter – e.g., $S_T = const \cdot T^{1/3}$. Andrews (1991) provides further discussions on practical choices of kernel and bandwidth.

We remark that, since $\widehat{M}_t$ is a nonlinear estimator, hence we use a joint limit theory together with $T/N \to 0$ to obtain its limiting distribution. This condition requires that the cross-section dimension grows slightly faster than the time-series to establish the results.[9] Nevertheless, the use of large panels is common in the nonlinear panel data literature, see e.g., Arellano and Hahn (2007). It is possible to relax the requirement on the sample growth to $T/N \to k < \infty$. Proposition 6 in the Appendix A.2 provides the result. However, in such case the limiting distribution has contributions of both the numerator and denominator terms. Hence, inference depends on a nuisance parameter. In this case, a bootstrap procedure for inference would be recommended.

We present numerical Monte Carlo simulations to evaluate the finite sample performance of the proposed methods in the Appendix A.4. We generate data using a no-arbitrage dynamic (consumption) capital asset-pricing model and employ it to investigate how close our proposed stochastic discount factor (SDF) estimator is to actual SDFs across simulations for medium-size panel-data samples. The model is mid-sized 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. This result survives robustness-analysis exercises on several dimensions.

[9] It is possible to obtain the exact same result in Proposition 3 using the sequential asymptotic approach proposed by Phillips and Moon (1999, 2000), when first $N \to \infty$, and then $T \to \infty$. Sequential limits are not always the same as joint limits, under some circumstances one can establish a relationship between sequential limits and joint limits, see Lemma 6 Phillips and Moon (1999, p.1067) for details on the conditions.

2.2 Discussing Identification

A No-arbitrage Estimator

The Asset-Pricing Equation holds with a valid $M_t > 0$ if and only if no arbitrage holds. Using it, 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. Intuitively, m_t leaves its imprint in every asset being priced in the cross-sectional dimension. Our job was to filter from $r_{i,t}$ data the single factor component related to a valid SDF in order to finally identify M_t.

Take the cross sectional average of (14), multiply the result by minus one, and take probability limits, to obtain:

$$\operatorname*{plim}_{N\to\infty} -\frac{1}{N}\sum_{i=1}^{N} r_{i,t} = \operatorname*{plim}_{N\to\infty} -\frac{1}{N}\sum_{i=1}^{N} \beta_i m_t + \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}\left[m_t\right]\right) - \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \xi_{it}. \tag{24}$$

Given Assumptions 2 and 3 and the results in Lemma 1, $\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \delta_i = 0$, and $\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \xi_{it} = 0$. Using the Continuous Mapping Theorem, we are left with:

$$\operatorname*{plim}_{N\to\infty} \overline{R}_t^G \;=\; M_t^{-\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\beta_i} \exp\left(\frac{1}{2}\overline{\sigma}^2\right) = M_t \exp\left(\frac{1}{2}\overline{\sigma}^2\right), \tag{25}$$

where in the first equality we employ the result that $\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \delta_i = 0$, and in the second we use $\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \beta_i = -1$.

Equation (25) is critical: the asset pricing equation is the only tool at hand to identify M_t, but, it only holds with unit power for M_t itself. Indeed, this is the power we need to identify and estimate $\exp\left(\frac{1}{2}\overline{\sigma}^2\right)$ using it, the final step to identify M_t as discussed above.

Alternative Weighting Schemes for Returns

One may wonder whether it would be possible to achieve identification in any other way. The simple answer is yes. Indeed, Assumption 3 imposes the restriction that $\mathbb{E}^i\left[\varepsilon_{it}\right] = 0$

to establish identification. This assumption was natural, given we employ the setup in Hansen and Singleton (1983), which assumes it. Nevertheless, one could think about using a weighted average instead. This would require a zero-weighted average of the innovations, i.e., $\mathbb{E}^{i\omega}[\varepsilon_{it}] := \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\omega_i\varepsilon_{it}\right] = 0$, where ω_i is a positive deterministic weight. In this case, to establish identification, one would take a weighted average of equation (3), and a corresponding law of large numbers, see, e.g., Wright, Platt, and Robertson (1977) and Chandra and Ghosal (1996) for sufficient conditions to apply law of large numbers for weighted averages. These modifications would imply a different identification of a valid SDF in which $\mathbb{E}^{i}[r_{i,t}]$ would be replaced by $\mathbb{E}^{i\omega}[r_{i,t}]$. Moreover, the corresponding estimator for the SDF would also depend on the weights. This feature of identification would require the estimation of these weights, which would substantially complicate practical applications, since the number of parameters to estimate would be large, and increasing with the sample size. Nevertheless, it could be a bridge between our approach and the SDF approach of Sandulescu, Trojani, and Vedolin (2021), and of Korsaye, Quaini, and Trojani (2021), where these weights are drawn from a previous optimization step.

An additional issue would be that, if we had used weights ω_i, $i = 1, 2, \cdots, N$, then, our combination of (log) returns in (24) would be, in the limit, $-\mathbb{E}^{i\omega}[r_{i,t}]$. From (25) an important issue is whether or not $-\frac{1}{N}\sum_{i=1}^{N}\omega_i\beta_i \xrightarrow{p} -1$, so that we can employ the asset pricing equation for identification and consistent estimation of $\exp\left(\frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}\omega_i\sigma_i^2\right)$.

One of the possible choices for weights is to set:

$$\omega_i = \frac{1}{\beta_i}, \text{ leading to } \quad -\frac{1}{N}\sum_{i=1}^{N}\omega_i\beta_i \xrightarrow{p} -1.$$

Another possibility is to group assets by some characteristic and weight simple within-group averages by their relative importance, with group participation adding up to 1. In

this case, for every asset i within the jth group, we have the following weights:

$$\frac{1}{N} = \frac{N_j}{N}\frac{1}{N_j},$$

where N_j is the total number of assets in the j-th group, $\frac{N_j}{N}$ measures the proportion of assets in the j-th group in population and $\frac{1}{N_j}$ is used to form within-group j averages. The number of groups could be finite or even infinite, but N_j and N have to diverge to infinity with $\frac{N_j}{N}$ converging to the proportion of group j in population.

We must stress that the fact that there may exist other weighting schemes that could be used in identification does not invalidate using the label *no-arbitrage identification* for our approach. Identification does not require the existence of a *unique* approach. Instead, it requires that a *given* identification approach leads to a unique parameter, or random variable – in our case, the stochastic process of a valid SDF $\{M_t\}$.

Factor and Factor-Loading Identification

In the context of equation (14), which has a single factor, we know from the seminal work of Lawley and Maxwell (1971) that there exists an indeterminacy problem for factor models that could be explained as follows:

$$\widetilde{\beta}_i\widetilde{m}_t = \left(\widetilde{\beta}_i\kappa^{-1}\right)(\kappa\widetilde{m}_t) = \beta_i m_t, \tag{26}$$

i.e., a factor and its respective loadings are identifiable up to multiplication by non-zero finite scalar κ. Clearly, κ determines the mean and *scale* of m_t and also the mean and *scale* of β_i. So, it is an important choice for identification of the factor and its loadings. The choice of κ is not an innocuous *normalization* in this context. On the contrary, it is an identification assumption that is tied to Assumption 3, where the location and scale of m_t and β_i are determined due to the identifying assumption that $\mathbb{E}^i\left[\varepsilon_{i,t}\right] = 0$, or, equivalently, that $\mu_t^m = -\mathbb{E}^i\left[\mu_{i,t}^r\right]$, where μ_t^m and $\mu_{i,t}^r$ are innovations in m_t and $r_{i,t}$ respectively.

Since $M_t = \exp(m_t)$, the scale of β_i determines the *power* of M_t. As argued above, this power is critical for identification of M_t. Indeed, $\mathbb{E}^i[\varepsilon_{i,t}] = 0$ implies $\mathbb{E}_i[\beta_i] = -1$, which allows identification and consistent estimation of $\exp\left(\frac{1}{2}\bar{\sigma}^2\right)$ using the Asset-Pricing Equation, and finally identification and consistent estimation of M_t.

We note that the implication that $\mathbb{E}^i[\beta_i] = -1$ can be tested using a simple Wald test for the mean slope based on estimates of a one-factor model. We implement this test empirically in Section 3.1.

Finally, although our model results in a single factor, it contains information on multiple returns trough the SDF, since the latter is a function of R_{it} for $i = 1, ..., N$ and $t = 1, ..., T$. We stress that, as pointed out by Cochrane (2005, Section 6.3), there is an equivalence between a single factor M_t and a linear combination of multiple factors such that

$$M_t = b' f_t, \tag{27}$$

where f_t is a vector stacking multiple factors including a constant term, and b stacks the loadings of such factors. In empirical work, one could either look at the right-hand side of equation (27), or at its left-hand side. We chose the latter. In the empirical section of the paper, we investigate how our single factor model fares with well-known multiple factor models commonly used in empirical work. Empirical results confirm that the single factor model is superior to price assets in sample and out of sample.

2.3 Economic Properties of the M_t Estimator and its Population Counterpart

The first property of our estimator, $\widehat{M_t}$, is that it is a non-parametric non-linear estimator, and a function of asset-return data alone. Indeed, 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^*. Hence, $\widehat{M}_t$ converges to this unique pricing kernel. Under incomplete markets, no-arbitrage implies that there exists at least one valid 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, $\widehat{M}_t$ converges to one of those valid M_t. Of course, this poses no problem, even if there exists an infinite number of valid SDFs, since all of them have identical pricing properties.

Third, from a population point of view, it is straightforward to verify that our identified M_t was constructed to obey:

$$\begin{aligned}\mathbb{E}\left[M_t R_{i,t}\right] &= \mathbb{E}\left[\frac{\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)}{\mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)R_{i,t}\right]}R_{i,t}\right] \\ &= \frac{\mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)R_{i,t}\right]}{\mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)R_{i,t}\right]} = 1, \quad i = 1, 2, , N, \end{aligned} \tag{28}$$

where the expectation on the left-hand side of (28), and the outer expectation of the numerator and denominator of the right-hand-side are taken across *time* and *assets*. This result is a natural property arising from the moment restrictions entailed by the Asset-Pricing Equation (2).[10] Moreover, since the identified M_t correctly prices every return in the economy, it should also correctly price the simple average of returns.

2.4 Comparisons with the Literature

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 show 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

[10] We thank an anonymous referee for suggesting this exercise to us.

project M_t onto the space of returns, which makes it straightforward to express M_t^*, the mimicking portfolio, only as a function of observable returns – similar to what we have here – a model-free estimate of M_t.

Still related to admissibility, there is a literature focused on model-free empirical SDFs, imposing a given set of asset pricing constraints and minimizing some alternative target notion of SDF variability. For example, SDFs minimizing entropy are given by a nonlinear transformation of gross returns (see, e.g., Sandulescu, Trojani, and Vedolin (2021)). The key point of this literature is that these procedures produce natural SDF variability lower bounds for testing asset pricing models. In addition, Korsaye, Quaini and Trojani (2021) have explored the ability of model-free SDFs in this family to produce out-of-sample pricing performance under various regularization schemes mitigating overfitting concerns.

The main focus of Hansen and Jagannathan, and the literature that followed, has been on the admissibility of SDF proxies. Our approach has in common with this literature the fact that M_t is a function of asset returns alone. We also believe that our approach delivers admissibility, but, from a different angle. We start by assuming the existence of a valid SDF, therefore admissible. Then, we prove its identification and propose a consistent estimator for it. So, from a population perspective we achieve admissibility.

More recently, there is an empirical SDF literature using machine learning and deep learning constructing model-free valid SDFs, that also builds directly on Hansen and Jagannathan. For example, Kozak, Nagel, and Santosh (2020) propose a robust SDF by shrinking the cross-section of return predictors. Feng, Giglio, and Xiu (2020) propose model selection methods to systematically evaluate the contribution to asset pricing of any new factor trying to tame the factor zoo. Giglio, Kelly, and Xiu (2022) and Kelly and Xiu (2023) survey machine-learning approaches that seek to extract admissible SDFs from high-dimensional return data. Chen, Pelger, and Zhu (2024) employ deep learning to estimate flexible, return-based SDFs. Here, rather than relying on shrinkage or flexible statistical learning, we provide a simple, theory-based, non-parametric estimator grounded

on no-arbitrage and panel-data asymptotics. Hence, our method can be viewed as a large-sample analogue of the Hansen and Jagannathan's mimicking-portfolio, but with testable restrictions and asymptotic properties. Our paper is also related to Barillas and Shanken (2017), who stress the equivalence between factor models and SDF admissibility tests.

