Journal of Financial Economics

Arbitrage-based recovery[☆]

Ferenc Horvath [*]

University of Liverpool, United Kingdom

ARTICLE INFO

JEL classification:
G1
G12
G13
G17

Keywords:
Recovery theorem
Physical probabilities
Stochastic discount factor
No-arbitrage

ABSTRACT

We develop a novel recovery theorem based on no-arbitrage principles. To implement our Arbitrage-Based Recovery Theorem empirically, one needs to observe the Arrow–Debreu prices only for one single maturity. We perform several different density tests and mean prediction tests using more than 26 years of S&P 500 options data, and we find evidence that our method can correctly recover the probability distribution of the S&P 500 index return on a monthly horizon, despite the presence of a non-trivial permanent SDF component.

1. Introduction

Extracting information about the physical probabilities of future events from market data has been the focus of finance research for several decades. A huge variety of applications in financial economics relies on knowledge of physical probabilities, such as portfolio choice, risk management, and asset pricing, just to name a few. Market practice usually hinges upon estimating physical probabilities from historical data. Nevertheless, estimates are only as reliable as good of a representative the historical data are.[1] Therefore, extracting physical probabilities from real-time market data in a forward-looking manner (as opposed to estimating them from backward-looking historical data) would be of paramount importance.

However, since the seminal works of Black and Scholes (1973) and Merton (1973) (and even earlier, of Bachelier, 1900), we know that asset prices, per se, depend only on risk-neutral probabilities, but not on physical probabilities. Therefore, extracting information about physical probabilities from asset pricing data is not possible without imposing further assumptions. In a recent seminal paper, a set of such assumptions was stated and the corresponding physical probability recovery theory developed by Ross (2015). In a finite-state framework Ross demonstrates that if we are able to observe the entire Arrow–Debreu price matrix and the stochastic discount factor (SDF) is transition independent,[2] then the physical probability measure can be recovered. Since in reality we can observe only one row of the Arrow–Debreu price matrix (and not the entire matrix), Ross proposes an approach where observing the transition state prices for as many maturities as the number of possible states, the entire Arrow–Debreu price matrix can be reconstructed.[3]

☆ Nikolai Roussanov was the editor for this article. We thank Zsolt Bihary, Jens C. Jackwerth, Christian S. Jensen, Yongjin Kim, Péter Kondor, Yaxuan Qi, Alexander Szimayer, Jianfeng Xu, and Steven Utke for useful comments.

[1] Already Markowitz (1952) warns about basing one's investment decision on expected returns and (co)variances which have been estimated from historical data: "*... (W)e must have procedures for finding reasonable* μ_i *and* σ_{ij}*. These procedures (...) should combine statistical techniques and the judgment of practical men. (...) (T)he statistical computations should be used to arrive at a tentative set of* μ_i *and* σ_{ij}. Judgment should then be used in increasing or decreasing some of these μ_i and σ_{ij} *on the basis of factors or nuances not taken into account by the formal computations.*"

[2] The stochastic discount factor is transition independent if it can be written as the product of a constant and a fraction, the numerator of which is a positive scalar-valued function of the state variable evaluated at the arrival state, and the denominator of which is the same function evaluated at the initial state. I.e., a transition-independent SDF has the functional form $m_{i,j} = \delta \times h\left(\theta_j\right) / h\left(\theta_i\right)$, where i is the initial state, j is the arrival state, $\delta \in \mathbb{R}_{>0}$, θ is the state variable, and h is a positive scalar-valued function.

[3] As Jensen et al. (2019) point out, the recovered Arrow–Debreu matrix is almost surely unique. I.e., the set of parameters for which there exists a continuum of recovered Arrow–Debreu matrices (instead of a unique recovered Arrow–Debreu matrix) has a measure of zero.

Received 16 July 2022; Received in revised form 25 October 2024; Accepted 28 October 2024

Available online 8 November 2024

Ross's recovery theorem has initiated a lively academic debate both in the theoretical and in the empirical finance literature. Several years before the emergence of the recovery theorem, Alvarez and Jermann (2005) and Hansen and Scheinkman (2009) showed that the stochastic discount factor process can be decomposed into the product of a transition-independent factor (the *transitory component*) and a martingale (the *permanent component*). Building on this factorization, Borovička et al. (2016) point out that Ross's approach recovers the true physical probabilities if and only if the martingale component of the SDF process is constant. This implicit assumption of a constant martingale component is, however, incompatible with several mainstream asset pricing models, for example, with the long-run risk model of Bansal and Yaron (2004). Using options on the 30-year Treasury bond futures, Bakshi et al. (2018) confirm empirically that the martingale component is unlikely to be constant. Qin et al. (2018) arrive at the same conclusion, using U.S. Treasury data. Dillschneider and Maurer (2019), after extending the recovery theorem to continuous state spaces, use S&P 500 index options to argue that Ross's recovery seems to be misspecified. Tran and Xia (2018) highlight the importance of the Arrow–Debreu price matrix dimension when implementing Ross's recovery theorem. Martin and Ross (2019) apply the idea of the recovery theorem to study the properties of the spot yield curve. Schneider and Trojani (2019) recover the conditional minimum variance projection of the pricing kernel on tradeable realized moments of market returns. Pazarbasi et al. (2021) amend the recovery framework with investor heterogeneity. In an early paper, Carr and Yu (2012) extend Ross's recovery to a bounded diffusion context, and Walden (2017) derives necessary and sufficient conditions of when recovery is possible with unbounded diffusion processes. A (non-exhaustive) list of further studies on recovery includes Dubynskiy and Goldstein (2013), Huang and Shaliastovich (2014), Liu (2014), Massacci et al. (2016), Park (2016), Qin and Linetsky (2016), Ghosh and Roussellet (2020), Jensen (2021), and Heston (2021).

Jackwerth and Menner (2020) perform a thorough empirical analysis to assess whether the physical probabilities recovered by Ross's Recovery Theorem and other approaches are indeed equal to the true physical probabilities. Using monthly observations of more than 30 years of S&P 500 European-style options, they recover the one-month physical probability distributions of the S&P 500 index level. Then, based on several different density tests and mean- and variance prediction tests, Jackwerth and Menner reject the hypothesis that realized S&P 500 index values are drawn from the recovered physical probability distributions.

The empirical analysis of Jackwerth and Menner (2020) also points out that applying Ross's recovery theorem to reconstruct the entire Arrow–Debreu price matrix can lead to a recovered transition state price matrix with highly counterintuitive features. E.g., even though one would expect high recovered state prices on the main diagonal of the Arrow–Debreu price matrix, the recovered Arrow–Debreu price matrix of Jackwerth and Menner tends to have high state prices in states which are far off-diagonal. Furthermore, the model-implied risk-free rates are often negative or their magnitude is as high as several hundred percent.

Besides assuming a constant SDF martingale component, empirical implementations of Ross's recovery theorem also assume that transition state prices are time-homogeneous. Jensen et al. (2019) develop the *Generalized Recovery Theorem*, where they relax this assumption. They show that as long as we are able to observe the transition state prices for at least as many maturities as the number of possible states, physical probabilities (corresponding to the assumed SDF) can almost surely be recovered, without assuming time-homogeneity.

Jensen et al. (2019) also demonstrate when (and how) *Generalized Recovery* can be implemented even if the number of possible states grows over time. Standard examples for such a case include the models of Mehra and Prescott (1985) and of Cox et al. (1979). Basically, one needs to make use of the fact that the SDF can be expressed as a function (with a *fixed* number of parameters) of the state variable. Recovery then means recovering the parameter values. Then, after calculating the SDF values, recovering the physical probabilities is straightforward. In this structured version of the *Generalized Recovery* framework, one needs to observe the transition state prices for at least as many maturities as the number of unknown parameters.

An unattractive requirement of each of the currently available recovery theorems is that one needs to observe the transition state prices for several different maturities in order to recover the physical probabilities for one single maturity. Even in the simplest case when the only state variable is the asset price, one needs to observe the transition state prices for at least as many different maturities as many possible values the asset price can take. For example, if we are interested in the physical probabilities over a one-month horizon and we can observe the monthly transition state prices for up to one year of maturity, then (assuming non-overlapping observation periods) the asset price is allowed to take only twelve different values. As Ross (2015) notes, such a coarse grid leads to poorly discretized transition state prices and eventually poorly discretized recovered probabilities. Audrino et al. (2019) use overlapping observation periods and transition state prices of $1/10$ of a month, but without additional restrictions, the recovered one-month transition state prices still exhibit counterintuitive features, as it is demonstrated by Jackwerth and Menner (2020). An adapted version of the Generalized Recovery approach of Jensen et al. (2019) suggests a way out: if the SDF can be written as a known function (with a small number of unknown parameters) of the state variables, then it is enough to observe the transition state prices for only as many different maturities as the number of parameters (which is, ideally, much lower than the number of possible values of the state variable). There is, however, a trade-off: the modeler has to exogenously provide the functional form of the SDF. The validity of the recovered physical probabilities will then also depend on the validity of the assumed SDF functional form. Furthermore, one still needs to observe the transition state prices for several different maturities, if the SDF has more than one parameter.

When recovery theorems are empirically implemented in the literature, the price of a traded asset is usually assumed to constitute the state variable. The goal is then to extract a "flexible" form of the SDF in the sense that the SDF is restricted only by the transition independence assumption. In this paper, we show that such a flexibility is actually spurious, and the recovered SDF cannot be anything else but the reciprocal of the gross asset return. This result follows from an additional no-arbitrage restriction ignored by the existing literature.

First, we develop our Arbitrage-Based Recovery approach when the state space (spanned by the price of a traded asset) is discrete. In such a framework, the pricing operator is represented by a positive square matrix (the *Arrow–Debreu price matrix*), and the existence and uniqueness of our recovered probability measure relies on the Perron–Frobenius theorem. Then, we show that the implications of Arbitrage-Based Recovery also hold when the state space is continuous and the state variable (the price) can take any real value in a closed and bounded interval. Since in such a framework the pricing operator is an integral operator (instead of a square matrix) and one cannot rely on the Perron–Frobenius theorem, we prove existence and uniqueness of our recovered probability measure directly. Finally, we show that our results still hold and Arbitrage-Based Recovery can also be applied when the state variable is the price–dividend ratio instead of the price itself.

Besides contributing to the literature on recovery theorems, our paper also contributes to the emerging literature on the factorization of the SDF into a transitory and a permanent component. Namely, we show that as long as the state variable is the price of a traded asset and the state space is bounded and closed, under a new numeraire the transitory SDF component is equal to the inverse of the realized gross return on that asset. Then, we show that the same result holds when the state variable is the price–dividend ratio (instead of the price itself) under another new numeraire.

Our results on the SDF factorization lead us directly to the next contribution of our paper. Namely, we add to the recent literature on estimating the expected return on the market in a forward-looking manner, based on options data. Martin (2017) shows that as long as the *"negative correlation condition"* is satisfied, a lower bound for the expected return on the market can be derived in real time, based on option prices. Then, Martin demonstrates empirically that the lower bound appears to be tight in the sense that the correlation in question seems to be close to zero. Our paper contributes to these results in two aspects. First, we show that Martin's negative correlation condition is actually a condition on the martingale component of the SDF under a new numeraire when a bounded and closed price process spans the state space. Namely, it imposes that the martingale component of the SDF under the so-called *dividend account numeraire* is negatively correlated with the total market return under the dollar numeraire.[4] Second, we show that the probabilities recovered by the Arbitrage-Based Recovery Theorem imply the same expected return on the market as Martin's lower bound.