3 Empirical Applications

This section presents two empirical exercises. First, we estimate a valid SDF using a large N and T panel and then estimate our proposed factor model using a multi-effect linear regression model that allows for parameter heterogeneity. Using regression results we test directly our key identifying assumption. In this exercise, we take into account the fact that our regressor was estimated using asset-return data, i.e, we have to deal with a generated-regressor problem.

Second, in a comprehensive investigation, we assess how well our SDF factor prices the cross-sectional returns of Fama-French portfolios using Fama-MacBeth regressions. We employ a natural list of competitors, consider pricing well-known anomalies, and evaluate competing models in-sample and out-of-sample.

3.1 Factor Model Evaluation: A Panel-Data Analysis

We use 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 $(3,986)$. The original panel is unbalanced. For example, some firms do not survive and are dropped from the stocks class. Some government bonds are short-lived and therefore are dropped from the data base as well, etc.

All return data used in this exercise came from either CRSP or Bloomberg. Mutual-Fund return data comes from the CRSP Mutual Fund Database, which reports open-

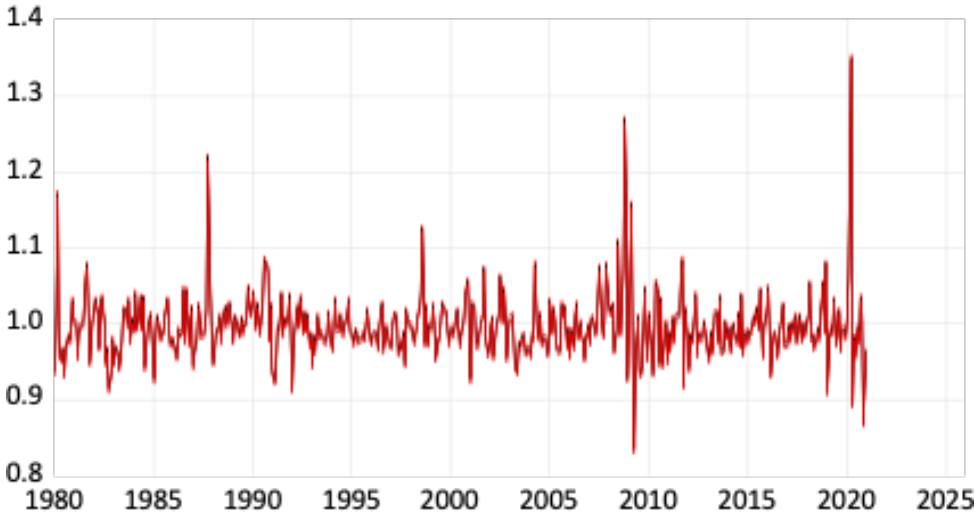
ended mutual-fund returns using survivor-bias-free data. Stock return data comes from the CRSP U.S. Stock and CRSP U.S. Indices, which collect returns from NYSE, AMEX, NASDAQ, and, more recently, NYSE Arca. Real-Estate return data comes from the CRSP/Ziman Real Estate Data Series. It collects return 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%).[11] 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, 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. As a robustness check, we changed these weights (from 5 up to 20 percentage points for individual categories), computing again $\widehat{M_t}$, but 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 1 in black, which includes 95% asymptotic confidence intervals in red.

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

Figure 1: Stochastic Discount Factor Panel-Data Estimate (Black) and 95% Asymptotic Confidence Bands (Red)



Note: Stochastic Discount Factor constructed using $N = 102,698$ asset returns from four different asset classes: Stocks (CRSP U.S. Stock and CRSP U.S. Indices), Bonds (CRSP Monthly Treasury U.S. Database and Bloomberg) , Mutual Funds (CRSP Mutual Fund Database) and REITS (CRSP/Ziman Real Estate Data Series) at the monthly frequency from 1980:1 through 2020:12 ($T = 492$). 95% asymptotic confidence intervals are in red. The estimate of M_t is in black.

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:

$$r_{i,t} = -\left(\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}\left[m_t\right]\right) + \beta_i m_t + \xi_{i,t}, \qquad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T. \tag{29}$$

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:

$$
\begin{aligned}
r_{i,t} &= -\left(\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}\left[m_t\right]\right) + \beta_i m_t + \xi_{i,t}, \qquad \text{where} \qquad (30)\\
\frac{1}{2}\sigma_i^2 + \delta_i \mathbb{E}\left[m_t\right] &= \frac{1}{2}\sigma_0^2 + \delta_0 \mathbb{E}\left[m_t\right] + \mu_i, \quad \mu_i \sim \mathcal{N}\left(0, \sigma_\mu^2\right),\\
\beta_i &= \beta_0 + \nu_i, \quad \nu_i \sim \mathcal{N}\left(0, \sigma_\nu^2\right).
\end{aligned}
$$

Following the *mixed-effect* panel-data literature, we estimate (30) by maximum likelihood assuming a (zero-mean) Gaussian *random-effect* specification for μ_i and ν_i. To implement estimation we constructed a balanced panel across asset classes.

The beta-model in (29) allows us to empirically test the identification condition in Assumption 3 above through the implication that $\mathbb{E}^i\left[\beta_i\right] = -1$. Note that the mixed model in equation (30) allows for heterogeneous slopes, and the parameter β_0 captures the mean $\mathbb{E}^i\left[\beta_i\right]$, that is, $\mathbb{E}^i\left[\beta_i\right] = \beta_0$. We employ a Wald test for the null hypothesis $H_0 : \beta_0 = -1$, based on the model estimates. Results for estimates and the testing procedure are presented in Table 1, below. We report cross-section clustered standard errors based on the four categories of assets. We note that SDF estimates enter the model as a covariate. To address a potential problem of generated regressors, in the addition to the regular standard errors, we report bootstrapped standard errors. We employ a nonparametric cross-sectional bootstrap as described in Kapetanios (2008), where we keep the time-series dimension intact and resample with replacement from the cross-sectional units.

In Table 1, the estimate for β_0 – the mean of β_i – is -1.03 and it is statistically equal to -1 using a Wald test, regardless of how we estimate standard errors. This is a specification test for our identification condition in Assumption 3 above, since $\mathbb{E}^i\left[\varepsilon_{i,t}\right] = 0$ for all t implies $\mathbb{E}^i\left[\beta_i\right] = -1$. Note that, when we computed $\widehat{M_t}$, we did not impose this restriction on the factor loadings, so it could indeed be used as a specification test of the identifying assumption. Using a standard one-sided Wald test for the null of zero and

Table 1: *Mixed-effect* Model Estimation

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

Weights on (T,S,RE,MF)	β_0 (SE)	$\frac{1}{2}\sigma_0^2 + \delta_0 \mathbb{E}\left[m_t\right]$ (SE)	σ_μ^2 (SE)	σ_ν^2 (SE)	$H_0 : \beta_0 = -1$ (P-Value)
(0.20, 0.15, 0.60, 0.05)	-1.0317	0.0058	0.0026	0.5392	0.0092
	(0.3305)	(0.0012)	(0.0012)	(0.2067)	(0.9235)
	[0.4562]	[0.0018]	[0.0015]	[0.2120]	[0.85145]

Note: Model estimated by Maximum Likelihood. A logarithmic version of $\widehat{M}_t$ is used as the regressor in the factor model. Robust standard errors in parentheses employ cross-sectional clusters based on the four categories of assets. Standard errors in brackets use a cross-sectional bootstrap procedure.

alternative of positive variance, estimates of σ_μ^2 and σ_ν^2 are both statistically different from zero and away from the boundary.

Next, we perform a sensitivity analysis of the results in Table 1 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 2. There is very little change in results – either quantitatively or qualitatively, which is reassuring.

Table 2: *Mixed-effect* Model Estimation – Sensitivity Analysis

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

Weights on				β_0	$\frac{1}{2}\sigma_0^2 + \delta_0 \mathbb{E}\left[m_t\right]$	$H_0 : \beta_0 = -1$
Treasuries	Stocks	REITS	Mutual Funds	(SE)	(SE)	(P-Value)
0.10	0.15	0.70	0.05	-1.16708	0.0062	0.6513
				(0.2070)	(0.0009)	(0.41966)
				[0.4197]	[0.0022]	[0.8388]
0.10	0.20	0.60	0.10	-1.1241	0.0061	0.3261
				(0.2173)	(0.0013)	(0.5680)
				[0.3654]	[0.0023]	[0.9304]
0.15	0.10	0.60	0.15	-1.0299	0.0052	0.0100
				(0.2983)	(0.0014)	(0.9200)
				[0.3914]	[0.0018]	[0.7218]
0.15	0.10	0.70	0.05	-1.1049	0.0058	0.1394
				(0.2810)	(0.0009)	(0.7087)
				[0.4569]	[0.0017]	[0.9443]
0.15	0.20	0.50	0.15	-1.0188	0.0055	0.0046
				(0.2774)	(0.0016)	(0.9459)
				[0.3555]	[0.0021]	[0.9343]
0.15	0.20	0.60	0.05	-1.0938	0.0062	0.1240
				(0.2664)	(0.0012)	(0.7246)
				[0.4319]	[0.00217]	[0.7855]
0.20	0.10	0.60	0.10	-0.9997	0.0053	0.0000
				(0.3421)	(0.0013)	(0.9995)
				[0.4353]	[0.0016]	[0.8629]
0.20	0.15	0.50	0.15	-0.9567	0.0051	0.0173
				(0.3288)	(0.0016)	(0.8951)
				[0.3860]	[0.0019]	[0.8456]
0.20	0.15	0.60	0.05	-1.0317	0.0058	0.0092
				(0.3304)	(0.0012)	(0.9235)
				[0.4560]	[0.0018]	[0.8514]
0.20	0.20	0.50	0.10	-0.9886	0.0057	0.0012
				(0.3202)	(0.0015)	(0.9717)
				[0.4028]	[0.0020]	[0.9369]
0.25	0.10	0.50	0.15	-0.8945	0.0047	0.0784
				(0.3763)	(0.0015)	(0.7793)
				[0.4206]	[0.0016]	[0.7404]
0.25	0.10	0.60	0.05	-0.9696	0.0054	0.0061
				(0.3889)	(0.0012)	(0.9377)
				[0.4830]	[0.0016]	[0.9473]
0.25	0.15	0.50	0.10	-0.9265	0.0053	0.0393
				(0.3705)	(0.0015)	(0.8428)
				[0.4282]	[0.0017]	[0.9796]
0.25	0.20	0.50	0.05	-0.9585	0.0058	0.0128
				(0.3658)	(0.0015)	(0.9096)
				[0.4553]	[0.0020]	[0.7963]

Note: See estimation details in Table 1. Variance estimates omitted for the sake of space. Robust standard errors in parentheses employ cross-sectional clusters based on the four categories of assets. Standard errors in brackets use a cross-sectional bootstrap procedure.

3.2 The Cross-Sectional Distribution of Asset Returns

Our second empirical exercise is extracted from papers closely related to this one – Issler, Martins and Moog (2023) and Issler and Moog (2025) – who employed Fama-MacBeth regressions to measure the in-sample and out-of-sample fit of the cross-sectional distribution of different asset-pricing models.[12] Issler, Martins and Moog work with a monthly and a quarterly databases consisting of the 25 Fama-French portfolios both downloaded from the Fama-French library. Data ranges from July 1963 to June 2019 (672 monthly observations) and from July, 1963 through June 2019 (224 quarterly observations).

Using monthly data on excess returns, models are evaluated on an equal basis using the same sample. For every model, the first step regressions regresses returns on factors, estimating betas. In the second step, returns are regressed on the estimated betas for each factor and lambdas are estimated for each model. The number of factors determines the number of lambdas, and the model fit is measured by the $\overline{R}^2$ statistic. The following asset-pricing models are evaluated: the capital asset-pricing model (CAPM), the Fama and French (1993) 3-factor model (FF3), the Fama and French (2015) 5-factor model (FF5), and the model in this paper (AGI). Results are presented in Table 3.

Consistent with previous results by Lettau and Ludvigson (2001) and others, the CAPM has virtually no explanatory power for the cross-section of asset returns. The Fama and French 3-factor model is a big improvement over the CAPM, with a $\overline{R}^2 = 0.61$. Fama and French's 5-factor model raises the bar to $\overline{R}^2 = 0.71$. Up to now, we have been raising the number of factors as we present competing models. However, the model in this paper has only one (non-linear) factor, but it raises again the bar to $\overline{R}^2 = 0.95$. It is also worth noting that not all the lambdas are statistically significant. Indeed, individually, 40% of them are not.

[12] We thank an anonymous referee for suggesting the out-of-sample analyses.

Table 3: Fama-MacBeth Regressions – Monthly Fama-French Data (2nd-Step). Excess Return Data

Model	Constant	λ_{AGI}	λ_{vw}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CMA}	R^2	$\overline{R}^2$	Rank test
CAPM	1.44	-	-0.31	-	-	-	-	0.05	0.01	H_0 : rank ≤ 1
	(4.85)	-	(-1.14)	-	-	-	-	-	-	149.69 > [8.34]
FF3	1.48	-	-0.54	0.19	0.33	-	-	0.65	0.61	H_0 : rank ≤ 3
	(3.83)	-	(-1.45)	(3.63)	(5.42)	-	-	-	-	33.77 > [1.51]
FF5	1.10	-	-0.21	0.28	0.28	0.43	0.07	0.77	0.71	H_0 : rank ≤ 5
	(2.75)	-	(-0.54)	(5.19)	(5.05)	(2.53)	(0.33)	-	-	22.58 > [1.28]
AGI	0.63	0.96	-	-	-	-	-	0.96	0.95	H_0 : rank ≤ 1
	(27.81)	(22.26)	-	-	-	-	-	-	-	486.50 > [126.28]

Note: Second-stage Fama-MacBeth regressions using different proxies of a valid stochastic discount factor explaining the cross-sectional return variation of 25 monthly portfolios – July, 1963 through June 2019 (672 monthly observations). Dependent variables are excess returns. CAPM is the capital asset pricing model. FF3 is Fama and French's (1993) three-factor model. FF5 is Fama and French's (2015) five-factor model. AGI uses the consistent estimate of a valid SDF proposed in this paper. This table was extracted from Issler, Martins and Moog (2023, p. 21). White robust t-statistics in parentheses. The last column presents the rank test in Chen and Fang (2019) for the second-step of the procedure. It provides the null hypothesis (H_0), the test statistic, and the 5% level of significance critical value inside brackets.

The regressions presented in Table 3 are a two-step procedure, and the validity of the Fama-MacBeth results and of the associated $\overline{R}^2$ requires a full rank condition for the regressor matrix of the second step. There is a potential weak-identification problem in the cross-sectional regression under rank deficiency. To address this problem, the last column of Table 3 presents the rank test in Chen and Fang (2019) for the regressor matrix in the second-step of the Fama-MacBeth procedure. We test the null hypothesis of a deficient rank. The table provides the corresponding null hypotheses for each regression, the test statistic, and the associated 5% level of significance critical value (inside brackets). The results indicate that the null hypotheses of reduced rank are strongly rejected in all cases.

The second exercise performed by Issler, Martins and Moog is identical to the one reported in Table 3, but it used quarterly data and included additional SDF estimates – the Lettau and Ludvigson's conditional CAPM model with *cay* (Mkt×CAY), and the conditional CAPM model with *cay* augmented with labor income (HC×CAY). Results are presented in Table 4.

In Table 4, the CAPM model performed poorly once again. Both the conditional

CAPM model with *cay* and the conditional CAPM model with *cay* augmented with labor income performed much better than the CAPM model, but performed worse than the Fama and French 3-factor model – a $\overline{R}^2 = 0.66$. The Fama-French 5 factor model raised the bar to $\overline{R}^2 = 0.79$, but the estimate in this paper achieved the best fit, with a $\overline{R}^2 = 0.90$. The last column of Table 4 presents the rank test in Chen and Fang (2019) testing the null hypothesis of a deficient rank in the second-step regression. At the 5% level, results indicate that the null hypotheses of reduced rank are strongly rejected for the CAPM model, FF3, FF5, and the AGI model. Nevertheless, results for Mkt×CAY, and HC×CAY do not reject the null hypothesis of rank deficiency.

Table 4: Fama-MacBeth Regressions – Quarterly Fama-French Data (2nd-Step). Excess Returns Data.

Model	Constant	λ_{AGI}	λ_{vw}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CMA}	λ_{HC}	λ_{CAY}	$\lambda_{MKT\times CAY}$	$\lambda_{HC\times CAY}$	R^2	$\overline{R}^2$	Rank
CAPM	3.49	-	-0.01	-	-	-	-	-	-	-	-	0.00	-0.04	H_0 : rank ≤ 1
	(4.63)	-	(-0.02)	-	-	-	-	-	-	-	-	-	-	81.40 > [6.33]
FF3	3.80	-	-0.90	0.55	1.02	-	-	-	-	-	-	0.70	0.66	H_0 : rank ≤ 3
	(3.10)	-	(-0.76)	(4.04)	(5.80)	-	-	-	-	-	-	-	-	9.08 > [1.70]
FF5	1.30	-	1.45	0.71	0.87	1.10	0.66	-	-	-	-	0.83	0.79	H_0 : rank ≤ 5
	(1.09)	-	(1.25)	(6.00)	(5.98)	(2.90)	(1.50)	-	-	-	-	-	-	5.93 > [0.84]
AGI	2.13	0.96	-	-	-	-	-	-	-	-	-	0.91	0.90	H_0 : rank ≤ 1
	(22.02)	(15.00)	-	-	-	-	-	-	-	-	-	-	-	573.89 > [310.14]
Mkt×CAY	3.66	-	-0.69	-	-	-	-	-	0.02	-0.12	-	0.55	0.48	H_0 : rank ≤ 3
	(4.01)	-	(-0.74)	-	-	-	-	-	(1.60)	(-2.39)	-	-	-	23.65 < [45.98]
HC×CAY	2.53	-	0.44	-	-	-	-	0.66	0.01	-0.07	0.01	0.71	0.63	H_0 : rank ≤ 5
	(2.95)	-	(0.51)	-	-	-	-	(2.39)	(1.41)	(-1.53)	(0.58)	-	-	19.41 < [39.03]

Note: Second-stage Fama-MacBeth regressions using different proxies of a valid stochastic discount factor explaining the cross-sectional return variation of 25 quarterly portfolios – July, 1963 through June 2019 (224 quarterly observations). Dependent variables are excess returns. CAPM is the capital asset pricing model. FF3 is Fama and French's (1993) three-factor model. FF5 is Fama and French's (2015) five-factor model. AGI uses the consistent estimate of a valid SDF proposed in this paper. This table was extracted from Issler, Martins and Moog (2023, p. 16). White robust t-statistics in parentheses. The last column presents the rank test proposed by Chen and Fang (2019) for the second-step of the procedure. It provides, in that order, the null hypothesis (H_0), the test statistic, and the 5% level of significance critical value in brackets.

The next exercise, extracted from Issler and Moog (2025), evaluates the in-sample and out-of-sample performance of different estimates of M_t in pricing the cross-sectional distribution of asset returns. The framework is similar to that of Issler, Martins and Moog, but now Fama-French portfolios are sorted by the well-known anomalies identified by Hou, Xue and Zhang (2015), which makes it harder to forecast the cross-section of returns. Moreover, an extended group of competitors is now considered in the analysis.

In-sample evaluation means that, whenever possible, the same data used in constructing SDFs are used in the Fama and MacBeth regressions. Whenever these two group of assets differ, in cases that they could have been the same, the analysis is labeled as out-of-sample. Here, we report only the out-of-sample results. In-sample results are reported in Appendix A.3.

A key insight of Issler and Moog relates to the fact that the mimicking portfolio M_t^* of Hansen and Jagannathan (1991, 1997) has a perfect fit in explaining the the cross-sectional distribution of asset returns in-sample. This happens by construction, since M_t^* is designed to deliver a zero-mean pricing error across time and across assets. This property gives hope that a single factor estimate, which requires a zero-mean pricing error across assets will also do well in comparison to estimates designed to fit specific characteristics of the cross-sectional distribution of returns.

The testing ground uses portfolio returns sorted by the 34 anomalies (out of 80 identified in the literature) that Hou, Xue and Zhang (2015) deemed relevant in statistical testing. As in Lettau and Ludvigson (2001), the $\overline{R}^2$ statistic in the second-stage Fama and MacBeth regressions is used to compare the explanatory power of different SDF models. To accommodate Lettau and Ludvigson's (2001) conditional CAPM model, quarterly data is used from 1972:1 2016:4 – a total of 178 quarters.

Issler and Moog explain their approach: "[I]n essence, each anomaly in this paper represents a different database, with portfolio returns being sorted according to that anomaly. Since there are 34 anomalies - with two of them comprised of 9 portfolios and the rest comprised of 10 portfolios - in total ... the models in our paper ... price 338 portfolio returns... Moreover, we will also use the 25 portfolios of Fama and French (1993), which are based on Size and Book-to-Market Value, in robustness tests..."

Table 5, extracted from Issler and Moog, displays the $\overline{R}^2$ statistic of the second-stage Fama and MacBeth regression for each anomaly. The evaluated SDF models are: CAPM, CAPM Cay (Lettau and Ludvigson, 2001), Fama-French 3 (Fama and French,

1993), Fama-French 5 (Fama and French, 2015), Q Model (Hou, Xue, and Zhang, 2015), Mispricing (Stambaugh and Yuan, 2017), Behavioral (Daniel, Hirshleifer, and Sun, 2020), and AGI out-of-sample (this paper). For the AGI case, instead of using the same portfolios that are being evaluated for each anomaly, the M_t estimate uses the 25 Fama-French portfolios sorted by size and book-to-market value, i.e., an out-of-sample exercise.

Results in Table 5 show that AGI out-of-sample (AGI OOS) produces the highest $\overline{R}^2$ for 76% of the anomalies tested, CAPM Cay produces the highest $\overline{R}^2$ for 12% of these anomalies, and the rest of the highest $\overline{R}^2$ statistics are achieved by the Fama-French 5 model and by the Q model. In addition, the AGI out-of-sample is very robust to different anomalies and performs well in pricing even when it is not the best pricing model.

Issler and Moog also perform an in-sample forecasting exercise, presented in Table 6 in Appendix A.3. There, the AGI model (this paper) produces the highest $\overline{R}^2$ for 79% of the anomalies tested in Table 6. The Fama-French 5 and CAPM Cay models both produce the highest $\overline{R}^2$ for 9% of the anomalies, whereas the Q Model produces the highest $\overline{R}^2$ for 3% of them.

All in all, results show that employing a theory-based approach, based on a simple, non-parametric, non-linear, consistent estimate of M_t is able to price well the cross-sectional distribution of asset returns in-sample and out-of-sample. The approach follows closely the mimicking portfolio M_t^* framework of Hansen and Jagannathan – a function of return data alone. Regarding the number of factors, we do not increase their number to fit the cross-sectional distribution of returns. Rather, we increase indefinitely the number of assets used in constructing the single-factor M_t to fit our asymptotic setup. In that sense, our M_t estimate is an antidote to the *factor zoo* currently an important issue in the asset-pricing literature; see, e.g., Feng, Giglio, and Xiu (2020).