Ross (2015) and other recovery theorems do not specify the state variable. The state space is allowed to be spanned by any variable, as long as options are traded on that state variable. The vast majority of the empirical literature on recovery, however, focuses on state spaces spanned by prices. The reason is pragmatic: most (liquid) options are written on prices (e.g., on the S&P 500 index level).[5] The fact that the state variable is a price itself, however, can be used to our advantage, and this aspect of the empirical implementation of recovery theorems has been so far largely overlooked by the literature. In this paper, we demonstrate that – after an appropriate change of numeraire – the transitory component of the SDF can be directly inferred from a theoretical recursive no-arbitrage restriction which the asset price spanning the underlying state must satisfy. In a recent paper, Zhu (2020) formulates a critique of Ross's Recovery Theorem based on its ignoring the Fundamental Theorem of Asset Pricing when developing the Recovery Theorem itself. However, when the state variable is not a price, the only no-arbitrage requirement is that the Arrow–Debreu price matrix be non-negative – which in itself is assumed by Ross's Recovery. The Fundamental Theorem of Asset Pricing does not necessitate imposing any additional assumptions in general. The additional connection between Ross's Recovery Theorem and the Fundamental Theorem of Asset Pricing proposed by Zhu (2020) exists only in the special case when the state variable is the price of a traded asset and the asset does not pay any intermediate cash flow. In that case, the Perron–Frobenius eigenvector is indeed equal to the vector of possible prices and the corresponding eigenvalue is equal to one, as argued by Zhu (2020). Still, Ross's Recovery Theorem remains valid and it is not inconsistent with the Fundamental Theorem of Asset Pricing; rather, the latter pins down the unique solution of Ross's recovery problem. However, when the underlying asset potentially pays any intermediate cash flow (e.g., dividend), as it is the case for stocks or the S&P 500 stock market index, this implication does not hold any more under the dollar numeraire. The numeraire has to be changed in order for it to reflect the effects of the intermediate cash flow. Importantly, the recovered stochastic discount factor will be transition independent only under the new numeraire (and not the dollar), and the recovery exercise can be interpreted as the long-term factorization of the SDF only under the new numeraire. A pragmatic approach to handle intermediate cash flows might be to use the cash-flow-reinvestment account value as the state variable, but this still requires making an assumption about the paid cash flow in future states, because options are almost always written on post-cash-flow prices, e.g., stock options are written on post-dividend stock prices.

As an empirical contribution, we implement our Arbitrage-Based Recovery Theorem using more than 26 years of S&P 500 stock market index options data. We test the hypothesis that the realized monthly returns of the S&P 500 market index are indeed drawn from the probability distributions recovered by our Arbitrage-Based Recovery Theorem. We find that this hypothesis cannot be rejected when using our recovery approach, while it is confidently rejected when Jackwerth and Menner (2020) test it using other recovery approaches. We also demonstrate that this result does not contradict the findings of the empirical literature on the importance of the permanent SDF component. We show that the restrictive assumptions imposed by empirical implementations of recovery theorems (e.g., that a closed and bounded price process spans the space) imply an SDF which is very unlikely to be equivalent to the true transitory SDF component. Hence, the fact that the permanent SDF component is very volatile empirically (as we confirm in Section 8.1) does not contradict the empirical success of our Arbitrage-Based Recovery Theorem.

The paper is organized as follows. In Section 2, we develop our Arbitrage-Based Recovery Theorem in a discrete-state-space framework. In Section 3, we discuss several alternative interpretations of the probabilities recovered by Arbitrage-Based Recovery. In Section 4, we connect Arbitrage-Based Recovery to the long-term factorization of Alvarez and Jermann (2005) and Hansen and Scheinkman (2009). In Section 5, we extend our model into a continuous-state-space framework, while in Section 6 we show how our theory can be implemented if the state variable is the price–dividend ratio (instead of the price itself). In Section 7, we implement our Arbitrage-Based Recovery Theorem empirically using more than 26 years of S&P 500 index options data. We test the hypothesis that the realized monthly returns of the S&P 500 market index are indeed drawn from the probability distributions recovered by our Arbitrage-Based Recovery Theorem, and we also perform several robustness checks. In Section 8, we demonstrate how the empirical success of our recovery approach can be reconciled with the literature's findings that the permanent SDF component is empirically very volatile. Section 9 concludes. In Appendix A, we provide methodological details on how we attach left and right tails to the risk-neutral probability density functions when we implement our Arbitrage-Based Recovery empirically. In the accompanying Online Appendix, we give a brief overview of the currently available recovery theorems, focusing on how Arbitrage-Based Recovery offers itself as the next logical step in the evolutionary pathway of recovery theorems; we provide a detailed treatment of specific dividend payment structures; and we also provide there the proofs which are not included in the paper.

2. Arbitrage-based recovery

Since the original Recovery Theorem was developed by Ross (2015), a sequence of alternative recovery approaches has emerged. In the Online Appendix, we provide a detailed comparison of the different recovery approaches and the exact assumptions they pose.

In this section, we develop a novel recovery theorem which does not require the observation of Arrow–Debreu prices for several different maturities in order to recover the probabilities over a single period. The key element of our approach is a change of numeraire, which enables us to link the solution of the recovery (eigenvalue–eigenvector) problem to an additional no-arbitrage criterion. Under the new numeraire, we retain Ross's assumption of a transition-independent stochastic discount factor. By applying our version of the recovery theorem, we need to observe the Arrow–Debreu prices only for the current state as initial state and only for one single maturity, and from those observations we can recover the physical transition probabilities corresponding to that maturity. Our model can be applied to recover the physical probabilities of the price movement of any asset, as long as the state space is spanned by the price of that asset. A schematic demonstration of our recovery theory can be found in Fig. 1.

[4] We provide the relevant definitions and we discuss these results in details in Section 3.4.

[5] Examples of options the underlying variable of which is not a price include, among others, interest rate options and weather options.

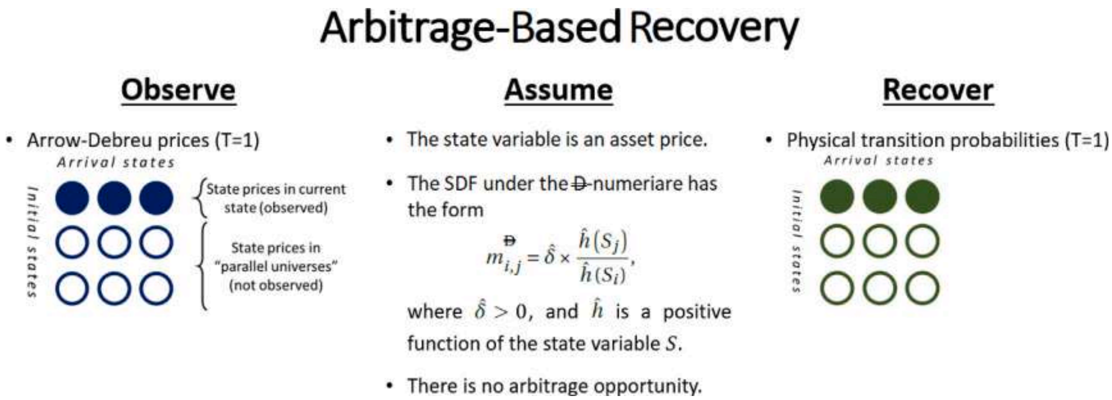

Fig. 1. Arbitrage-based recovery. One needs to observe the Arrow–Debreu prices (where the current state is the initial state) for only one maturity. If the SDF under the dividend account numeraire takes the form $\hat{\delta} \times \hat{h}(S_j) / \hat{h}(S_i)$, where $\hat{\delta}$ is a positive constant, S is the asset price and the only state variable, and $\hat{h}$ is a positive function of S, then in the absence of arbitrage opportunities the physical transition probabilities can be recovered.

Fig. 2. Timing in our model. At time t, we observe the stock price in the current state i. Then, at time $t+1$ in state j the dividend $D_{i,j}$ is paid. Right afterwards, we observe the stock price S_j.

2.1. Framework

Consider a discrete-time complete financial market. For the sake of concreteness, we consider here the market for a stock, but our approach can readily be applied for other long-term financial assets as well. The only state variable in our model is the stock price, which can take n possible values, collected in the $n \times 1$ column vector $\boldsymbol{S} \in \mathbb{R}^n_{>0}$. The $n \times n$ matrix of one-period Arrow–Debreu prices is $\boldsymbol{A}^{\$} \in \mathbb{R}^{n\times n}_{>0}$. The dollar sign in the superscript emphasizes that the Arrow–Debreu price matrix corresponds to the dollar numeraire. The stock pays dividend in one period, which is allowed to be stochastic. Denoting by $D_{i,j} \in \mathbb{R}_{\geq 0}$ the dollar amount of dividend paid in one period if we go from state i to state j, we introduce the $n \times n$ gross dividend yield matrix $\boldsymbol{F} \in \mathbb{R}^{n\times n}_{\geq 1}$ with elements

$$F_{i,j} = \frac{D_{i,j} + S_j}{S_j}, \tag{1}$$

where S_j is the stock price in state j, observed just after the dividend has been paid. The timeline in Fig. 2 demonstrates the timing in our model. In our theoretical discussion, we do not restrict the form of the dividend yield matrix $\boldsymbol{F}$ apart from requiring dividends to be non-negative (and therefore the elements of $\boldsymbol{F}$ to be greater than or equal to one).[6]

Now, we introduce the concept of a *dividend account*. A dividend account has an initial (time-zero) value of \$1. Then, after one period, the account is increased by multiplying its value by the realized gross dividend yield. So, its value after one period will be $\$1 \times F_{i,j}$ if we go from state i to state j. Using the value of the dividend account as numeraire, we can define a "currency" (denoted by Ð) whose value is always equal to the dividend account value.

Next, consider the Hadamard (element-wise) product $\boldsymbol{A}^{Ð} \triangleq \boldsymbol{A}^{\$} \odot \boldsymbol{F}$. Since $A^{\$}_{i,j}$ is the current price in dollars (in state i) of receiving \$1 in one period in state j (and nothing in other states), $A^{Ð}_{i,j}$ is the current price in dollars (in state i) of receiving $1 \times F_{i,j}$ dollars in one period in state j (and nothing in other states). Alternatively, we can interpret $A^{Ð}_{i,j}$ as the current price in units of Ð (in state i) of receiving one unit of Ð in one period in state j (and nothing in other states).[7] Hence, $\boldsymbol{A}^{Ð}$ is the Arrow–Debreu price matrix under the Ð-numeraire.

The stochastic discount factor (SDF) under the dollar numeraire, $\boldsymbol{m}^{\$} \in \mathbb{R}^{n\times n}_{>0}$, is the (unique) $n \times n$ matrix such that

$$\boldsymbol{A}^{\$} = \boldsymbol{m}^{\$} \odot \boldsymbol{\pi} \tag{2}$$

holds, where $\boldsymbol{\pi} \in \mathbb{R}^{n\times n}_{>0}$ is the physical transition probability matrix. Similarly, the SDF under the Ð-numeraire is the unique $n \times n$ matrix $\boldsymbol{m}^{Ð} \in \mathbb{R}^{n\times n}_{>0}$ such that

$$\boldsymbol{A}^{Ð} = \boldsymbol{m}^{Ð} \odot \boldsymbol{\pi} \tag{3}$$

holds.[8] Substituting $\boldsymbol{A}^{Ð} = \boldsymbol{A}^{\$} \odot \boldsymbol{F}$ in (3) and then dividing both sides (element-wise) by $\boldsymbol{\pi}$, we obtain that the relationship between the dollar SDF and the Ð-SDF is

$$\boldsymbol{m}^{Ð} = \boldsymbol{m}^{\$} \odot \boldsymbol{F}. \tag{4}$$

Following Ross (2015), we now introduce the definition of a *transition-independent stochastic discount factor*.