Table 5: Out-of-Sample (OOS) $\overline{R}^2$ Statistic for different M_t Estimates Across Anomalies

	CAPM	CAPM Cay	FF3	FF5	Q Model	Mispricing	Behavioral	AGI OOS
SUE-1	-0.07	0.81	0.68	0.80	0.79	0.88	0.65	0.99
SUE-6	-0.04	0.92	0.59	0.92	0.83	0.68	0.41	0.99
Abr-1	-0.11	0.80	-0.37	0.98	0.32	0.90	0.63	0.94
Abr-6	-0.08	0.49	-0.10	-0.24	0.54	0.25	0.70	0.74
R6-6	-0.12	0.95	0.85	0.98	0.99	0.92	0.98	0.95
RE-1	-0.12	0.88	0.45	0.43	0.83	0.62	0.58	0.96
RE-6	-0.04	0.97	0.57	0.84	0.83	0.77	0.67	0.90
R11-1	-0.09	0.96	0.80	0.99	0.91	0.84	0.81	0.96
IM6	-0.14	0.94	0.96	0.97	0.97	0.90	0.75	0.99
B/M	0.21	0.83	0.95	0.99	0.78	0.95	0.88	0.99
E/P	0.33	0.80	0.65	0.86	0.83	0.50	0.29	0.97
NO/P	0.18	0.65	0.36	0.55	0.21	0.52	0.83	0.93
Dur	0.34	0.98	0.64	0.57	0.91	0.71	0.55	0.96
CP	0.12	0.62	0.55	0.64	0.61	0.42	-0.14	0.96
I/A	-0.01	0.77	0.57	0.52	0.51	0.41	0.31	0.90
NOA	-0.09	0.90	0.03	0.72	0.57	0.77	0.66	0.99
$\Delta PI/A$	0.09	0.56	0.17	0.26	0.73	0.55	0.83	0.97
IG	-0.10	0.56	0.41	0.24	0.53	0.49	0.67	0.92
NSI	0.10	0.83	-0.10	0.40	0.75	0.68	0.74	0.98
CEI	0.14	0.70	0.41	0.22	0.71	0.55	0.82	0.95
IvG	-0.10	0.54	-0.16	0.14	0.60	0.47	0.58	0.92
IvC	-0.10	0.94	0.03	0.54	0.32	0.32	0.05	0.90
OA	0.40	0.74	0.53	0.56	0.33	0.33	0.84	0.96
POA	0.28	0.76	0.44	0.88	0.61	0.57	0.61	0.95
PTA	-0.12	0.61	0.00	0.27	0.77	0.41	0.63	0.79
ROE	0.61	0.99	0.95	0.99	0.91	0.94	0.83	0.99
ROA	0.66	0.87	0.73	0.74	0.85	0.76	0.78	0.99
GP/A	0.01	0.89	0.34	0.64	0.44	0.43	0.69	0.99
NEI	0.17	0.35	0.31	0.83	0.76	0.14	-0.10	0.99
FP6	0.51	0.89	0.51	0.41	0.90	0.70	0.82	0.93
OC/A	0.02	0.96	0.57	0.95	0.87	0.90	0.84	0.99
Ad/M	-0.06	0.83	0.62	0.84	0.86	0.66	0.59	0.96
RD/M	0.78	0.88	0.74	0.72	0.80	0.75	0.71	0.98
OL	0.01	0.81	0.16	0.76	0.76	0.93	0.69	0.99

Note: $\bar{R}^2$ in the second-stage Fama and MacBeth (1973) regression for each anomaly. Regressions employ excess returns as dependent variables. Models being evaluated: CAPM, CAPM Cay (Lettau and Ludvigson, 2001), FF3 (Fama and French, 1993), FF5 (Fama and French, 2015), Q Model (Hou, Xue, and Zhang, 2015), Mispricing (Stambaugh and Yuan, 2017), Behavioral (Daniel, Hirshleifer, and Sun, 2020), AGI out-of-sample (This paper). List of anomalies. Momentum anomalies: SUE-1, SUE-6, R6-6, R11-1, Abr-1, Abr-6, RE-1, RE-6, I-Mom. Value Vs. Growth anomalies: B/M, E/P, CF/P, NO/P, Dur. Investment anomalies: I/A, NOA, $\Delta PI/A$, IG, NSI, CEI, IvG, IvC, OA, POA, PTA. Profitability anomalies: ROE, ROA, GP/A, NEI, FP. Intangible anomalies: OC/A, Ad/M, RD/M, OL. Sources: Hou, Xue, and Zhang (2015) and Issler and Moog (2025).

4 Conclusion

This paper proposes a no-arbitrage approach to asset pricing with the following ingredients. First, we employ a novel identification strategy for a valid stochastic discount factor based on the Asset-Pricing Equation. Our strategy relies heavily on the cross-section dimension for priced returns, which plays a pivotal role rarely exploited in the literature. We also 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 common elements of (log) asset returns, allowing a common-idiosyncratic decomposition of these returns. Our key identifying assumption implies restrictions on this factor model, which can be tested to validate the former. Second, we derive an estimator of a valid SDF in a panel-data framework. We prove its consistency and asymptotic normality 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 SDF estimator depends exclusively on appropriate cross-sectional averages of asset returns (geometric and arithmetic), which makes its computation a simple and direct exercise.

The techniques discussed in this paper were applied to two issues in finance and one in macroeconomics. In our first application, we 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. We employed the *mixed-effect* panel-data model to assess the fit of our one-factor model to the data. We implemented a Wald test on the model's mean beta that verified the validity of our key identifying assumption – that the cross-sectional mean of the betas is equal to -1.

Our second application, extracted from Issler, Martins and Moog (2023), and from Issler and Moog (2025), investigated the ability of our proposed one-factor model to explain the cross-section of returns in-sample and out-of-sample by means of Fama-MacBeth regressions. In-sample results (Issler, Martins and Moog) show that our one-factor model has the best fit for the cross-section of returns considering the 25 Fama-French portfolios kept by the Fama-French library. Competitors included the CAPM, Lettau and Ludvigson's *cay* and *cay* augmented with labor income, and Fama-French's three- and five-factor models. Our estimate of M_t performed best overall in terms of $\overline{R}^2$ with monthly and quarterly data. Out-of-sample results (Issler and Moog) considered an extended exercise regrading pricing well-known anomalies, with an extended set of M_t estimates. Again, our single-factor approach showed the best out-of-sample fit overall regarding competitors using three- or five-factor models. So, in this case: *Less is More*.

Our last exercise – a mid-sized Monte-Carlo simulation, implemented in Appendix A.4 – uses a no-arbitrage dynamic (consumption) capital asset-pricing model with 100 time-series observations and $1,000$ cross-sectional observations to verify the small-sample fit our asymptotic M_t estimate. 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 across time and assets. The biases of the mean estimate and of the variance estimate are negligible in both cases, showing promise for its use in empirical work.

A Appendix

A.1 Proofs of the Main Results

In this section we provide proofs for the results in the paper.

Proof of Lemma 1.

From Assumptions 1 and joint log-Normality for $\ln(M_t)$ and $\ln(R_t)$, we have the following model

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

Take average over the cross-section

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

Notice that by Assumption 2 and applying the Minkowski's inequality to (3), for all i, $\sup_t \mathbb{E}\left[\varepsilon_{i,t}^2\right] < \infty$. Thus, from a WLLN for triangular arrays, $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it} \xrightarrow{p} \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,t}\right] \equiv \mathbb{E}^i\left[\varepsilon_{i,t}\right]$. Also, by the log-normality condition, $\mathbb{E}(\varepsilon_{i,t}) = 0$, for all i, and under Assumption 3, $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 = \overline{\sigma}^2 < \infty$. Thus, from the above display we obtain:

$$\mathbb{E}^i\left[r_{i,t}\right] = -m_t - \frac{1}{2}\overline{\sigma}^2,$$

such that

$$m_t + \frac{1}{2}\overline{\sigma}^2 = -\mathbb{E}^i\left[r_{i,t}\right].$$

Taking exponential on both sides

$$\exp\left(\frac{1}{2}\overline{\sigma}^2\right) M_t = \exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right),$$

such that

$$M_t = \frac{\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)}{\exp\left(\frac{1}{2}\overline{\sigma}^2\right)}.$$

Now, we have to write $\exp\left(\frac{1}{2}\overline{\sigma}^2\right)$ as a function of data. Multiply equation (2) in the main text by $\exp\left(\frac{1}{2}\overline{\sigma}^2\right)$, to obtain:

$$\mathbb{E}_{t-1}\left[\exp\left(\frac{1}{2}\overline{\sigma}^2\right) M_t R_{i,t}\right] = \exp\left(\frac{1}{2}\overline{\sigma}^2\right).$$

Take the unconditional mean of the previous equation across time and assets and use the law of iterated expectations to obtain:

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

Recall from above that $\exp\left(\frac{1}{2}\overline{\sigma}^2\right) M_t = \exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)$. So, we obtain:

$$\exp\left(\frac{1}{2}\overline{\sigma}^2\right) = \mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right) R_{i,t}\right].$$

Therefore, substituting back we are able to write M_t as a simple function of observables

as:

$$M_t = \frac{\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)}{\mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right) R_{i,t}\right]},$$

and from Assumption 3 the ratio is well defined, such that the result follows.

The second part of the lemma follows directly from the OLS projection in equation (13) together with Assumption 3. Section 2 above provides the details.

■

Before we establish the limiting properties of $\widehat{M_t}$, we recall the notation. Since we are dealing with panel data, we can compute mean, variances, and covariances across time and across assets. Across time, conditional means use the operator $\mathbb{E}_{t-1}\left[\cdot\right]$, unconditional means use the operator $\mathbb{E}\left[\cdot\right]$, and unconditional variances and covariances use the operator VAR$(\cdot)$ and COV$(\cdot)$, respectively. Across assets, unconditional means use the operator $\mathbb{E}^i\left[\varepsilon_{i,t}\right] \equiv \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \varepsilon_{i,t}\right]$.

Proof of Proposition 2. Recall the $\widehat{M_t}$ estimator in (22):

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

where

$$\overline{R}_t^G = \prod_{i=1}^N R_{i,t}^{-\frac{1}{N}} \quad \text{and} \quad \overline{R}_t^A = \frac{1}{N}\sum_{i=1}^N R_{i,t}.$$

First, consider the numerator of $\widehat{M_t}$. By Assumption 2, and a Weak Law of Large Numbers for triangular arrays (e.g., Theorem 2.2.6 in Durrett (2019, p.51)), we have that, for all t:

$$\frac{1}{N}\sum_{i=1}^N r_{i,t} \xrightarrow{p} \mathbb{E}^i\left[r_{i,t}\right].$$

Hence, by the above display, the Continuous Mapping Theorem, and equation (5) together with Assumption 3

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

Now, consider the denominator of the estimator:

$$\frac{1}{T}\sum_{j=1}^T \overline{R}_j^G \overline{R}_j^A = \frac{1}{T}\sum_{t=1}^T \left(\prod_{i=1}^N R_{i,t}^{-\frac{1}{N}} \frac{1}{N}\sum_{i=1}^N R_{i,t}\right) = \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N \left(\prod_{i=1}^N R_{i,t}^{-\frac{1}{N}} R_{i,t}\right).$$

By Lemma 4.3 in Newey and McFadden (1994) (see also Theorem 3.4 in Christensen

(2017)), the previous steps, and Assumption 2, we have that:

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\prod_{i=1}^{N}R_{i,t}^{-\frac{1}{N}}R_{i,t}\right)\xrightarrow{p}\mathbb{E}\left[\exp\left(\frac{1}{2}\overline{\sigma}^{2}\right)M_{t}R_{i,t}\right].$$

Take now the unconditional expectation of equation (2), use the law-of-iterated expectations, such that $\mathbb{E}\left[M_{t}R_{i,t}\right]=1$. From this result and the above display, we have that

$$\begin{aligned}\frac{1}{T}\sum_{j=1}^{T}\overline{R}_{j}^{G}\overline{R}_{j}^{A}&\xrightarrow{p}\exp\left(\frac{1}{2}\overline{\sigma}^{2}\right)\mathbb{E}\left[M_{t}R_{i,t}\right]\\&=\exp\left(\frac{1}{2}\overline{\sigma}^{2}\right)\times 1.\end{aligned}$$

Finally, by the Continuous Mapping Theorem:

$$\widehat{M}_{t}=\frac{\overline{R}_{t}^{G}}{\frac{1}{T}\sum_{j=1}^{T}\overline{R}_{j}^{G}\overline{R}_{j}^{A}}\xrightarrow{p}\frac{\exp\left(\frac{1}{2}\overline{\sigma}^{2}\right)M_{t}}{\exp\left(\frac{1}{2}\overline{\sigma}^{2}\right)}=M_{t}.$$

■

Proof of Proposition 3. Recall the $\widehat{M}_{t}$ estimator:

$$\widehat{M}_{t}=\frac{\overline{R}_{t}^{G}}{\frac{1}{T}\sum_{j=1}^{T}\overline{R}_{j}^{G}\overline{R}_{j}^{A}},$$

where

$$\overline{R}_{t}^{G}=\prod_{i=1}^{N}R_{i,t}^{-\frac{1}{N}}\quad\text{and}\quad\overline{R}_{t}^{A}=\frac{1}{N}\sum_{i=1}^{N}R_{i,t}.$$

By taking the logs on both sides, we are able to rewrite $\widehat{M}_{t}$ as:

$$\ln\widehat{M}_{t}=\ln\overline{R}_{t}^{G}-\ln\left(\frac{1}{T}\sum_{j=1}^{T}\overline{R}_{j}^{G}\overline{R}_{j}^{A}\right).\tag{31}$$

Now, notice that regarding the first term of equation (31), by Assumption 2 and WLLN for triangular arrarys, as $N\to\infty$, for all t, we obtain:

$$\ln\overline{R}_{t}^{G}=\ln\left(\prod_{i=1}^{N}R_{i,t}^{-\frac{1}{N}}\right)=-\frac{1}{N}\sum_{i=1}^{N}r_{i,t}\xrightarrow{p}-\mathbb{E}^{i}\left[r_{i,t}\right],$$

such that

$$\overline{R}_{t}^{G}\xrightarrow{p}\exp\left(-\mathbb{E}^{i}\left[r_{i,t}\right]\right):=R_{t}^{G}.$$

In the same lines, for $\overline{R}_t^A$, we have that

$$\overline{R}_t^A = \frac{1}{N}\sum_{i=1}^{N} R_{i,t} \xrightarrow{p} \mathbb{E}^i\left[R_{i,t}\right] := R_t^A.$$

Note also that, by Assumption 2 and WLLN it follows from routine arguments (e.g., Lemma 4.3 of Newey and McFadden (1994) and Christensen (2017), replacing the law of large numbers by the ergodic theorem), that:

$$\frac{1}{T}\sum_{j=1}^{T}\overline{R}_j^G\overline{R}_j^A = \frac{1}{T}\sum_{t=1}^{T}\left[\left(\prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}}\right)\frac{1}{N}\sum_{i=1}^{N} R_{i,t}\right] = \frac{1}{T}\sum_{j=1}^{T} R_j^G R_j^A + o_p(1),$$

and thus, as $T \to \infty$:

$$\frac{1}{T}\sum_{j=1}^{T}\overline{R}_j^G\overline{R}_j^A \xrightarrow{p} \mathbb{E}\left[R_t^G R_t^A\right].$$

Thus, we are able to define the following population version of the estimator in (31) as:

$$\ln M_t = -\mathbb{E}^i\left[r_{i,t}\right] - \ln\left(\mathbb{E}\left[R_t^G R_t^A\right]\right).$$

The task now is to establish the asymptotic distribution of the estimator. From equation (31), subtract $\ln M_t$ on both sides, and then multiply both sides by $\sqrt{T}$ to obtain:

$$\sqrt{T}\left(\ln\widehat{M}_t - \ln M_t\right) = \sqrt{\frac{T}{N}}\sqrt{N}\left(\ln\overline{R}_t^G + \mathbb{E}^i\left[r_{i,t}\right]\right) - \sqrt{T}\left[\ln\left(\frac{1}{T}\sum_{j=1}^{T}\overline{R}_j^G\overline{R}_j^A\right) - \ln\left(\mathbb{E}\left[R_t^G R_t^A\right]\right)\right]. \tag{32}$$

Thus, using the previous result, we can write (32) as

$$\begin{aligned}\sqrt{T}\left(\ln\widehat{M}_t - \ln M_t\right) = {} & \sqrt{\frac{T}{N}}\sqrt{N}\left(\ln\overline{R}_t^G + \mathbb{E}^i\left[r_{i,t}\right]\right) \\ & - \sqrt{T}\left[\ln\left(\frac{1}{T}\sum_{j=1}^{T} R_j^G R_j^A\right) - \ln\left(\mathbb{E}\left[R_t^G R_t^A\right]\right)\right] + o_p(1).\end{aligned} \tag{33}$$

We study both terms on the right hand side of (33). Consider the first term. As $N \to \infty$, by Assumptions 1 and 2 and a CLT for triangular arrays (e.g., Theorem 3.4.10 in Durrett (2019, p.128)), we have the following convergence:

$$\sqrt{N}\left(\ln\overline{R}_t^G + \mathbb{E}^i\left[r_{i,t}\right]\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_t^G), \tag{34}$$

where

$$V_t^G = \mathbb{E}^i\left[\left(\ln R_{it} - \mathbb{E}^i \ln R_{it}\right)^2\right].$$

We have used a time index to indicate that V_t^G could be time-dependent.

Next, note that by assumption as $T/N \to 0$, such that

$$\sqrt{\frac{T}{N}} \to 0,$$

and from equation (34), $\sqrt{N}\left(\ln \overline{R}_t^G + \mathbb{E}^i\left[r_{i,t}\right]\right)$ is bounded in probability. Therefore, as $(N,T) \to \infty$ with $T/N \to 0$, the first term in (33) is negligible:

$$\sqrt{\frac{T}{N}}\sqrt{N}\left(\ln \overline{R}_t^G + \mathbb{E}^i\left[r_{i,t}\right]\right) = o(1)O_p(1).$$

Next, we examine the second term in (33), which can be written as

$$\sqrt{T}\left(\ln \widehat{M}_t - \ln M_t\right) = -\sqrt{T}\left[\ln\left(\frac{1}{T}\sum_{t=1}^{T} R_t^G R_t^A\right) - \ln\left(\mathbb{E}\left[R_t^G R_t^A\right]\right)\right] + o_p(1).$$

Consider the convergence property of the following term. As $T \to \infty$, applying a suitable CLT we obtain:

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} R_t^G R_t^A - \mathbb{E}\left[R_t^G R_t^A\right]\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V^{GA}),$$

where

$$V^{GA} \;=\; \text{VAR}\left(\frac{1}{T}\sum_{t=1}^{T} \overline{R}_t^G \overline{R}_t^A\right).$$

Then, by the Delta Method, we obtain:

$$\sqrt{T}\left(\ln \widehat{M}_t - \ln M_t\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, V\right),$$

where $V = \frac{V^{GA}}{\left(\mathbb{E}\left[R_t^G R_t^A\right]\right)^2}$, and (23) offers a consistent estimate of V^{GA}.

To derive the asymptotic distribution of the level $\widehat{M}_t$, instead of its logarithm $\left(\ln \widehat{M}_t\right)$, we employ the Delta Method to obtain

$$\sqrt{T}\left(\widehat{M}_t - M_t\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, M_t^2 V),$$

■

A.2 A General Identification Strategy for a Valid SDF

In this section we establish identification of M_t under more general conditions, that is, without the log-Normal assumption used in Section 2 of the main text.

Here, we first provide an auxiliary results that employs an exact Taylor-Expansion of the exponential function in order to characterize the Asset-Pricing Equation (2). The next lemma collects the results.

Lemma 4 *Under Assumptions 1-2, we can rewrite equation* (3) *as:*

$$r_{i,t} = -m_t - \gamma_i^2 + \sum_{j=0}^{\infty} b_{i,j} \upsilon_{i,t-j}, \qquad i = 1, 2, \ldots, N. \tag{35}$$

where $\upsilon_{i,t}$ *is white noise, the error term,* $\sum_{j=0}^{\infty} b_{i,j} \upsilon_{i,t-j}$*, potentially has serial correlation, and*

$$-\gamma_i^2 = \mathbb{E}\left[m_t\right] + \mathbb{E}\left[r_{i,t}\right].$$

Proof of Lemma 4. To establish identification of M_t under general assumptions, we consider a second-order Taylor Expansion of the exponential function around x, with increment h, as follows:

$$\begin{aligned} e^{x+h} &= e^x + he^x + \frac{h^2 e^{x+\lambda(h)\cdot h}}{2}, \\ \text{where} \quad & \lambda(h) : \mathbb{R} \to (0,1). \end{aligned} \tag{36}$$

It is important to stress that (36) 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 (36) 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 (36) by e^x yields:

$$e^h = 1 + h + \frac{h^2 e^{\lambda(h)\cdot h}}{2}, \tag{37}$$

showing that (37) does not depend on x. Therefore, we obtain 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 (37) with the Pricing Equation (2), we let $h = \ln(M_t R_{i,t})$ in (37) to obtain:

$$M_t R_{i,t} = 1 + \ln(M_t R_{i,t}) + \frac{\left[\ln(M_t R_{i,t})\right]^2 e^{\lambda(\ln(M_t R_{i,t}))\cdot \ln(M_t R_{i,t})}}{2}, \tag{38}$$

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 (38), namely $z_{i,t}$, which, due to its formula, is a non-negative random variable:

$$z_{i,t} \equiv \frac{1}{2} \times \left[\ln(M_t R_{i,t})\right]^2 e^{\lambda(\ln(M_t R_{i,t}))\cdot\ln(M_t R_{i,t})} \geq 0, \qquad i = 1, 2, \ldots, N.$$

Taking the conditional expectation of both sides of (38) gives:

$$\begin{aligned} \mathbb{E}_{t-1}\left[M_t R_{i,t}\right] &= 1 + \mathbb{E}_{t-1}\left[\ln(M_t R_{i,t})\right] + \mathbb{E}_{t-1}\left[z_{i,t}\right], \text{ or,} \\ 0 &= \mathbb{E}_{t-1}\left[M_t R_{i,t}\right] - 1 = \mathbb{E}_{t-1}\left[\ln(M_t R_{i,t})\right] + \mathbb{E}_{t-1}\left[z_{i,t}\right], \end{aligned} \tag{39}$$

where (39) is a direct consequence of the Asset-Pricing equation (2), since its left-hand side cancels out – $\mathbb{E}_{t-1}\left[M_t R_{i,t}\right] - 1 = 0$ – and then:

$$\mathbb{E}_{t-1}\left[z_{i,t}\right] = -\mathbb{E}_{t-1}\left[\ln(M_t R_{i,t})\right]. \tag{40}$$

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 $\mathbb{E}_{t-1}\left[\ln(M_t R_{i,t})\right]$ under the Pricing Equation. 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.

The dynamic setup that we work with next comes from the Wold representation for $m_t + r_{i,t}$. From Assumption 2, $\{\ln(M_t R_{i,t})\}$ is a covariance-stationary processes with finite first and second moments. So, by the Wold Theorem, it has a Wold Representation as follows:

$$m_t + r_{i,t} = \mathbb{E}\left[m_t\right] + \mathbb{E}\left[r_{i,t}\right] + \sum_{j=0}^{\infty} b_{i,j} \upsilon_{i,t-j}, \qquad i = 1, 2, \ldots, N. \tag{41}$$

This can be written as, (42)

$$r_{i,t} = -m_t - \gamma^2_{i,t|t-1} + \upsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \tag{43}$$

where, for all i, $b_{i,0} = 1$, $\sum_{j=0}^{\infty} b^2_{i,j} < \infty$, $\upsilon_{i,t}$ is a white-noise process, and, due to (40):

$$\begin{aligned} -\gamma^2_{i,t|t-1} &= -\mathbb{E}_{t-1}\left[z_{i,t}\right] = \mathbb{E}_{t-1}\left[m_t + r_{i,t}\right] \\ &= \mathbb{E}\left[m_t\right] + \mathbb{E}\left[r_{i,t}\right] + \sum_{j=1}^{\infty} b_{i,j} \upsilon_{i,t-j}. \end{aligned}$$

Note that equation (43) generalizes equation (3), where for the latter we had assumed log-Normality. But, to derive (43), we first use the Wold Representation for $m_t + r_{i,t}$ in equation (42) and then derive $-\gamma^2_{i,t|t-1} = -\mathbb{E}_{t-1}\left[z_{i,t}\right] = \mathbb{E}_{t-1}\left[m_t + r_{i,t}\right]$ consistent with it. Indeed, we have combined two less restrictive ingredients:

1. The Wold representation for $m_t + r_{i,t}$ in equation (42);

2. The restriction that $\mathbb{E}_{t-1}\left[M_t R_{i,t}\right] - 1 = 0$, imposed on the exact Taylor expan-

sion in (38), and then applied to (42), which yields $-\gamma_{i,t|t-1}^2 = -\mathbb{E}_{t-1}[z_{i,t}] = \mathbb{E}_{t-1}[m_t + r_{i,t}] = \mathbb{E}[m_t] + \mathbb{E}[r_{i,t}] + \sum_{j=1}^{\infty} b_{i,j}\upsilon_{i,t-j}$.