Definition 1. Consider a discrete-time finite-state complete-market model where the only state variable is the post-dividend stock price S. The stochastic discount factor $\boldsymbol{m}^{\$}$ is said to be transition independent if it can be expressed as

$$m^{\$}_{i,j} = \delta \times \frac{h\left(S_j\right)}{h\left(S_i\right)}, \tag{5}$$

[6] In the Online Appendix, we discuss several specific forms of the gross dividend yield matrix $\boldsymbol{F}$ in details. In our empirical analysis in Section 7, we assume that the dividend is known one period ahead.

[7] To see this, recall that at time zero Ð1 = \$1.

[8] Note that the physical transition probability matrix $\boldsymbol{\pi}$ in (3) is the same as in (2).

where i is the initial state, j is the arrival state, δ is a positive scalar, and $h : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is a positive scalar-valued function.

Similarly, the stochastic discount factor $\boldsymbol{m}^{Đ}$ is said to be transition independent if it can be expressed as

$$\overset{Đ}{m}_{i,j} = \hat{\delta} \times \frac{\hat{h}\left(S_j\right)}{\hat{h}\left(S_i\right)}, \tag{6}$$

where $\hat{\delta}$ is a positive scalar, and $\hat{h} : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is a positive scalar-valued function.

2.2. No-arbitrage and recovered physical probabilities

The purpose of our recovery theory – just as it is the purpose of any other recovery theory – is to decompose the observed Arrow–Debreu prices into the SDF and the physical probabilities. The first building block of our approach, which we formalize in the following theorem, establishes the uniqueness of the recovered physical transition probabilities. Since the theorem holds under any numeraire, we do not specify the numeraire of the Arrow–Debreu price matrix.

Theorem 1. *Consider a discrete-time finite-state complete-market model with a positive square Arrow–Debreu price matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}_{>0}$. There exists a unique decomposition of the matrix $\boldsymbol{A}$ such that $\boldsymbol{A} = \phi \boldsymbol{Z}\boldsymbol{\Pi}\boldsymbol{Z}^{-1}$, where $\boldsymbol{Z} \in \mathbb{R}^{n\times n}_{\geq 0}$ is a diagonal matrix with positive diagonal elements, $\boldsymbol{\Pi} \in \mathbb{R}^{n\times n}_{>0}$ is a stochastic matrix, $\phi \in \mathbb{R}_{>0}$ is a positive scalar, and the uniqueness of $\boldsymbol{Z}$ is to be understood as "unique up to a positive scale factor".*

Proof. The proof follows the logic of the proof of Result 1 in Martin and Ross (2019). First, we prove the existence of $\boldsymbol{Z}$, $\boldsymbol{\Pi}$, and ϕ. According to the Perron–Frobenius theorem, there exist a positive vector $\boldsymbol{z} \in \mathbb{R}^{n}_{>0}$ and a positive scalar $\phi \in \mathbb{R}_{>0}$ such that $\boldsymbol{A}\boldsymbol{z} = \phi\boldsymbol{z}$. Define $\boldsymbol{\Pi} \triangleq \frac{1}{\phi}\boldsymbol{Z}^{-1}\boldsymbol{A}\boldsymbol{Z}$, where $\boldsymbol{Z}$ is a diagonal matrix with the elements of $\boldsymbol{z}$ in its diagonal. Since $\boldsymbol{\Pi}$ is positive and its elements in each of its rows sum to one, it is a stochastic matrix. As one can check, $\boldsymbol{A} = \phi\boldsymbol{Z}\boldsymbol{\Pi}\boldsymbol{Z}^{-1}$ holds. Thus, there indeed exist ϕ, $\boldsymbol{Z}$, and $\boldsymbol{\Pi}$ satisfying the appropriate claims in the theorem.

Now, we prove the uniqueness of $\boldsymbol{Z}$, $\boldsymbol{\Pi}$, and ϕ. Multiplying both sides of $\boldsymbol{A} = \phi\boldsymbol{Z}\boldsymbol{\Pi}\boldsymbol{Z}^{-1}$ by $\boldsymbol{Z}\boldsymbol{\iota}$ from the right (where $\boldsymbol{\iota}$ is the $n \times 1$ column vector of ones), we obtain $\boldsymbol{A}\boldsymbol{z} = \phi\boldsymbol{z}$. According to the Perron–Frobenius theorem, ϕ is unique and $\boldsymbol{z}$ is unique up to a positive scale factor. Hence, $\boldsymbol{Z}$ is also unique up to a positive scale factor. And if we rearrange $\boldsymbol{A} = \phi\boldsymbol{Z}\boldsymbol{\Pi}\boldsymbol{Z}^{-1}$ as $\boldsymbol{\Pi} = \frac{1}{\phi}\boldsymbol{Z}^{-1}\boldsymbol{A}\boldsymbol{Z}$, we find that therefore $\boldsymbol{\Pi}$ is also unique. ■

Now, we impose our additional no-arbitrage condition. If there is no arbitrage opportunity, then the time-t stock price must be equal to the sum of each Arrow–Debreu prices multiplied by the time-$(t+1)$ cash flows in the respective states. Since the stock cash flow at time $t+1$ is the sum of the dividend and the post-dividend stock price, this means that

$$S_i = \sum_{j=1}^{n} A^{\$}_{i,j} \times \left(F_{i,j} \times S_j\right) \qquad \forall i \in \{1, 2, \dots, n\} \tag{7}$$

must hold, if there is no arbitrage opportunity. Note that S is always the stock price which we can observe in the stock market *after dividend payment*: on the left-hand side S_i is the stock price which we observe at time t, and S_j on the right-hand side is the stock price which we observe at time $t+1$, right after we have received the dividend. Expressing (7) in matrix notation (and switching the left-hand side and the right-hand side), we find that

$$\left(\boldsymbol{A}^{\$} \odot \boldsymbol{F}\right)\boldsymbol{S} = \boldsymbol{S} \tag{8}$$

must hold, if there is no arbitrage opportunity. And since $\boldsymbol{A}^{Đ} = \boldsymbol{A}^{\$} \odot \boldsymbol{F}$ by definition, this no-arbitrage condition can be equivalently written as

$$\boldsymbol{A}^{Đ}\boldsymbol{S} = \boldsymbol{S}. \tag{9}$$

Hence, interestingly, we find that the no-arbitrage condition (9) is actually a condition on the Perron–Frobenius eigenvector and the Perron–Frobenius eigenvalue of $\boldsymbol{A}^{Đ}$, i.e., the Arrow–Debreu price matrix under the Đ-numeraire. We formalize this in the following lemma.

Lemma 1. *Consider a positive square Arrow–Debreu price matrix under the Đ-numeraire, $\boldsymbol{A}^{Đ} \in \mathbb{R}^{n\times n}_{>0}$, in a discrete-time finite-state complete-market model. The only state variable is the post-dividend stock price in dollars, the possible values of which are collected in the vector $\boldsymbol{S} \in \mathbb{R}^{n}_{>0}$. If there is no arbitrage opportunity, then the Perron–Frobenius eigenvector of $\boldsymbol{A}^{Đ}$ is $\boldsymbol{S}$ and its Perron–Frobenius eigenvalue is* 1.

Proof. The fact that the Perron–Frobenius eigenvector of $\boldsymbol{A}^{Đ}$ is $\boldsymbol{S}$ and its Perron–Frobenius eigenvalue is 1 readily follows from Eq. (9) and the Perron–Frobenius theorem. ■

Now, we are ready to state our main theorem, in which we provide our recovered physical transition probabilities.

Theorem 2 (*Arbitrage-Based Recovery Theorem*). *Consider a discrete-time finite-state complete-market model with a positive square Arrow–Debreu price matrix under the dollar numeraire, $\boldsymbol{A}^{\$} \in \mathbb{R}^{n\times n}_{>0}$, and a gross dividend yield matrix $\boldsymbol{F} \in \mathbb{R}^{n\times n}_{\geq 1}$. The state variable is the post-dividend stock price, the possible values of which are collected in the vector $\boldsymbol{S} \in \mathbb{R}^{n}_{>0}$. The stochastic discount factor under the Đ-numeraire is assumed to be transition independent. If there is no arbitrage opportunity, then the elements of the physical transition probability matrix $\boldsymbol{\Pi} \in \mathbb{R}^{n\times n}_{>0}$ are*

$$\Pi_{i,j} = A^{\$}_{i,j} \times \frac{S_j \times F_{i,j}}{S_i} \tag{10}$$

for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, n\}$.

Proof. According to Theorem 1, there exist a unique positive scalar ϕ, a unique diagonal matrix $\boldsymbol{Z}$ with positive diagonal elements, and a unique stochastic matrix $\boldsymbol{\Pi}$ such that $\boldsymbol{A}^{Đ} = \phi\boldsymbol{Z}\boldsymbol{\Pi}\boldsymbol{Z}^{-1}$. Rearranging this, we obtain $\boldsymbol{\Pi} = \frac{1}{\phi}\boldsymbol{Z}^{-1}\boldsymbol{A}^{Đ}\boldsymbol{Z}$. According to Lemma 1, $\phi = 1$ and $Z_{i,i} = S_i$. Substituting these values into the rearranged equation and using the identity $\boldsymbol{A}^{Đ} = \boldsymbol{A}^{\$} \odot \boldsymbol{F}$, we find $\Pi_{i,j} = A^{\$}_{i,j} \times F_{i,j} \times S_j / S_i$. This completes the proof. ■

The result of Theorem 2 offers itself as the next logical step in the evolutionary path of recovery theorems. From a practical perspective, according to the theorem, if there is no arbitrage opportunity and the SDF under the Đ-numeraire is transition independent, we can uniquely recover the physical transition probabilities in the state space spanned by the (post-dividend) asset price.

2.3. Inferred stochastic discount factor

Similarly to recovering the physical transition probabilities, we can also infer the unique transition-independent stochastic discount factor under the Đ-numeraire, $\boldsymbol{m}^{Đ}$. We formalize this in the following corollary.

Corollary 1. *Consider a discrete-time finite-state complete-market model with a positive square Arrow–Debreu price matrix under the dollar numeraire, $\boldsymbol{A}^{\$} \in \mathbb{R}^{n\times n}_{>0}$, and a gross dividend yield matrix $\boldsymbol{F} \in \mathbb{R}^{n\times n}_{\geq 1}$. The state variable is the post-dividend stock price, the possible values of which are collected in the vector $\boldsymbol{S} \in \mathbb{R}^{n}_{>0}$. The stochastic discount factor under the Đ-numeraire is assumed to be transition independent. If there is no arbitrage opportunity, then the stochastic discount factor under the Đ-numeraire is*

$$\overset{Đ}{m}_{i,j} = \frac{S_i}{S_j}. \tag{11}$$

Furthermore, the stochastic discount factor under the dollar numeraire is

$$m^{\$}_{i,j} = \frac{S_i}{F_{i,j} \times S_j}. \tag{12}$$

Proof. The inferred SDFs follow readily from the definition of the Arrow–Debreu price, the relationship $\boldsymbol{A}^{Đ} = \boldsymbol{A}^{\$} \odot \boldsymbol{F}$, and our recovered probabilities in (10). ∎

3. Alternative interpretations and further discussion

Although the focus of the current paper is the canonical problem of recovering the physical probabilities from observable asset prices, the theory of Arbitrage-Based Recovery is also related to several other branches of the literature. In this section, we discuss these connections and highlight several alternative interpretations of our recovered probabilities.

3.1. Log-utility agent

The form of our inferred dollar SDF in (12) lends itself for an alternative economic interpretation. Namely, it can be interpreted as the stochastic discount factor implied by a pseudo-representative investor who has a subjective discount rate of zero and log-utility over the realized total cash flow, and who invests her wealth fully in the stock.[9] Consequently, this pseudo-representative agent makes her investment decision based on our recovered physical probabilities, (10). The next corollary formalizes this.