To generate an equation in the spirit of (3), we have defined the time invariant part of $-\gamma_{i,t|t-1}^2$ as:

$$-\gamma_i^2 = \mathbb{E}[m_t] + \mathbb{E}[r_{i,t}].$$

If we incorporate $\sum_{j=1}^{\infty} b_{i,j}\upsilon_{i,t-j}$ into the error term, that leads to equation (35), which generalizes (3):

$$r_{i,t} = -m_t - \gamma_i^2 + \sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}, \qquad i = 1, 2, \ldots, N.$$

where the error term, $\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}$, potentially has serial correlation, capturing the dynamic nature of $\gamma_{i,t|t-1}^2$.

■

The next result is based on the previous lemma. We employ the joint weak-stationarity assumption to discuss identification in a *reduced-form* sense and in a *structural-form* sense.

Lemma 5 *Under the assumptions of Lemma 4 and under the additional identifying assumption that* $\mathbb{E}^i\left[\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}\right] = 0$ *and* $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \gamma_i^2 = \overline{\gamma}^2 < \infty$ *in* (35)*, we can identify* M_t *as:*

$$M_t = \frac{\exp\left(-\mathbb{E}^i[r_{i,t}]\right)}{\mathbb{E}\left[\exp\left(-\mathbb{E}^i[r_{i,t}]\right) R_{i,t}\right]},$$

where the expectation in the denominator is taken across assets and across time. Moreover, these assumptions ensure that there exists a single-factor model for $r_{i,t}$*, as follows:*

$$\begin{aligned} r_{i,t} &= -\left(\gamma_i^2 + \delta_{i,m}\mathbb{E}[m_t]\right) + \beta_i m_t + \xi_{i,t}, \qquad i = 1, 2, \ldots, N, \\ \text{where } \beta_i &= -1 + \delta_{i,m} \qquad i = 1, 2, \ldots, N. \end{aligned} \tag{44}$$

The term $\delta_{i,m}$ *is defined as:*

$$\delta_{i,m} = \frac{\text{COV}\left(\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}, m_t\right)}{\sigma_m^2}, \text{ for all } i = 1, 2, \ldots, N, \tag{45}$$

and it plays the same role as δ_i *in the main text; errors* $(\xi_{i,t})$ *are orthogonal to the single factor* (m_t)*; there is heterogeneity in the intercept* $(\gamma_i^2 + \delta_{i,m}\mathbb{E}[m_t])$ *and slope* $(\beta_i = -1 + \delta_{i,m})$*, although they are respectively constrained by* $\mathbb{E}^i[\delta_{i,m}] = 0$*, and* $\mathbb{E}^i[\beta_i] = -1$*; and the error term has zero mean across time* $(\mathbb{E}[\xi_{i,t}] = 0)$ *and assets* $(\mathbb{E}^i[\xi_{i,t}] = 0)$*.*

Proof of Lemma 5.

To obtain *reduced-from* identification, take now the cross-sectional mean of (35), together with the WLLN, to achieve:

$$\mathbb{E}^i\left[r_{i,t}\right] = -m_t - \overline{\gamma}^2 + \mathbb{E}^i\left[\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}\right].$$

With the assumption stated in Lemma 5 that $\mathbb{E}^i\left[\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}\right] = 0$, we get:

$$\mathbb{E}^i\left[r_{i,t}\right] = -m_t - \overline{\gamma}^2.$$

By exponentiating both sides, we can identify M_t as before, as long as we are able to identify $\exp\left(\overline{\gamma}^2\right)$ – a constant. But this can be done as in equation (6) above:

$$M_t = \exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)\exp\left(-\overline{\gamma}^2\right) = \frac{\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)}{\exp\left(\overline{\gamma}^2\right)}.$$

Note that the role previously played by $\sigma_i^2/2$ is now played by γ_i^2, but otherwise all else is the same, and we are able to follow the same steps as before. Multiply equation (2) in the main text by $\exp\left(\overline{\gamma}^2\right)$, to obtain:

$$\mathbb{E}_{t-1}\left[\exp\left(\overline{\gamma}^2\right) M_t R_{i,t}\right] = \exp\left(\overline{\gamma}^2\right).$$

Take the unconditional mean of the previous equation across *assets* and *time* and use the law of iterated expectations. Recall that $\exp\left(\overline{\gamma}^2\right) M_t = \exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)$ to obtain:

$$\exp\left(\overline{\gamma}^2\right) = \mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right) R_{i,t}\right]. \tag{46}$$

Therefore, we are able to write M_t as a simple non-linear function of observables alone:

$$M_t = \frac{\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right)}{\mathbb{E}\left[\exp\left(-\mathbb{E}^i\left[r_{i,t}\right]\right) R_{i,t}\right]},$$

and the identification result follows.

We now establish the single-factor model representation, i.e., what we have labeled as *structural-form* identification. As before, we will proceed by performing an OLS projection of:

$$\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}, \qquad i = 1, 2, \ldots, N, \tag{47}$$

onto $\left(m_t - \mathbb{E}\left[m_t\right]\right)$ in order to obtain a diversifiable error term in the cross-sectional dimension. Despite the dynamic nature of (47), the OLS projection will produce an idiosyncratic error term, devoid of elements captured only by m_t.

To deal with structural-form identification, we re-write equation (42) as:

$$(m_t - \mathbb{E}[m_t]) + (r_{i,t} - \mathbb{E}[r_{i,t}]) = \sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}, \qquad i = 1, 2, \ldots, N. \tag{48}$$

Take the cross-sectional average of (48), employing Assumptions 2, $\mathbb{E}^i\left[\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}\right] = 0$ and use a WLLN to obtain:

$$(m_t - \mathbb{E}[m_t]) + \mathbb{E}^i[r_{i,t} - \mathbb{E}[r_{i,t}]] = 0.$$

As before, this establishes that:

$$\begin{aligned} \text{COV}\left(\mathbb{E}^i[r_{i,t}], m_t\right) &= -\sigma_m^2 = \text{COV}(-m_t, m_t), \text{ and,} & (49)\\ \text{VAR}\left(\mathbb{E}^i[r_{i,t}]\right) &= \text{VAR}(m_t) = \sigma_m^2. & (50) \end{aligned}$$

To obtain a proper factor model for $r_{i,t}$, start from equation (35) and consider now the OLS projection of $\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}$ onto $(m_t - \mathbb{E}[m_t])$:

$$\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j} = \delta_{i,m}(m_t - \mathbb{E}[m_t]) + \xi_{i,t}, \qquad ,$$

where $\delta_{i,m}$ is defined in equation (45). This allows re-writing (35) as a proper factor model with orthogonality between m_t and $\xi_{i,t}$:

$$\begin{aligned} r_{i,t} &= -\left(\gamma_i^2 + \delta_{i,m}\mathbb{E}[m_t]\right) + \beta_i m_t + \xi_{i,t} \qquad i = 1, 2, \ldots, N,\\ \text{where } \beta_i &= -1 + \delta_{i,m} \qquad i = 1, 2, \ldots, N. \end{aligned} \tag{51}$$

A key consequence of the identifying Assumption 3 is the following:

$$\begin{aligned} \mathbb{E}^i[\delta_{i,m}] &= \frac{\mathbb{E}^i\left[\text{COV}\left(m_t, \sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}\right)\right]}{\sigma_m^2} = \frac{\mathbb{E}^i\text{COV}(m_t, (m_t + r_{i,t}))}{\sigma_m^2} \\ &= \frac{\sigma_m^2 + \mathbb{E}^i[\text{COV}(m_t, r_{i,t})]}{\sigma_m^2} = 1 + \frac{\text{COV}(m_t, \mathbb{E}^i[r_{i,t}])}{\sigma_m^2} = 0, \end{aligned} \tag{52}$$

where the last line uses the result in (49). So:

$$\mathbb{E}^i[\beta_i] = -1 + \mathbb{E}^i[\delta_{i,m}] = -1.$$

It is important to emphasize that controlling for the factor only requires to compute the OLS projection of $\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}$ onto $(m_t - \mathbb{E}[m_t])$. This fits exactly the same procedure done in Section 2.1.

Finally, since:

$$\xi_{i,t} = \sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j} - \delta_{i,m}\left(m_t - \mathbb{E}\left[m_t\right]\right), \tag{53}$$

because $\mathbb{E}^i\left[\delta_{i,m}\right] = 0$, and $\mathbb{E}^i\left[\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}\right] = 0$, we have:

$$\mathbb{E}^i\left[\xi_{i,t}\right] = 0, \text{ for all } t \text{ as well.}$$

Across time, $\xi_{i,t}$ also has a zero mean, which can be seen by taking the mean of (53) across time of to obtain $\mathbb{E}\left[\xi_{i,t}\right] = 0$.

Therefore, we end up with a proper factor model, where errors $(\xi_{i,t})$ are orthogonal to the single factor (m_t); where there is heterogeneity in the intercept $(\gamma_i^2 + \delta_{i,m}\mathbb{E}\left[m_t\right])$ and slope $(\beta_i = -1 + \delta_{i,m})$, although they are respectively constrained by $\mathbb{E}^i\left[\delta_{i,m}\right] = 0$, and $\mathbb{E}^i\left[\beta_i\right] = -1$. Finally, the error term has zero mean across time $(\mathbb{E}\left[\xi_{i,t}\right] = 0)$ and assets $(\mathbb{E}^i\left[\xi_{i,t}\right] = 0)$. Indeed, these are consequences of our key Assumptions 1-2 and the added assumption that $\mathbb{E}^i\left[\sum_{j=0}^{\infty} b_{i,j}\upsilon_{i,t-j}\right] = 0$.

■

Proposition 6 *Using the setup of Hansen and Singleton (1983), Assumptions 1-3, as* $(N,T) \to \infty$*, and* $T/N \to k < \infty$*, the following holds:*

$$\sqrt{T}\left(\widehat{M}_t - M_t\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, M_t^2 V_t),$$

where $V_t := \lim_{N,T\to\infty} \text{VAR}\left(\ln \overline{R}_t^G - \ln\left(\frac{1}{T}\sum_{j=1}^{T} \overline{R}_j^G \overline{R}_j^A\right)\right)$.

Proof of Proposition 6. Recall the estimator:

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

where

$$\overline{R}_t^G = \prod_{i=1}^{N} R_{i,t}^{-\frac{1}{N}} \quad \text{and} \quad \overline{R}_t^A = \frac{1}{N}\sum_{i=1}^{N} R_{i,t}.$$

The proof follows the same lines of that of Proposition 3. The task now is to establish the asymptotic distribution of the estimator. Recall the linear representation for $\ln \widehat{M}_t - \ln M_t$ in equation (33), which can be written as following:

$$\begin{aligned}\sqrt{T}\left(\ln \widehat{M}_t - \ln M_t\right) = &\sqrt{\frac{T}{N}}\sqrt{N}\left(\ln \overline{R}_t^G + \mathbb{E}^i\left[r_{i,t}\right]\right) \\ &- \sqrt{T}\left[\ln\left(\frac{1}{T}\sum_{j=1}^{T} R_j^G R_j^A\right) - \ln\left(\mathbb{E}\left[R_t^G R_t^A\right]\right)\right] + o_p(1).\end{aligned} \tag{54}$$

We study both terms on the right hand side of (54).

Consider the first term. Recall that $\ln \overline{R}_t^G = -\frac{1}{N}\sum_{i=1}^N r_{i,t}$, such that by Assumptions 1 and 2 and a CLT for triangular arrays (see, e.g., Durrett 2019, Theorem 3.4.10), we have the following convergence:

$$\sqrt{N}\left(\ln \overline{R}_t^G + \mathbb{E}^i\left[r_{i,t}\right]\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_t^G), \tag{55}$$

where

$$V_t^G \;=\; \text{VAR}\left(\frac{1}{N}\sum_{i=1}^N \ln R_{it}\right).$$

As $N, T \to \infty$ with $T/N \to k$, note that (55) converges to a normal distribution

$$\sqrt{N}\left(\ln \overline{R}_t^G + \mathbb{E}^i\left[r_{i,t}\right]\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_t^G/k^2). \tag{56}$$

We now examine the convergence in law of the second term on the right hand side of equation (54).

Recall from the proof of Proposition 3 above that, as $T \to \infty$, by applying a suitable CLT we obtain:

$$\sqrt{T}\left(\ln\left(\frac{1}{T}\sum_{t=1}^T R_t^G R_t^A\right) - \ln\left(\mathbb{E}\left[R_t^G R_t^A\right]\right)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, V\right), \tag{57}$$

where $\frac{V^{GA}}{\left(\mathbb{E}\left[R_t^G R_t^A\right]\right)^2}$ with

$$V^{GA} \;=\; \text{VAR}\left(\frac{1}{T}\sum_{t=1}^T \overline{R}_t^G \overline{R}_t^A\right).$$

From the convergence of the two terms above in (56) and (57), as $(N, T) \to \infty$, we have that

$$\sqrt{T}\left(\ln \widehat{M}_t - \ln M_t\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_t),$$

where $V_t = \text{VAR}\left(\ln \overline{R}_t^G - \ln\left(\frac{1}{T}\sum_{j=1}^T \overline{R}_j^G \overline{R}_j^A\right)\right)$.

Note that

$$V_t = \frac{1}{k^2}V_t^G + V - 2\rho\sqrt{\frac{1}{k^2}V^G V}.$$

Finally, to derive the asymptotic distribution of the level $\widehat{M}_t$, instead of its logarithm $\left(\ln \widehat{M}_t\right)$, one could employ the Delta Method,

$$\sqrt{T}\left(\widehat{M}_t - M_t\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_t M_t^2).$$

■

A.3 The Cross-Sectional Distribution of Asset Returns: A Comprehensive In-Sample Evaluation

This section presents the in-sample analysis, analogous to the out-of-sample analysis in Table 5 discussed in Section 3.2 of the main text. In-sample evaluation means that, whenever possible, the same return data used in constructing SDFs are explained by the Fama-MacBeth regressions.

Table 6, extracted from Issler and Moog (2025), displays the $\overline{R}^2$ statistic of the second-stage Fama and MacBeth regression for each anomaly investigated by Hou, Xue, and Zhang (2015). The evaluated SDF models are: CAPM, CAPM Cay (Lettau and Ludvigson, 2001), Fama-French 3 (Fama and French, 1993), Fama-French 5 (Fama and French, 2015), Q Model (Hou, Xue, and Zhang, 2015), Mispricing (Stambaugh and Yuan, 2017), Behavioral (Daniel, Hirshleifer, and Sun, 2020), and AGI in-sample (this paper). For the latter, data on the portfolios being tested for each anomaly are used to construct the estimate of M_t, therefore the label in-sample.

Overall, AGI in-sample produces the highest $\overline{R}^2$ for 79% of the anomalies tested in Table 6. The Fama-French 5 and CAPM Cay models both produce the highest $\overline{R}^2$ for 9% of the anomalies, whereas the Q Model produces the highest $\overline{R}^2$ for 3% of them. We want to highlight two relevant results for evaluating these different SDF models. First, the SDF estimate proposed here delivers very robust results across anomalies, rarely below 0.9. Indeed, its average $\overline{R}^2$ across anomalies is 0.95, above the second SDF model (CAPM Cay), whose average $\overline{R}^2$ is 0.79, and way above all other SDF models, whose median $\overline{R}^2$ is 0.64. Second, even when AGI in-sample does not beat an alternative SDF model, its $\overline{R}^2$ is close to the leading one.

Table 6: In-Sample $\overline{R}^2$ Statistic for different M_t Estimates Across Anomalies

	CAPM	CAPM Cay	FF 3	FF 5	Q Model	Mispricing	Behavioral	AGI In Sample
SUE-1	-0.07	0.81	0.68	0.80	0.79	0.88	0.65	0.99
SUE-6	-0.04	0.92	0.59	0.92	0.83	0.68	0.41	0.99
Abr-1	-0.11	0.80	-0.37	0.98	0.32	0.90	0.63	0.94
Abr-6	-0.08	0.49	-0.10	-0.24	0.54	0.25	0.70	0.73
R6-6	-0.12	0.95	0.85	0.98	0.99	0.92	0.98	0.94
RE-1	-0.12	0.88	0.45	0.43	0.83	0.62	0.58	0.97
RE-6	-0.04	0.97	0.57	0.84	0.83	0.77	0.67	0.94
R11-1	-0.09	0.96	0.80	0.99	0.91	0.84	0.81	0.96
IM6	-0.14	0.94	0.96	0.97	0.97	0.90	0.75	0.99
B/M	0.21	0.83	0.95	0.99	0.78	0.95	0.88	0.98
E/P	0.33	0.80	0.65	0.86	0.83	0.50	0.29	0.98
NO/P	0.18	0.65	0.36	0.55	0.21	0.52	0.83	0.92
Dur	0.34	0.98	0.64	0.57	0.91	0.71	0.55	0.98
CP	0.12	0.62	0.55	0.64	0.61	0.42	-0.14	0.98
I/A	-0.01	0.77	0.57	0.52	0.51	0.41	0.31	0.92
NOA	-0.09	0.90	0.03	0.72	0.57	0.77	0.66	0.98
$\Delta PI/A$	0.09	0.56	0.17	0.26	0.73	0.55	0.83	0.98
IG	-0.10	0.56	0.41	0.24	0.53	0.49	0.67	0.93
NSI	0.10	0.83	-0.10	0.40	0.75	0.68	0.74	0.97
CEI	0.14	0.70	0.41	0.22	0.71	0.55	0.82	0.95
IvG	-0.10	0.54	-0.16	0.14	0.60	0.47	0.58	0.91
IvC	-0.10	0.94	0.03	0.54	0.32	0.32	0.05	0.92
OA	0.40	0.74	0.53	0.56	0.33	0.33	0.84	0.93
POA	0.28	0.76	0.44	0.88	0.61	0.57	0.61	0.97
PTA	-0.12	0.61	0.00	0.27	0.77	0.41	0.63	0.81
ROE	0.61	0.99	0.95	0.99	0.91	0.94	0.83	0.98
ROA	0.66	0.87	0.73	0.74	0.85	0.76	0.78	0.98
GP/A	0.01	0.89	0.34	0.64	0.44	0.43	0.69	0.97
NEI	0.17	0.35	0.31	0.83	0.76	0.14	-0.10	0.99
FP6	0.51	0.89	0.51	0.41	0.90	0.70	0.82	0.94
OC/A	0.02	0.96	0.57	0.95	0.87	0.90	0.84	0.98
Ad/M	-0.06	0.83	0.62	0.84	0.86	0.66	0.59	0.98
RD/M	0.78	0.88	0.74	0.72	0.80	0.75	0.71	0.99
OL	0.01	0.81	0.16	0.76	0.76	0.93	0.69	0.99

Note: $\overline{R}^2$ in the second-stage Fama and MacBeth (1973) regression for each anomaly. Models being evaluated: CAPM, CAPM Cay (Lettau and Ludvigson, 2001), FF 3 (Fama and French, 1993), FF 5 (Fama and French, 2015), Q Model (Hou, Xue, and Zhang, 2015), Mispricing (Stambaugh and Yuan, 2017), Behavioral (Daniel, Hirshleifer, and Sun, 2020), AGI (This paper). List of anomalies. Momentum anomalies: SUE-1, SUE-6, R6-6, R11-1, Abr-1, Abr-6, RE-1, RE-6, I-Mom. Value Vs. Growth anomalies: B/M, E/P, CF/P, NO/P, Dur. Investment anomalies: I/A, NOA, $\Delta PI/A$, IG, NSI, CEI, IvG, IvC, OA, POA, PTA. Profitability anomalies: ROE, ROA, GP/A, NEI, FP. Intangible anomalies: OC/A, Ad/M, RD/M, OL. Sources: Hou, Xue, and Zhang (2015) and Issler, Martins and Moog (2025).

A.4 A Macro-Finance Simulation Study

This section uses numerical simulations to assess the finite sample performance of the proposed method. In particular, we investigate the empirical bias for the stochastic discount factor (SDF) estimator and for its variance. We generate data using a no-arbitrage dynamic model and employ it to investigate how close our proposed SDF estimate is to the actual SDF across simulations. We employ the (consumption) capital-asset-pricing model (CCAPM) setup of Hansen and Singleton (1983), where they assume that $\ln(M_t \mathbf{R}_t)$ has a conditional multivariate *Normal* distribution with homoskedastic variance in the time dimension. 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. These assumptions give us a log-linear system on $r_{i,t}$ and $\Delta \ln(c_t)$, $i = 1, \ldots, N$, as follows:

$$r_{i,t} - \gamma \Delta \ln(c_t) = -\ln(\beta) - \frac{\sigma_i^2}{2} + \varepsilon_{i,t}, \qquad i = 1, \ldots, N. \tag{58}$$

Hansen and Singleton show that the dynamic representation of the data follows a vector autoregression VAR(p) process. 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 $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\mathbf{v})$, and:

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

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

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, i.e., transmitted to $(r_{1,t}, r_{2,t}, \cdots, r_{N,t})'$, 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:

$$\mathbf{X}_t = \mathbf{A}_0^{-1}\mathbf{c}_0 + \mathbf{A}_0^{-1}\begin{pmatrix} \mathbf{0} \\ a_1 \end{pmatrix}\mathbf{X}_{t-1} + \cdots + \mathbf{A}_0^{-1}\begin{pmatrix} \mathbf{0} \\ a_p \end{pmatrix}\mathbf{X}_{t-p} + \mathbf{A}_0^{-1}\mathbf{v}_t, \quad \text{or,} \tag{59}$$

$$\mathbf{X}_t = \mathbf{c} + \mathbf{B}_1\mathbf{X}_{t-1} + \mathbf{B}_2\mathbf{X}_{t-2} + \cdots + \mathbf{B}_p\mathbf{X}_{t-p} + \eta_t, \tag{60}$$

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 $\eta_t = \mathbf{A}_0^{-1}\mathbf{v}_t$. It is important to note that $\mathbf{\Sigma_v}$ 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, despite $\mathbf{\Sigma_v}$ being a diagonal matrix, the reduced form variance-covariance matrix will not be diagonal, since $\mathbb{E}[\eta_t\eta_t'] = \left(\mathbf{A}_0^{-1}\right)\mathbf{\Sigma_v}\left(\mathbf{A}_0^{-1}\right)'$. 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 (60), 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 (14).

To simplify matters in our simulation, we employed a VAR(p) of order one, i.e., $p = 1$. The parameters of interest were set according to standard GMM estimates using U.S. quarterly data. The estimate for β was set as $\beta = 0.999$, to mimic monthly frequency estimates, whereas we set $\gamma = 1.30$. We have also set $\mathbb{E}[\Delta\ln(c_t)] = 0.0021$ at the monthly frequency, which yields an annual growth rate of 2.55%. We chose $\text{VAR}(\Delta\ln(c_t)) = 1.78E-5$.

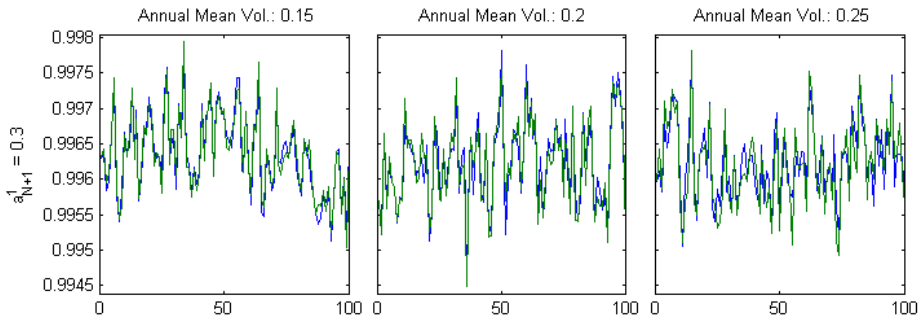
After a burning period of 30 months, we simulate the system with $N = 1,000$ assets, and $T = 100$ monthly observations. In our preferred setting, the autoregressive coefficient for consumption growth is set to 0.3, i.e., $a_{N+1}^1 = 0.3$, whereas the other coefficients in the last line of $\begin{pmatrix} \mathbf{0} \\ a_1 \end{pmatrix}$ are all relatively small, selected to be close to 10^{-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 – Figures 2, 3, 4, and 5. The first shows our benchmark case, with $a^1_{N+1} = 0.3$.