Corollary 2. *Consider a log-utility agent within a single-period finite-state complete-market arbitrage-free model, where the states are indexed by the possible values of the stock, expressed in dollars. She allocates her wealth at $t = 0$ so that her expected utility from the dollar value of her wealth at $t = 1$ is maximized. If she invests her wealth fully in the stock, then the probability measure used by this agent is the same as the probability measure recovered by the Arbitrage-Based Recovery Theorem.*

Proof. We provide the proof in the Online Appendix. ∎

3.2. Discounting by the return on the $S^{Đ}$-asset

Another alternative interpretation of our recovered probabilities is that they are used by an agent who discounts everything by the return on an asset whose next-period payoff (in state j) will be S_j units of Đ. We formalize this in the next corollary.

Corollary 3. *Consider an asset (the $S^{Đ}$-asset) whose next-period payoff in state j is S_j units of Đ. The probabilities recovered by the Arbitrage-Based Recovery Theorem correspond to the probabilities used by an agent who discounts the next-period cash flow by the return on the $S^{Đ}$-asset.*

Proof. The time-t price in state i of the $S^{Đ}$-asset is S_i units of Đ, due to the no-arbitrage condition (9). Thus, the one-period gross return on the $S^{Đ}$-asset under the Đ-numeraire is S_j / S_i. Comparing this to the recovered probabilities in (10), the result follows immediately. ∎

3.3. Positivity of Ross's subjective discount rate

In Ross's original recovery theorem, the Perron–Frobenius eigenvalue is often interpreted as the subjective discount factor of a pseudo-representative agent. The Perron–Frobenius theorem assures that this subjective discount factor is positive. Economic intuition further suggests that its value should be lower than one, in accordance with the concept of *time value of money*. Based on the logic of our Arbitrage-Based Recovery Theorem, now we prove rigorously that if the state variable is the price of a traded asset (e.g., the S&P 500 index level) and the dividend (or other similar cash flow) is for sure non-negative, then the Perron–Frobenius eigenvalue of the Arrow–Debreu (dollar) price matrix is for sure lower than or equal to one. Furthermore, if the dividend is strictly positive in at least one initial-state-arrival-state combination, then the Perron–Frobenius eigenvalue (and therefore the subjective discount factor of the pseudo-representative agent) is strictly lower than one.[10] We formalize this in the following theorem.

Theorem 3. *Consider a discrete-time finite-state complete-market model with a positive square Arrow–Debreu price matrix under the dollar numeraire, $\boldsymbol{A}^{\$} \in \mathbb{R}^{n\times n}_{>0}$, and a gross dividend yield matrix $\boldsymbol{F} \in \mathbb{R}^{n\times n}_{\geq 1}$. The state variable is the post-dividend stock price, the possible values of which are collected in the vector $\boldsymbol{S} \in \mathbb{R}^{n}_{>0}$. If there is no arbitrage opportunity, then the Perron–Frobenius eigenvalue of the matrix $\boldsymbol{A}^{\$}$ is less than or equal to one. Furthermore, if at least one element of the gross dividend yield matrix $\boldsymbol{F}$ is strictly greater than one, then the Perron–Frobenius eigenvalue of the matrix $\boldsymbol{A}^{\$}$ is strictly less than one.*

Proof. First, consider the case when the underlying asset does not pay any dividend (i.e., when each element of the gross dividend yield matrix $\boldsymbol{F}$ is equal to one). Then, by definition, $\boldsymbol{A}^{Đ} = \boldsymbol{A}^{\$}$. Hence, due to Lemma 1, the Perron–Frobenius eigenvalue of the matrix $\boldsymbol{A}^{\$}$ is equal to one.

Now, consider the case when at least one element of $\boldsymbol{F}$ is strictly greater than one. Then $\boldsymbol{A}^{\$} \leq \boldsymbol{A}^{Đ}$ and $\boldsymbol{A}^{Đ} \neq \boldsymbol{A}^{\$}$. Due to Corollary 6.16 in Zhan (2013), this implies that the Perron–Frobenius eigenvalue of $\boldsymbol{A}^{\$}$ is strictly less than the Perron–Frobenius eigenvalue of $\boldsymbol{A}^{Đ}$. Since according to Lemma 1 the Perron–Frobenius eigenvalue of $\boldsymbol{A}^{Đ}$ is equal to one, the Perron–Frobenius eigenvalue of $\boldsymbol{A}^{\$}$ must then be strictly less than one. This completes the proof. ∎

3.4. Negative correlation condition of Martin (2017)

In a model-free framework, Martin (2017) shows that the conditional expected excess return on the market can be decomposed into two terms. Concretely,

$$\mathrm{E}_t\left(R^{\$}_{t+1}\right) - R^{\$}_{f,t+1} = \frac{1}{R^{\$}_{f,t+1}} var^{*}_t\left(R^{\$}_{t+1}\right) - cov_t\left(M^{\$}_{t+1} R^{\$}_{t+1}, R^{\$}_{t+1}\right), \quad (13)$$

where $R^{\$}_{t+1}$ and $R^{\$}_{f,t+1}$ are the market return and the risk-free rate over the period from t to $t+1$, the SDF corresponding to the same time period is $M^{\$}_{t+1}$, the asterisk means that the variance is calculated under the risk-neutral probability measure, the t subscript denotes conditioning on the information available at time t, and the dollar sign in the superscript emphasizes that the numeraire is the dollar (and not the dividend account). The strength of the decomposition in (13) – besides its model-free nature – lies in the fact that the first term on the right-hand side can be observed directly in the market in real time (via option prices), while the sign of the second term can be "controlled" by economic reasoning. As Martin (2017) argues, in all mainstream asset pricing models the sign of the covariance in (13) is negative. As a consequence, the risk-neutral variance of the market return scaled by the risk-free rate serves as a lower bound for the expected excess market return. Then, Martin (2017) shows that empirically the magnitude of the covariance term is small, thus the lower bound is a good approximation for the true expected excess market return.

Our Arbitrage-Based Recovery helps understand the essence of Martin's negative correlation condition. Concretely, we can show that the negative correlation condition is actually a condition on the correlation between the martingale component of the Đ-SDF (in a framework

[9] But, importantly, we do not assume any form of preference or even the existence of a representative agent. We only assume that the SDF $\boldsymbol{m}^{Đ}$ is transition independent. The particular functional form of our inferred SDF (both under the dollar and the Đ-numeraire) is then a direct consequence of the no-arbitrage condition.

[10] This is a special case of a more general, recent result by Borovička and Stachurski (2021). We thank the referee for pointing this out.

where the state is spanned by a closed, bounded price process) and the market return. To see this, note that the true SDF under the Đ-numeraire can be written as

$$M^{Đ}_{t+1} = \frac{S_t}{S_{t+1}} \times H^{Đ}_{t+1}, \tag{14}$$

where $H^{Đ}_{t+1}$ is a positive random variable with expected value of one, also known as the *martingale component* of the Đ-SDF. Dividing both sides by the gross dividend yield F_{t+1}, we obtain

$$\frac{M^{Đ}_{t+1}}{F_{t+1}} = \frac{S_t}{S_{t+1} \times F_{t+1}} \times H^{Đ}_{t+1}, \tag{15}$$

which can be equivalently written as

$$M^{\$}_{t+1} \times R^{\$}_{t+1} = H^{Đ}_{t+1}. \tag{16}$$

Hence, Martin's covariance term can be written as

$$cov_t\left(M^{\$}_{t+1} R^{\$}_{t+1}, R^{\$}_{t+1}\right) = cov_t\left(H^{Đ}_{t+1}, R^{\$}_{t+1}\right). \tag{17}$$

The negative correlation condition is thus, in essence, equivalent to the condition that the martingale component of the SDF (under the dividend account numeraire) is negatively correlated with the total market return (under the dollar numeraire).

Since Arbitrage-Based Recovery assumes that the permanent component of the Đ-SDF is constant, its implied probabilities correspond to an expected return which is exactly equal to the lower bound of Martin (2017). Hence, Arbitrage-Based Recovery can also be considered as an extension of Martin's approach: while Martin (2017) recovers the first moment of the stock market return from option prices, Arbitrage-Based Recovery extracts a consistent probability distribution of the stock market return from the same option prices. And just as Martin's recovered expected returns are hard to reject empirically, in Section 7 we similarly find that the probability distributions implied by our Arbitrage-Based Recovery Theorem are difficult to reject empirically as well.

4. Long-term factorization

Ross's Recovery Theorem is closely related to the branch of the literature initiated by Alvarez and Jermann (2005) which concerns the long-term factorization of the stochastic discount factor and the corresponding long-forward probability measure. As Alvarez and Jermann (2005) and Hansen and Scheinkman (2009) show, the stochastic discount factor can be decomposed into the product of a *transitory component* and a *permanent component*. The transitory SDF component, in turn, is equal to the reciprocal of the return on the *long bond*, i.e., on a zero-coupon bond the maturity of which is in the infinity.[11] Since Ross's Recovery Theorem actually recovers the transitory SDF component (as shown by Borovička et al., 2016), the SDF recovered by Ross is also equal to the reciprocal of the return on the long bond. In this section, we show that this interpretation carries over to our Arbitrage-Based Recovery approach, but under the dividend numeraire Đ (instead of the dollar numeraire). In other words, our inferred SDF is equal to the reciprocal of the return on a zero-coupon bond denominated in units of Đ and with maturity in the infinity. In the remainder of this section, we demonstrate this in more details.

If we take the dividend account as the numeraire, the equivalent of a zero-coupon bond with maturity in T periods is a security which will pay one unit of the dividend account in each state in T periods. We call this asset the *T-period zero-coupon Đ-bond*. For example, if the initial state at $t = 0$ is i and in T periods we arrive at state m by taking the path $(i \to j \to k \to \ldots \to l \to m)$, then the T-period zero-coupon Đ-bond will pay $F_{i,j} \times F_{j,k} \times \cdots \times F_{l,m}$ dollars, which is equivalent to one unit of the dividend account then. Now, we show that the recovered probabilities in (10) indeed correspond to the long-forward probabilities under the Đ-numeraire.

Corollary 4. *The probabilities recovered by the Arbitrage-Based Recovery Theorem correspond to the long-forward probabilities under the Đ-numeraire.*

Proof. If we go from state i to state j, the one-period gross return on the Đ long bond is

$$R^{Đ}_{\infty,i,j} = \lim_{T\to\infty} \frac{\sum_{k=1}^{n}\left[\left(A^{Đ}\right)^{T-1}\right]_{j,k}}{\sum_{k=1}^{n}\left[\left(A^{Đ}\right)^{T}\right]_{i,k}}, \tag{18}$$

where we raised the one-period Arrow–Debreu price matrix $\boldsymbol{A}^{Đ}$ to the power of T and $T-1$, to obtain the T- and the $(T-1)$-period Arrow–Debreu price matrices. According to the Perron–Frobenius theorem, we have

$$\lim_{T\to\infty}\left(\boldsymbol{A}^{Đ}\right)^{T} = \boldsymbol{S}\boldsymbol{w}', \tag{19}$$

where $\boldsymbol{w}$ is the left Perron–Frobenius eigenvector of $\boldsymbol{A}^{Đ}$. Hence, the return in (18) can be written as

$$R^{Đ}_{\infty,i,j} = \frac{S_j \sum_{k=1}^{n} w_k}{S_i \sum_{k=1}^{n} w_k} = \frac{S_j}{S_i}. \tag{20}$$

Using the identity $\boldsymbol{A}^{Đ} = \boldsymbol{A}^{\$} \odot \boldsymbol{F}$, we can express the recovered probabilities in (10) as

$$\Pi_{i,j} = A^{Đ}_{i,j} \times \frac{S_j}{S_i}. \tag{21}$$

Substituting $R^{Đ}_{\infty,i,j}$ in place of S_j/S_i in (21) and then dividing both sides by $R^{Đ}_{\infty,i,j}$, we readily see that the recovered probabilities indeed correspond to the long-forward probabilities under the Đ-numeraire. This completes the proof. ■

5. Arbitrage-based recovery in continuous space

So far, we have assumed that the only state variable in our model is the stock price, and that it can take only finitely many values. Consequently, our pricing operator has been represented by a square matrix of positive Arrow–Debreu prices, and our recovery approach has relied on the uniqueness of the positive eigenvector (and the corresponding positive eigenvalue) of this matrix.

Although one might argue that stock prices are quoted using discrete price increments ("ticks"), there is no economic rationale behind assuming that prices can take only certain (discrete) values. The good news is that our framework can readily be extended to accommodate a continuous state space represented by the price of a traded asset, as we now show in this section.

We still consider a discrete-time financial market where the only state variable is the stock price, but this price can now take any real value in the interval $\left[\underline{S}, \overline{S}\right]$ where $0 < \underline{S} < \overline{S}$. There exists a linear positive operator (the *pricing operator*) $\mathscr{A}^{\$}$ mapping the set of bounded real-valued functions with domain $\left[\underline{S}, \overline{S}\right]$ into the same set. The pricing operator can be represented as

[11] The essence of this result predates the emergence of the literature on recovery theorems and on the SDF decomposition and was first pointed out by Kazemi (1992). Importantly, the result itself does not rely on either Markovianity of the model or the Perron–Frobenius Theorem. Further discussion can be found in Alvarez and Jermann (2005) and Martin and Ross (2019). Qin and Linetsky (2017) generalize the result to semimartingale settings.

$$\left(\mathcal{A}^{\$} f\right)(x) = \int_{\underline{S}}^{\overline{S}} A^{\$}(x, y) f(y) dy, \tag{22}$$

where $A^{\$}(x, y) : \left[\underline{S}, \overline{S}\right] \times \left[\underline{S}, \overline{S}\right] \to \mathbb{R}_{>0}$ is the dollar state-price density going from state $S_t = x$ to $S_{t+1} = y$. Similarly to the gross dividend yield matrix $\boldsymbol{F}$ defined in (1) in the discrete-state case, now there exists a gross dividend yield function $F(x, y) : \left[\underline{S}, \overline{S}\right] \times \left[\underline{S}, \overline{S}\right] \to \mathbb{R}_{\geq 1}$ representing the gross dividend yield if we go from state $S_t = x$ to $S_{t+1} = y$. The state-price density under the Đ-numeraire is defined as $A^{Đ}(x, y) = A^{\$}(x, y) \times F(x, y)$, which, in turn, defines the pricing operator under the Đ-numeraire, $\mathcal{A}^{Đ}$, accordingly. If there is no arbitrage opportunity in the market, then

$$\mathcal{A}^{Đ} f(x) = f(x) \tag{23}$$

must hold for $f(x) = x$. I.e., absence of arbitrage implies that $f(x) = x$ is an eigenfunction of the operator $\mathcal{A}^{Đ}$ with the corresponding eigenvalue being equal to one.

To show that our results developed in a discrete state space also hold when the state space is continuous (i.e., when the stock price can take any real value in the interval $\left[\underline{S}, \overline{S}\right]$), it suffices to show that $f(x) = x$ is the unique (up to a multiplicative constant) positive eigenfunction of the operator $\mathcal{A}^{Đ}$. We formalize this in the following theorem.

Theorem 4. *The pricing operator $\mathcal{A}^{Đ}$ has a unique (up to a multiplicative constant) positive eigenfunction, namely, $f(x) = x$. Furthermore, the eigenvalue corresponding to this eigenfunction is equal to one.*

Proof. We provide the proof in the Online Appendix. ■

When the stock price is not restricted to the interval $\left[\underline{S}, \overline{S}\right]$, but it can take any positive real value, the recovery problem can still be formulated. Then, however, the pricing operator $\mathcal{A}^{Đ}$ might have several different positive eigenfunctions, in which case unique recovery is not possible. Whether there exists a unique (up to a multiplicative constant) positive eigenfunction, depends on the exact properties of the pricing operator $\mathcal{A}^{Đ}$, besides its linearity and positivity. Importantly, even if the recovery problem has several distinct solutions, the solution proposed by our Arbitrage-Based Recovery approach will always be among the solutions. Furthermore, if the recovery problem has a unique solution, then our approach will provide this unique solution.

6. Price–dividend ratio as the state variable

In our frameworks so far, we have assumed that the only state variable is the price of a traded asset, and the state space is bounded. The assumption that prices can take values only in a given interval is, however, not supported empirically. Assuming instead that the price–dividend ratio can take values in an interval is more plausible. The good news is that the principles of our Arbitrage-Based Recovery approach can also be applied in an environment where the state variable is the price–dividend ratio (instead of the price itself). Furthermore, our recovered probability distribution of the price–dividend ratio is consistent with our previously recovered probability distribution of the price. In this section, we show how Arbitrage-Based Recovery can be adapted to a framework with the price–dividend ratio as the state variable. We only show the continuous-state-space case; the discrete-state-space case can be treated analogously.

Consider a discrete-time economy where the only state variable is the price–dividend ratio, $X \in \mathbb{R}_{>0}$, which can take any real value in the interval $\left[\underline{X}, \overline{X}\right]$ where $0 < \underline{X} < \overline{X}$. There exists a linear positive operator (the *pricing operator*) mapping the set of bounded real-valued functions with domain $\left[\underline{X}, \overline{X}\right]$ into the same set. This operator can be represented as

$$\left(\mathcal{A}^{\$} f\right)(x) = \int_{\underline{X}}^{\overline{X}} A^{\$}(x, y) f(y) dy, \tag{24}$$

where $A^{\$}(x, y) : \left[\underline{X}, \overline{X}\right] \times \left[\underline{X}, \overline{X}\right] \to \mathbb{R}_{>0}$ is the dollar state-price density if we go from state x to state y. There exists a dividend growth function $h(x, y) : \left[\underline{X}, \overline{X}\right] \times \left[\underline{X}, \overline{X}\right] \to \mathbb{R}_{>0}$ representing the dividend growth D_{t+1}/D_t if we go from state x (at time t) to state y (at time $t+1$).

Similarly to previous sections, also now we propose a change of numeraire. Since our state variable is the post-dividend stock price scaled by the dividend itself (i.e., $X_t = S_t/D_t$), we can define a corresponding state price density $A^{\$}(x, y) \times h(x, y)$ which also uses the dividend as numeraire. While $A^{\$}(x, y)$ expresses the rate of exchange between one dollar today and one dollar in a future state, $A^{\$}(x, y) \times h(x, y)$ is the rate of exchange between one unit of dividend today and one unit of dividend in a future state. We can also go one step further and define the state price density $A^{Ǥ}(x, y) \triangleq A^{\$}(x, y) \times h(x, y) \times F(y)$, where $F(y) = 1 + 1/y$ corresponds to the $\boldsymbol{F}$ matrix in our discrete-state-space framework in Section 2 and to the F function in our continuous-state-space framework in Section 5, evaluated at the arrival state y. Since the state variable now is the price–dividend ratio itself (as opposed to the earlier sections, where the state variable was the price), F is a function of the arrival state only, and it is independent of the departure state. The state price density $A^{Ǥ}(x, y)$ expresses the rate of exchange between one unit of dividend today and F units of dividend in a future state. The pricing operator under the Ǥ-numeraire, $\mathcal{A}^{Ǥ}$, is defined accordingly as

$$\left[\mathcal{A}^{Ǥ} f\right](x) = \int_{\underline{X}}^{\overline{X}} A^{Ǥ}(x, y) f(y) \mathrm{d}y. \tag{25}$$

Recovery under the Ǥ-numeraire entails solving the eigenvalue–eigenfunction problem

$$\left[\mathcal{A}^{Ǥ} g\right](x) = \gamma g(x) \tag{26}$$

for $g(x) > 0$ and $\gamma > 0$. Note that if there is no arbitrage opportunity, then

$$x = \int_{\underline{X}}^{\overline{X}} A^{\$}(x, y) h(x, y) F(y) y \,\mathrm{d}y \tag{27}$$

must hold, which can equivalently be expressed as

$$x = \int_{\underline{X}}^{\overline{X}} A^{Ǥ}(x, y) y \,\mathrm{d}y. \tag{28}$$

Therefore, absence of arbitrage implies that $g(x) = x$ and $\gamma = 1$ is a solution of the eigenfunction–eigenvalue problem (26). Using the same reasoning as in Theorem 4, it can be shown that this is the only solution. The corresponding conditional probability density of the price–dividend ratio is

$$\pi(x, y) = A^{Ǥ}(x, y) \times \frac{y}{x}, \tag{29}$$

where x is the price–dividend ratio at the departure state and y is the price–dividend ratio at the arrival state. This density, however, can equivalently be expressed as

$$\pi(x, y) = A^{\$}(x, y) \times R(x, y), \tag{30}$$

where $R\left(X_t, X_{t+1}\right) = \left(D_{t+1} + P_{t+1}\right)/P_t$ is the gross total return on the security itself, if we go from state X_t to state X_{t+1}. This recovered probability measure is therefore consistent with the implications of our Theorem 2 (and, of course, our results in Section 5).

If we relax the assumption that the price–dividend ratio can take values only in the interval $\left[\underline{X}, \overline{X}\right]$ and allow it to take any positive real value, the recovery problem does not necessarily have a unique solution. Whether a unique solution exists depends on the exact properties of the pricing operator $\mathcal{A}^{Ǥ}$, besides its linearity and positivity. However, if the solution is unique, then it will be equal to our solution, otherwise there would be an arbitrage opportunity in the market. When multiple solutions exist and we assume that the price–dividend ratio is stationary and ergodic under the true physical probability measure, there is at most one solution under which the price–dividend ratio remains stationary and ergodic, as shown by Borovička et al. (2016). Whether there is such a solution (which retains stationarity and ergodicity) and whether it corresponds to our solution depends, again, on the exact properties of the operator $\mathscr{A}^{\mathcal{G}}$.

7. Empirical tests of arbitrage-based recovery

From a theoretical perspective, our Arbitrage-Based Recovery Theorem has several appealing properties. Not only does it require observing only one row of the Arrow–Debreu price matrix for only one single maturity, but it also explicitly excludes arbitrage opportunities.

Nonetheless, the proof of the pudding is in the eating, and the defense of a theory resides in its empirical validity. Jackwerth and Menner (2020) systematically test to what extent the different versions of recovery theorems are supported empirically. They test three different recovery theorem versions (Ross Original, Ross Basic, Ross Stable/Generalized Recovery), with and without imposing economic restrictions on the Arrow–Debreu price matrix. Based on 32 years of S&P 500 options data (from April 1986 to December 2017) and using several different density tests (the Berkowitz test, two forms of the Knüppel test, and the Kolmogorov–Smirnov test), they find strong evidence that the one-month S&P 500 returns do not follow the distribution predicted by the recovery theorems. These findings are also supported by mean-prediction and variance-prediction tests. We test our Arbitrage-Based Recovery Theorem following an approach similar to Jackwerth and Menner (2020), using more than 26 years of S&P 500 options data. Our results are summarized in Table 5.