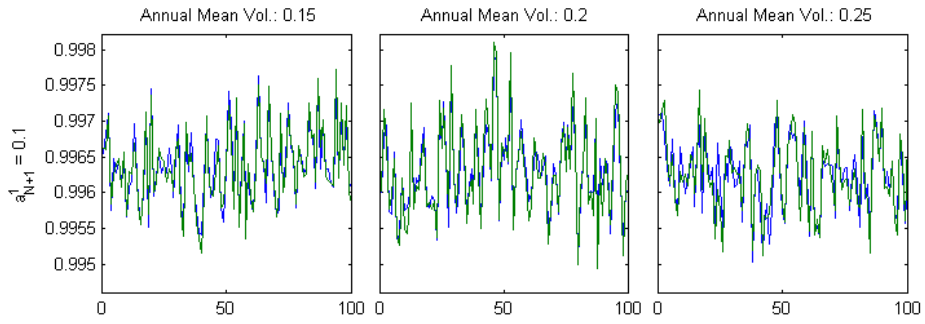
Figure 2: Average Simulated Stochastic Discount Factor of CCAPM (Blue) and Average Estimate in This Paper (Green).



Note: Monte-Carlo simulations using model described in equations (58)–(60). The total number of simulations is 100.

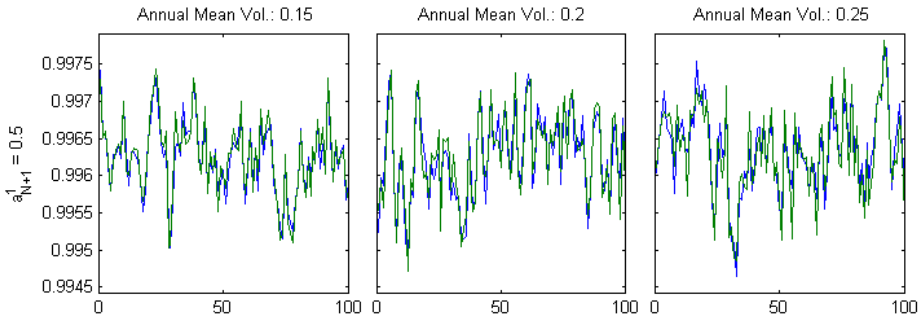
Figures 3 and 4 below also show simulation results for alternative settings, where the only change was on the autoregressive coefficient for consumption growth, from 0.3 to either 0.1 or 0.5. In them, we see very little difference between the average actual SDF, $M_t = \beta \left(\frac{c_t}{c_{t-1}}\right)^{-\gamma}$, and the average across simulations of $\widehat{M_t} = \frac{\overline{R}^G_t}{\frac{1}{T}\sum_{j=1}^{T} \overline{R}^G_j \overline{R}^A_j}$. 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).



Note: See Note in Figure 2.

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

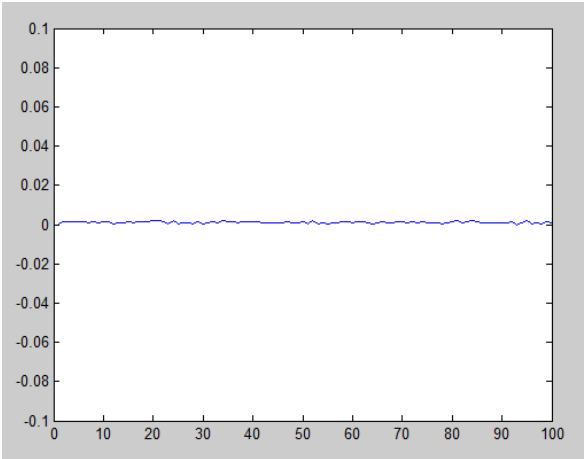


Note: See Note in Figure 2.

Overall, the results in Figures 2, 3 and 4 show numerical evidence that the SDF estimator is approximately unbiased in finite samples.

Next, we investigate the finite sample performance of the variance estimate in Proposition 3. The results for the case of the volatility 0.25 are presented in Figure 5. We omit the other cases for sake of space, but the results are similar. We can observe from the figure that the variance estimator is approximately unbiased in finite samples.

Figure 5: Empirical Bias for the Variance of the SDF Estimate Across Simulations.



References

[1] Almeida, C., Ardison, K., and Garcia, R. (2020), "Nonparametric Assessment of Hedge Fund Performance," *Journal of Econometrics*, 214 (2), 349–378.

[2] Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59 (3), 817–858.

[3] Araujo, F., Issler, J.V., and Fernandes, M. (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 Issler, J.V. (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] Arellano M., and Hahn, J. (2007), "Understanding Bias in Nonlinear Panel Models: Some Recent Developments,"In: Blundell, R., Newey, W., and Persson, T., eds. Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress. Econometric Society Monographs. Cambridge University Press; 381–409.

[6] 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*, 164 (1), 116–129.

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

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

[9] Bai, J. (2009), "Panel Data Models with Interactive Fixed Effects", *Econometrica*, 77, 1229–1279.

[10] Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models,"*Econometrica*, 70, 191–221.

[11] Bai, J., and Ng, S. (2004), "A Panic Attack on Unit Roots and Cointegration," *Econometrica*, 72, 1127–1177.

[12] Bansal, R., and Yaron, A. (2004), "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *The Journal of Finance*, 59, 1481–1509.

[13] Barillas, F., and Shanken, J. (2017), "Which Alpha?" *Review of Financial Studies*, 30(4), 1316–1338.

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

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

[16] Campello, M., Galvao, 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] Chandra, T. K., and Ghosal, S. (1996), "The Strong Law of Large Numbers for Weighted Averages Under Dependence Assumptions," *Journal of Theoretical Probability*, 9, 3, 797–809.

[18] Chen, Q., and Fang, Z. (2019), "Improved inference on the rank of a matrix," *Quantitative Economics*, 10, 1787–1824.

[19] Chen, L., Pelger, M., and Zhu, J. (2024), "Deep Learning in Asset Pricing, " *Management Science*, 70 (2), 714-750.

[20] Christensen, T.M. (2017), "Nonparametric Stochastic Discount Factor Decomposition," *Econometrica*, 85 (5), 1501–1536.

[21] Daniel, K., Hirshleifer, D., and Sun, L. (2020), "Short- and Long-Horizon Behavioral Factors," *The Review of Financial Studies*, 33(4), 1673—1736.

[22] Durrett, R. (2019). *Probability: Theory and Examples*, Cambridge University Press, 5th edition.

[23] Engle, R.F., and Issler, J.V. (1995), "Estimating Common Sectoral Cycles," *Journal of Monetary Economics*, 35, 83–113.

[24] Engle, R.F., and Kozicki, S. (1993), "Testing for Common Features," *Journal of Business and Economic Statistics*, 11(4), 369–380.

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

[26] 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.

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

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

[29] Fama, E.F., and MacBeth, J.D. (1973), "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81 (3), 607-–636.

[30] Feng, G., Giglio, S., and Xiu, D. (2020), "Taming the Factor Zoo: A Test of New Factors," *Journal of Finance*, 75 (3), 1327–1370.

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

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

[33] 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*, 100, 830–840.

[34] Gaglianone, W.P., and Issler, J.V. (2023), "Microfounded Forecasting," *Mimeo, FGV.*

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

[36] Giglio, S., Kelly, B., and Xiu, D. (2022), "Factor Models, Machine Learning, and Asset Pricing," Annual Review of Financial Economics, 14, 337-368.

[37] Gomes, F.A.R., and Issler, J.V. (2017), "Testing Consumption Optimality using Aggregate Data," *Macroeconomic Dynamics*, 21 (5), 1119–1140.

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

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

[40] Hansen, L.P., and Richard, S.F. (1987), "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–613.

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

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

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

[44] Hansen, L.P., and Singleton, K. (1984), "Errata: Generalized Instrumental Variables Estimation of Nonlinear Expectations Models," *Econometrica*, 52 (1), 267–268.

[45] Harrison, J.M., and Kreps, D.M. (1979), "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory*, 20, 381–405.

[46] Hecq, A., Palm, F.C., Urbain, J.-P. (2006), "Testing for Common Cyclical Features in VAR Models with Cointegration," *Journal of Econometrics*, 132 (1), 117–141.

[47] Hou, K., Xue, C., and Zhang, L., (2015), "Digesting Anomalies: An Investment Approach," *Journal of Financial Studies*, 28 (3), 650–705.

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

[49] Issler, J.V., Martins, T.G., and Moog, L. (2023), "Less is More: An Empirical Investigation of Asset Pricing Models in the Cross-Sectional Dimension," Mimeo, FGV.

[50] Issler, J.V., and Moog, L. (2025), "An Out-of-Sample Investigation of Asset Pricing Models," Mimeo, FGV.

[51] Issler, J.V., and Vahid, F. (2001), "Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates," *Journal of Monetary Economics*, 47, 449–475.

[52] 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*, 132 (1), 281–303.

[53] Kapetanios, G. (2008), "A Bootstrap Procedure for Panel Data Sets with Many Cross-Sectional Units," *Econometrics Journal*, 11, 377–395.

[54] Korsaye, S.A., Quaini, A., and Trojani, F. (2021), "Smart Stochastic Discount Factors," Swiss Finance Institute Research Paper No. 21-51, Available at SSRN: `https://ssrn.com/abstract=3878300`.

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

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

[57] 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), 1238–1287.

[58] Lewbel, A. (2019), "The Identification Zoo: Meanings of Identification in Econometrics," *Journal of Economic Literature*, 57, 835–903.

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

[60] Newey, W. K., and McFadden, D. L. (1994), "Large Sample Estimation and Hypothesis Testing," in Handbook of Econometrics, ed. by R. F. Engle and D. L. McFadden, Elsevier, vol. 4, chap. 36, 2111–2245.

[61] Newey, W. K., and West, K. D. (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

[62] Phillips, P.C.B., and Moon, H.R. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica* 67, 1057–1111.

[63] Phillips, P.C.B., and Moon, H.R. (2000), "Nonstationary Panel Data Analysis: An Overview of Some Recent Developments," *Econometric Reviews*, 19, 263–286.

[64] Pukthuanthong, K., Roll, R.W., and Wang, J.L. (2021), "An Agnostic and Practically Useful Estimator of the Stochastic Discount Factor," Available at SSRN: `https://ssrn.com/abstract=3503974`.

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

[66] Ross, S.A. (1976), "The Arbitrage Pricing of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

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

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

[69] Sandulescu, M., Trojani, F., and Vedolin, A. (2021), "Model-Free International Stochastic Discount Factors," *Journal of Finance*, 76 (2), 935–976.

[70] Stambaugh, R., and Yuan, Y. (2017), "Mispricing Factors," *Journal of Financial Studies*, 30 (4), 1270–1315.

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

[72] Vahid, F., and Engle, R.F. (1993), "Common Trends and Common Cycles," *Journal of Applied Econometrics*, 8, 341–360.

[73] Vahid, F., and Engle, R.F. (1997), "Codependent Cycles," *Journal of Econometrics*, 80, 199–221.

[74] Vahid, F., and Issler, J.V. (2002), "The Importance of Common Cyclical Features in VAR Analysis: A Monte Carlo Study," *Journal of Econometrics*, 109, 341–363.

[75] Wright, F.T., Platt, R.D., and Robertson, T. (1977), "A Strong Law for Weighted Averages of Independent, Identically Distributed Random Variables with Arbitrarily Heavy Tails," *Annals of Probability*, 5, 4, 586–590.