Before describing our empirical exercise in details, we would like to point out a duality regarding how our empirical results can be interpreted. Having access to a time series of the one-month-ahead S&P 500 index level state price density, we apply our methodology to recover the probability distribution of the index level in one month. As our theoretical results in Sections 5 and 6 show, this empirical exercise can be motivated and approached from two different perspectives. Concretely, either we assume that the state variable is the index level itself or we assume that the state variable is the price–dividend ratio. Since the dividend is assumed to be known one period in advance (as we will explain later), the recovered probability densities will be exactly the same, only the random variable (i.e., the index level or the price–dividend ratio) is scaled by a constant. Similarly, the results of our empirical tests will be exactly the same, regardless of whether the index level or the price–dividend ratio is the state variable. For brevity, in this section we will refer to our exercise as one attempting to recover the probability distribution of the index level itself, keeping in mind that the exercise and its results can also be interpreted as one attempting to recover the price–dividend ratio probability distribution.

7.1. Data

We collect prices of European-style options written on the S&P 500 index from January 19, 1996 to May 19, 2022. The options expire at the market opening of the third Friday[12] of each month, and we observe their bid and ask prices 29 days before their expiry.[13] Our source of option price data from January 19, 1996 to December 23, 2021 is OptionMetrics, and for 2022 observations it is Bloomberg. We discard all options with a bid price lower than \$0.50, and then we calculate the mid option prices. We also collect the S&P 500 price index values from OptionMetrics[14] and the S&P 500 total return index values from Yahoo Finance for each observation date and for each option exercise date.

Altogether, we have 317 observation dates, one in each month of our sample period. On each observation date, we determine the 28-day continuously-compounded risk-free interest rate by linearly interpolating between the two closest available maturities.[15] Furthermore, using the S&P 500 price index and the S&P 500 total return index data, we calculate the dividend received during each 28-day period.

7.2. Risk-neutral probabilities

On each observation date, we first obtain the risk-neutral probability density function (pdf) characterizing the risk-neutral probability distribution of the S&P 500 index value 28 days later. To this end, we first transform each option price into its implied volatility, using the Black–Scholes option pricing formula.[16] As standard in the literature, we use only out-of-money (OTM) options. To handle the "jump" of the implied volatility function at the moneyness level of one, we adapt an approach similar to Figlewski (2010) to smooth the implied volatilities around the current S&P 500 index level. Then, applying a slightly modified version of the *Fast and Stable Method* of Jackwerth (2004),[17] we fit a smooth curve on the "observed" implied volatilities for each observation day, using a fine grid of $\Delta = \$0.10$ along the strike price dimension between the lowest and the highest traded strike prices. For demonstration purposes, in Fig. 3, we depict our implied volatility curves (together with the "observed" implied volatilities) for three observation dates: August 20, 2004; November 21, 2008; and April 16, 2020.

Transforming back each implied volatility along the grid into a European call option price and then taking the (numeric) second derivative of the option price with respect to the strike price, we obtain the state-price density for each observation day. For details on why the second derivative of a European call option price with respect to the strike price is equal to the state-price density, we refer to Ross (1976) and Breeden and Litzenberger (1978). Dividing the state-price densities by the risk-free discount factors, we obtain the risk-neutral densities. Finally, to add tails to the risk-neutral density functions below the lowest and above the highest traded strike prices, we follow Figlewski (2010) and choose appropriately parameterized generalized extreme value (GEV) distributions. We provide more details on our methodology of adding tails to the risk-neutral density curves in Appendix A. In Fig. 4, we plot the risk-neutral probability density functions (as dotted curves) for the same three dates as we used in Fig. 3 for the implied volatilities. The square markers indicate the lowest and the highest traded strike prices for each date.

7.3. Recovering the physical probabilities

The Arrow–Debreu price is the product of the physical probability and the stochastic discount factor, and our ultimate goal is to separately

[12] Until February 15, 2015, our observed options expire at the market opening of the Saturday immediately following the third Friday of the month.

[13] Due to the options expiring at the market opening, this corresponds to a horizon of 28 days.

[14] For 2022, our S&P 500 price index value data is obtained from Yahoo Finance.

[15] For observations between 1996 and 2021, our spot yield curve data is from OptionMetrics. For 2022, we use the 4-week Treasury Bill secondary market rate, provided by the St. Louis Fed.

[16] Importantly, we do not assume that the assumptions of the Black–Scholes model hold. We use the Black–Scholes formula only as a tool which provides a one-to-one mapping between option prices and implied volatilities.

[17] When calculating the numerical derivatives of the fitted volatility curve, Jackwerth (2004) assumes that the volatility corresponding to an exercise price just one grid step outside the boundary is equal to the implied volatility on the boundary. To allow for larger flexibility, our boundary condition instead assumes that the second derivatives on the boundaries are zero.

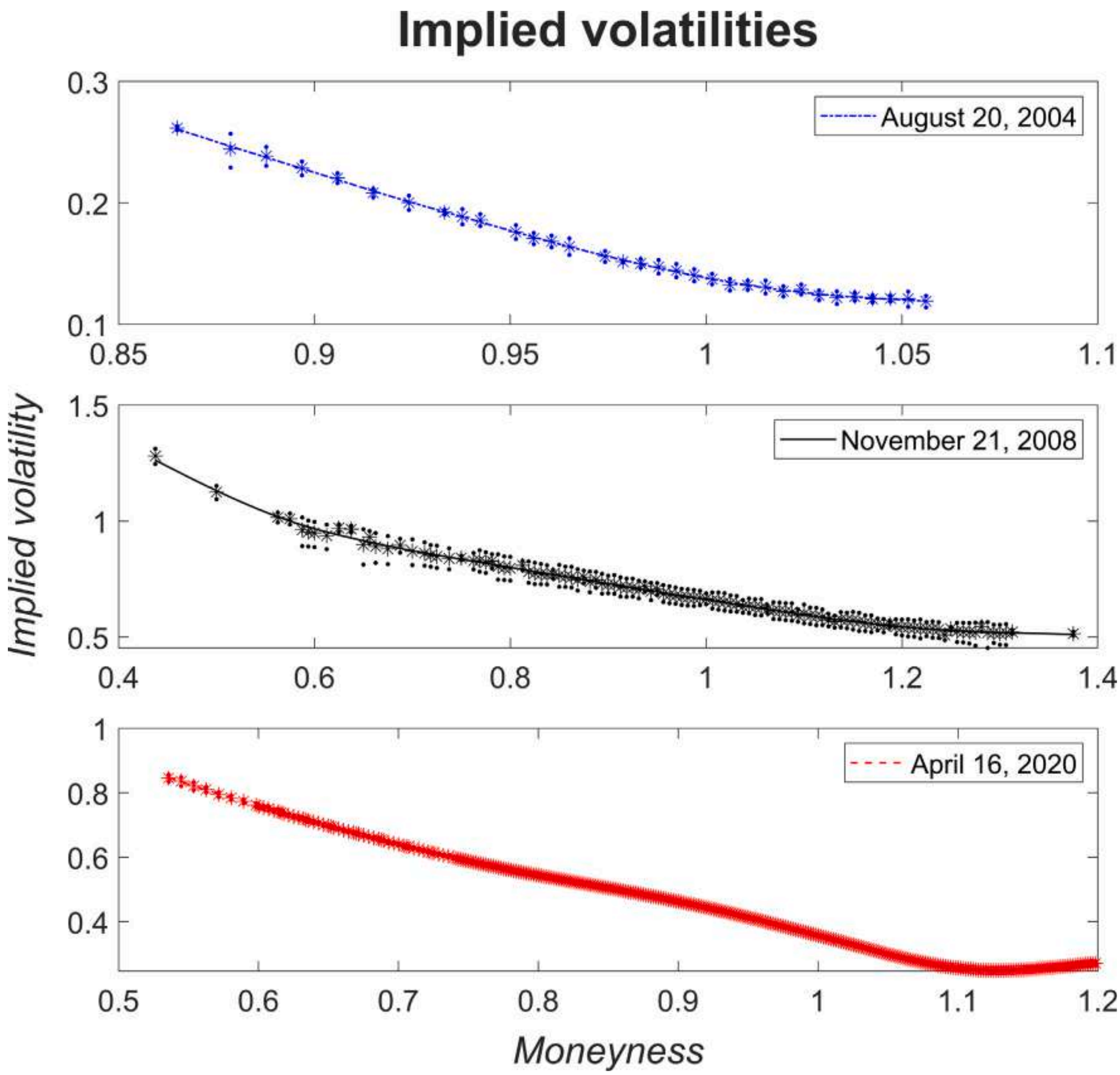

Fig. 3. Implied volatility curves on three dates, calculated from European-style S&P 500 index options with 28 days to maturity. The "*" markers indicate implied volatilities corresponding to mid option prices, while the "·" markers denote implied volatilities corresponding to bid and ask option prices. Implied volatilities are calculated by the Black–Scholes option pricing formula. The 28-day risk-free interest rates are obtained by linearly interpolating between the two risk-free rates with the closest maturities (e.g., 25 days and 31 days). The dividend yield parameter of the Black–Scholes formula is calculated under the assumption that the dividend is received just before the option exercise date, and the dollar amount of this dividend is known on the observation date. To interpolate between the observation points on the implied volatility curves, we use a slightly modified version of the *Fast and Stable Method* of Jackwerth (2004).

identify these two components, based on our Arbitrage-Based Recovery Theorem. In the previous step, we have determined the Arrow–Debreu prices.[18] But to apply our theorem, we still need to determine the gross dividend yield $F_{i,j}$. Since we want to recover the S&P 500 index level over a one-month horizon, we can reasonably assume that the dividend to be received in one month is deterministic and known at the time of observation.[19] In Fig. 4, we show the physical probability density functions recovered by our model on the three dates used in the previous subsection. For comparison purposes, we also show the risk-neutral densities as dotted curves. The first date, August 20, 2004, reflects relaxed market conditions, with implied volatilities between 12 and 26 percent for the observed options. The recovered physical probability density function looks "standard": it is centered around a level slightly higher than one, and it features a slight left skew. The two other dates correspond to crisis periods, with much higher implied volatilities. By 21 November, 2008, the severity of the global financial crisis became evident, and the strong negative skewness of the density function indicates that the market was prepared for further serious declines in the S&P 500 level. A high (more than 10% per month) increase in the index level had also much higher probability than on the "relaxed" date of 20 August, 2004, which reflects the huge uncertainty perceived by the market. The situation was similar on 16 April, 2020, by when the seriousness of the COVID-19 pandemic on a global level became apparent. Comparing the two crises dates, the market during the pandemic attributed a much lower probability to a large positive monthly jump in the S&P 500 index level, which is reflected in the difference between the right tails of the two "crisis" distributions. The effect of risk adjustment is much more apparent on the two crisis dates, reflected in the observable differences between the risk-neutral and the physical density functions. Risk adjustment has a more substantial effect in "very bad" and "very good" states (corresponding to low and high moneyness levels, respectively), and its effect is more negligible around a moneyness level of one. This is intuitive from an economic

[18] We can readily calculate the Arrow–Debreu prices from the state price densities by integrating the state price density function between any two strike prices (or moneyness levels).

[19] As Chetty et al. (2005) note, firms usually announce dividend payments about four to six weeks before the actual payment takes place.

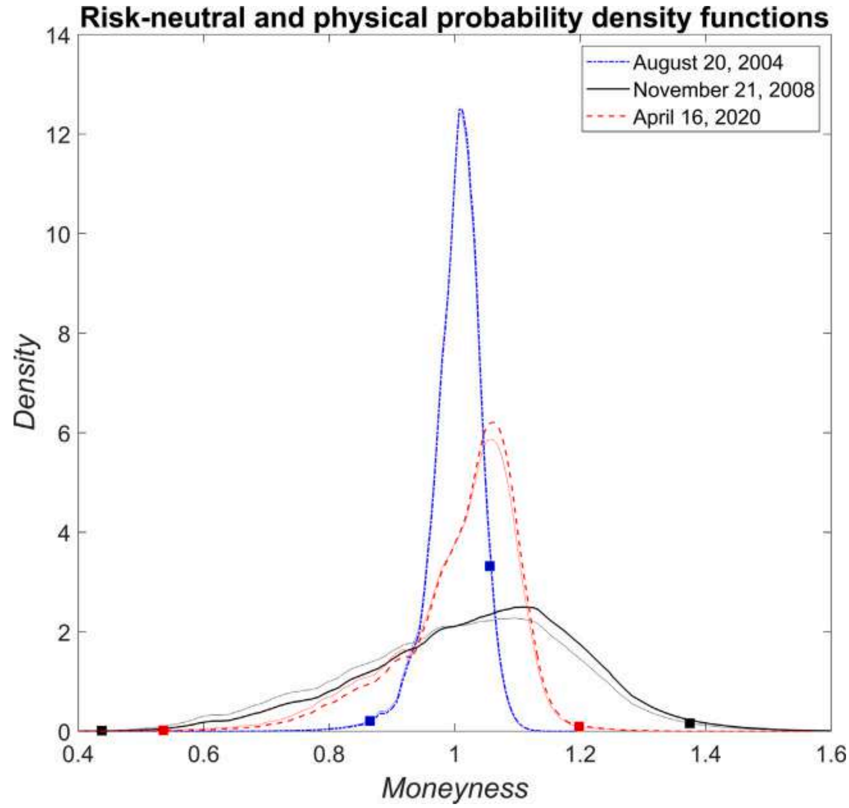

Fig. 4. Risk-neutral and physical probability density functions on three dates, calculated from European-style S&P 500 index options with 28 days to maturity. The risk-neutral probability density functions are shown as dotted curves, while the dash-dotted, the solid, and the dashed curves correspond to the physical probability density functions. The square markers indicate the lowest and the highest traded strike prices for each date. Between these two strike prices, the risk-neutral densities are obtained as the second derivatives of the call option price with respect to the strike price, divided by the risk-free discount factor. Call price curves are obtained as transforms of the implied volatility curves in Fig. 3, and the 28-day risk-free discount factors are obtained by linearly interpolating between the two risk-free interest rates with the closest maturities (e.g., 25 days and 31 days). Then, following Figlewski (2010), we add tails to the risk-neutral probability density functions by choosing appropriately parameterized generalized extreme value distributions. Afterwards, we apply the Arbitrage-Based Recovery Theorem to transform these risk-neutral probability density functions into physical probability density functions. During this transformation, we assume that the dividend is received just before the option exercise date, and the dollar amount of this dividend is known on the observation date.

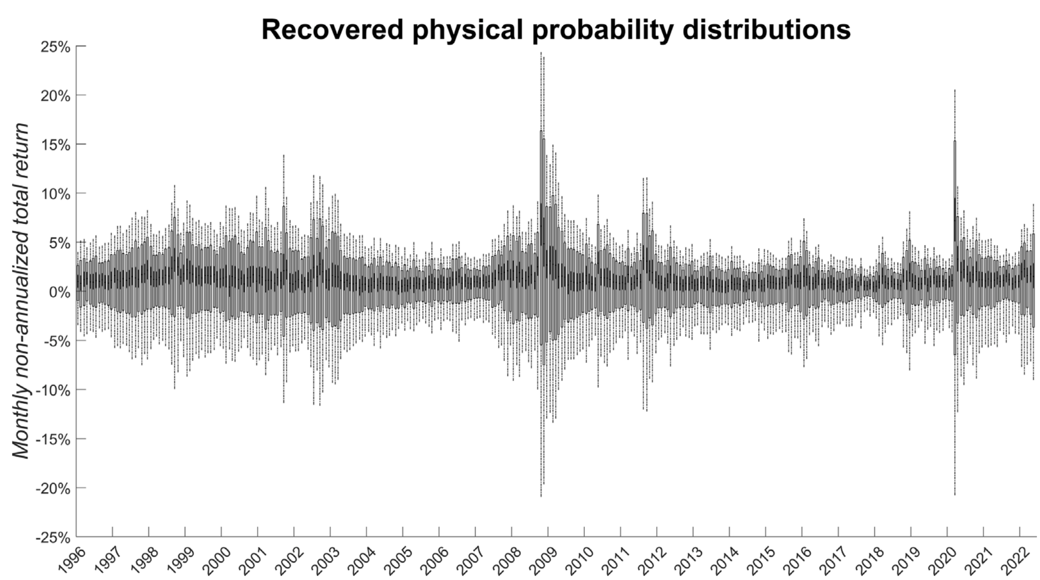

Fig. 5. Recovered physical probability distributions of the 1-month non-annualized total returns on the S&P 500 market index, calculated from European-stye S&P 500 A.M.-settled index options with 29 days to maturity. The dash-dotted lines denote returns lying between the 10th and the 90th percentiles, the boxes denote returns between the 25th and the 75th percentiles, and the solid lines inside the boxes correspond to returns between the 45th and the 55th percentiles. First, we obtain a smooth implied volatility curve for each observation date by applying a slightly modified version of the *Fast and Stable Method* of Jackwerth (2004), which we then transform back into risk-neutral probability density functions. We attach tails to these risk-neutral density functions by choosing appropriately parameterized generalized extreme value distributions, following Figlewski (2010). Then, we transform the risk-neutral probability distributions into physical probability distributions by applying the Arbitrage-Based Recovery theorem. Throughout, we assume that the dividend is received just before the option exercise date, and the dollar amount of the dividend is known on the observation date. The 28-day risk-free interest rates are obtained by linearly interpolating between the two risk-free rates with the closest maturities (e.g., 25 days and 31 days) (until 2021). For 2022, we use the 4-week Treasury Bill secondary market rate as the risk-free rate.

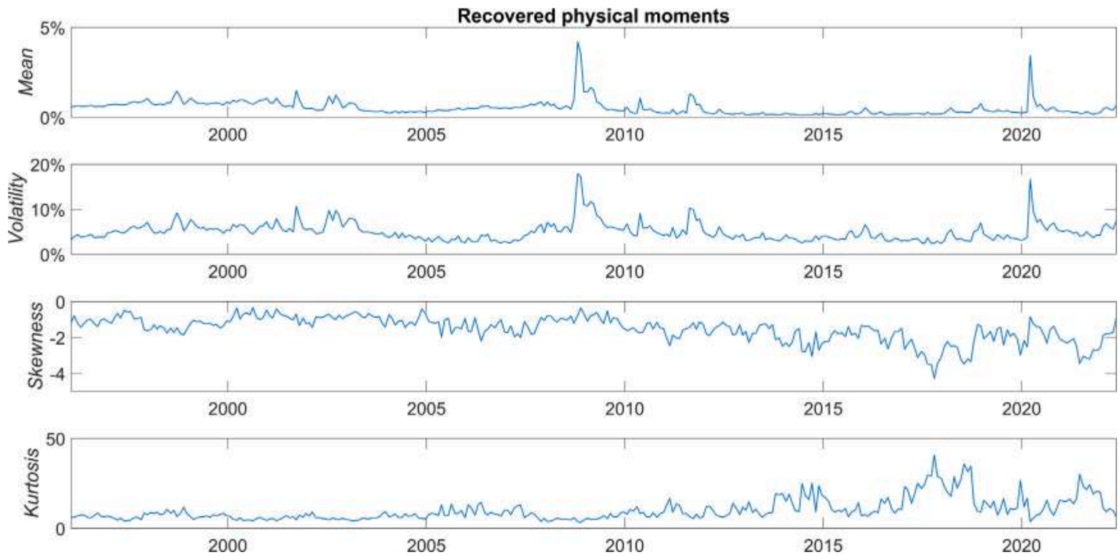

Fig. 6. Recovered physical moments of the 1-month non-annualized total return on the S&P 500 market index. We obtain the physical probability distributions by applying the Arbitrage-Based Recovery Theorem, as shown in Fig. 5. Then, for each observation date, we calculate the expected value (mean), the standard deviation (volatility), the skewness, and the kurtosis of the return. The recovered physical moments suggest that throughout our observation period, the probability distribution of the market return is left-skewed and it features fat tails. Both its mean and its volatility are highly volatile. Furthermore, both the magnitude and the volatility of its skewness and kurtosis increased substantially after the 2008 financial crisis. The mean and the volatility of the return are highly correlated, which is in accord with economic intuition that higher risk (volatility) should be compensated with higher reward (mean).

Table 1
Summary statistics of recovered physical moments of the 1-month non-annualized total return on the S&P 500 market index. From January 1996 to May 2022, we obtain the physical probability distributions with monthly frequency by applying the arbitrage-based recovery theorem, as shown in Fig. 5. Then, for each observation date, we calculate the expected value (mean), the standard deviation (volatility), the skewness, and the kurtosis of the return.

	Median	Expected value	Standard deviation	Minimum	Maximum
Mean	0.42%	0.52%	0.43%	0.10%	4.20%
Volatility	4.64%	5.01%	2.12%	2.38%	17.90%
Skewness	−1.42	−1.53	0.69	−4.27	−0.34
Kurtosis	8.03	10.14	6.29	3.37	40.78

Table 2
Cross-correlations of recovered physical moments of the 1-month non-annualized total return on the S&P 500 market index. From January 1996 to May 2022, we obtain the physical probability distributions with monthly frequency by applying the Arbitrage-Based Recovery Theorem, as shown in Fig. 5. Then, for each observation date, we calculate the expected value (mean), the standard deviation (volatility), the skewness, and the kurtosis of the return.

	Mean	Volatility	Skewness	Kurtosis
Mean	1	–	–	–
Volatility	0.87	1	–	–
Skewness	0.45	0.45	1	–
Kurtosis	−0.42	−0.44	−0.94	1

perspective: in "very bad" states (corresponding to low moneyness levels) market participants' marginal utility is high, hence they value cash flows in those states highly, which corresponds to attributing much higher risk-neutral probabilities to those states than physical probabilities. The opposite is true for "very good" states with high moneyness levels. These effects of the risk adjustment can be seen by comparing the risk-neutral density functions (dotted curves) and the recovered physical density functions (dash-dotted, solid, and dashed curves) in Fig. 4.

In Fig. 5, we show the recovered physical probability distributions of the 1-month non-annualized total returns on the S&P 500 market index for each month between January 1996 and May 2022. The day of recovery is always 29 days before the Saturday immediately following the third Friday of the month (until February 2015) or 29 days before the third Friday of the month (after February 2015). Due to all options being AM-settled, this corresponds to a 28-day recovery period. The dash-dotted lines denote returns between the 10th and the 90th percentiles, the boxes denote returns between the 25th and the 75th percentiles, and the solid lines inside the boxes correspond to returns between the 45th and the 55th percentiles. Return distributions are in general left skewed with a median slightly above 0%, and they clearly exhibit time-varying volatility. This is further confirmed in Fig. 6, where we plot the time series of the recovered physical moments; and in Table 1, where we provide summary statistics of the recovered physical moments.

The expected (total) return on the market index is on average 0.52% per month (about 6% per year), and it is extremely volatile, ranging from 0.10% per month during relaxed market conditions to 4.20% per month in October 2008. Furthermore, in untabulated results, we confirm the finding of Martin (2017) that the distribution of the expected return itself is right skewed (its skewness is 4.21). The volatility of the (monthly) market return is also highly volatile, ranging from 2.38% to 17.90%, with an average level of 5.01%. We find a strong positive correlation (Table 2) between recovered physical expected returns and volatilities, which is in line with economic intuition: a higher risk (higher volatility) must be compensated by a higher expected return.

The recovered monthly return distributions exhibit negative skewness and positive kurtosis throughout our observation period, which is in accordance with the empirical literature. The skewness and the kurtosis are strongly negatively correlated, which suggests that fat tails tend to coincide with left skews. Interestingly, return distributions which are more left skewed and which exhibit fatter tails tend to occur during low-volatility and low-expected-return periods. This is in accord with the recent findings of Gormsen and Jensen (2020). Furthermore, as Fig. 6 suggests, both the magnitude and the volatility of the return skewness and kurtosis increased substantially after the 2008 financial crisis.

Fig. 7 plots the Sharpe ratio of the monthly non-annualized total return on the S&P 500 index, calculated on each observation date under the physical probabilities implied by the Arbitrage-Based Recovery Theorem. First, for each observation date, we calculate the expected

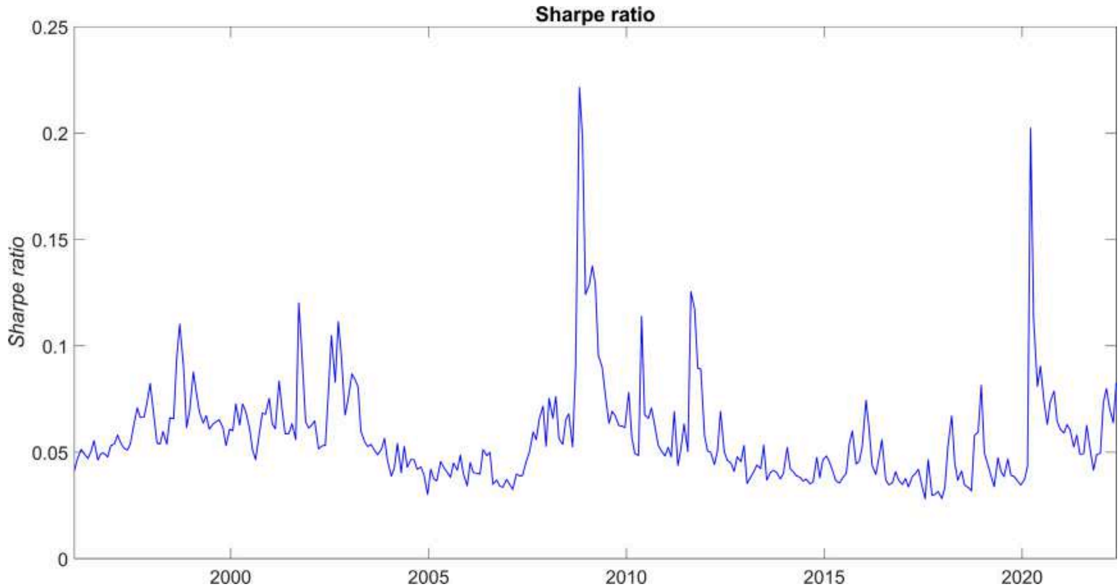

Fig. 7. Sharpe ratio of the 1-month non-annualized total return on the S&P 500 market index. We obtain the physical probability distributions by applying the Arbitrage-Based Recovery Theorem, as shown in Fig. 5. Then, for each observation date, we calculate the expected value and the standard deviation of the market return. The Sharpe ratio is then obtained by subtracting the 1-month risk-free rate from the expected market return and dividing the difference by the volatility of the market return. The Sharpe ratio is volatile, and its value is especially high at the beginning of crisis periods (October and November 2008, and March 2020).

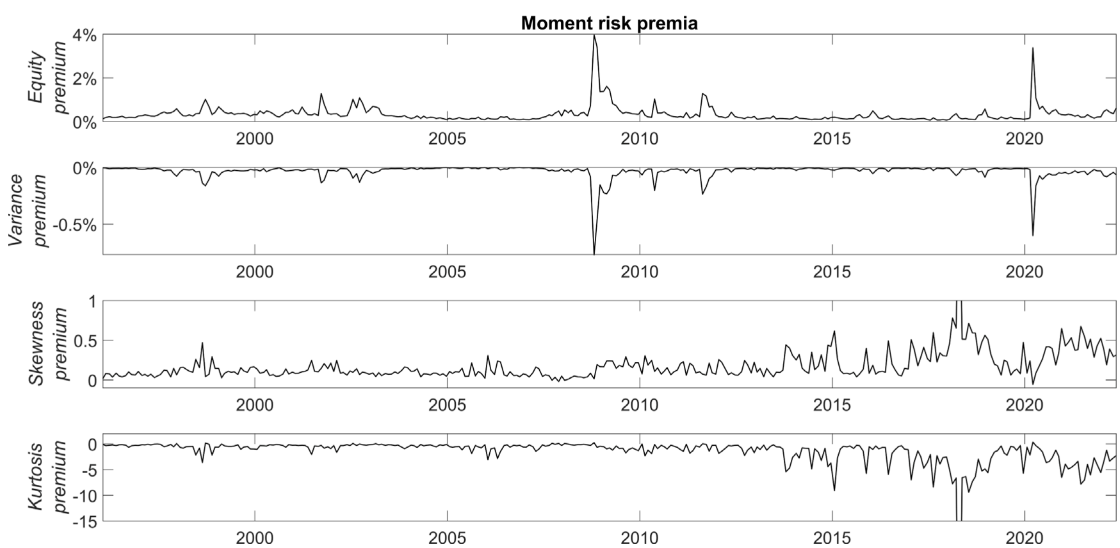

Fig. 8. Moment risk premia of the 1-month non-annualized total return on the S&P 500 market index. We obtain the physical probability distributions by applying the Arbitrage-Based Recovery Theorem, as shown in Fig. 5. Then, for each observation date, we calculate the expected value, the variance, the skewness, and the kurtosis of the return under both the risk-neutral and the recovered physical probability measures. The moment premia are then obtained as the difference between the physical and the risk-neutral moments. In accord with economic intuition, the equity premium is positive and the variance premium is negative throughout our observation period, and the skewness premium is almost always positive while the kurtosis premium is almost always negative. The equity premium and the variance premium are strongly negatively correlated, just as the skewness premium and the kurtosis premium. Similarly to the Sharpe ratio, the magnitudes of the equity premium and the variance premium are especially high at the beginning of crisis periods (October and November 2008, and March 2020). Furthermore, similarly to the physical skewness and kurtosis, both the magnitude and the volatility of the skewness premium and the kurtosis premium increased substantially after the 2008 financial crisis.

value and the standard deviation of the non-annualized market return under the recovered probability measure. The Sharpe ratio is then obtained by subtracting the 1-month non-annualized risk-free rate from the expected market return and dividing the difference by the volatility of the market return. The Sharpe ratio is highly volatile, and its value is especially high at the beginning of crisis periods (October and November 2008, and March 2020).

We also calculate the risk premia of the first four moments (equity premium, variance premium, skewness premium, and kurtosis premium), defined as the difference between the physical and the risk-neutral moments of the 1-month non-annualized total returns on the S&P 500 market index. Fig. 8 shows the time series of the moment risk premia, while Tables 3 and 4 show their summary statistics and their cross-correlations, respectively. The signs of the moment risk premia are in accordance with economic intuition (see, e.g., Ebert, 2013): odd moments have positive risk premia, and even moments have negative risk premia.[20] Each of the four moment risk premia features strong time variation, their standard deviations being at least as high as the

[20] On every single observation date, the equity premium is positive and the variance premium is negative. The skewness premium is positive on more than 98% of our observation dates, while the kurtosis premium is negative on more

Table 3
Summary statistics of moment risk premia of the 1-month non-annualized total return on the S&P 500 market index. From January 1996 to May 2022, we obtain the physical probability distributions with monthly frequency by applying the arbitrage-based recovery theorem, as shown in Fig. 5. Then, for each observation date, we calculate the expected value, the variance, the skewness, and the kurtosis of the return under both the risk-neutral and the recovered physical probability measures. The moment premia are then obtained as the difference between the physical and the risk-neutral moments.

	Median	Expected value	Standard deviation	Minimum	Maximum
Equity premium	0.24%	0.34%	0.40%	0.07%	3.96%
Variance premium	−0.02%	−0.04%	0.07%	−0.77%	−0.002%
Skewness premium	0.11	0.18	0.21	−0.05	2.90
Kurtosis premium	−0.61	−1.55	3.60	−56.68	0.35

Table 4
Cross-correlations of moment risk premia of the 1-month non-annualized total return on the S&P 500 market index. From January 1996 to May 2022, we obtain the physical probability distributions with monthly frequency by applying the Arbitrage-Based Recovery Theorem, as shown in Fig. 5. Then, for each observation date, we calculate the expected value, the variance, the skewness, and the kurtosis of the return under both the risk-neutral and the recovered physical probability measures. The moment risk premia are then obtained as the difference between the physical and the risk-neutral moments.

	Equity premium	Variance premium	Skewness premium	Kurtosis premium
Equity premium	1	–	–	–
Variance premium	−0.96	1	–	–
Skewness premium	−0.10	0.00	1	–
Kurtosis premium	0.11	−0.02	−0.96	1

absolute value of their expected values. Similarly to the Sharpe ratio, the magnitudes of the equity premium and the variance premium are especially high at the beginning of crisis periods (October and November 2008, and March 2020). Furthermore, similarly to the physical skewness and kurtosis, both the magnitude and the volatility of the skewness premium and the kurtosis premium increased substantially after the 2008 financial crisis. The equity premium and the variance premium are strongly negatively correlated, as it is also confirmed empirically by Bollerslev et al. (2009). The skewness premium and the kurtosis premium are also strongly negatively correlated, which is in agreement with economic intuition: when the market requires a high compensation for a stronger left skew, it tends to require a high compensation for fatter tails as well.

To demonstrate a possible practical application of our Arbitrage-Based Recovery Theorem, we also perform a simple market-timing exercise. Starting with $1 on January 19, 1996, we invest our wealth according to three different investment strategies with monthly rebalancing, and we plot the evolution of our wealth (on a logarithmic scale) in Fig. 9. According to the first investment strategy (shown as the blue dashed curve), the total wealth is invested in the one-month risk-free asset and upon maturity it is always reinvested at the actual one-month risk-free interest rate. The second strategy (shown as the red dotted curve) corresponds to a buy-and-hold investment policy, where the total wealth is invested in the S&P 500 total return index on January 19, 1996, and it is kept there indefinitely. The third investment policy (shown as the black solid curve) is based on a market-timing strategy using the Arbitrage-Based Recovery Theorem. Concretely, we invest in the one-month risk-free asset and in the S&P 500 total return index so that the portfolio weight of the S&P 500 total return index is proportional to the Sharpe ratio of its one-month return, calculated by the Arbitrage-Based Recovery Theorem. To make the second (buy-and-hold) and the third (market-timing) strategies more directly comparable, we choose the constant of proportionality in the third strategy so that the average (over the observation period) portfolio weight of the market index is equal to one. We re-evaluate the portfolio weights monthly (on each observation date), and re-balance our portfolio accordingly. The difference between the wealth paths of the market-timing strategy and the other two strategies can be attributed to the effect of market timing based on the Arbitrage-Based Recovery Theorem. As we see in Fig. 9, the market-timing strategy outperforms not only the rolled-over (locally) risk-free portfolio, but also the total return market index.[21]

than 96% of our observation dates. In the few instances when the skewness premium is negative and/or the kurtosis premium is positive, their magnitude is always very small.

7.4. Density tests

Now, we test whether realized S&P 500 index levels on the option exercise dates are indeed drawn from our model-implied distributions. Following Jackwerth and Menner (2020), we perform several density tests and mean-prediction tests. Our null hypothesis is that the S&P 500 index levels observed on the option exercise dates are realizations of a random variable with our model-implied probability density function.

To perform our density tests, first we observe the S&P 500 index level on each option exercise day.[22] Then, we transform our model-implied physical probability density functions into cumulative distribution functions (CDFs). For each option, we determine the CDF value corresponding to the realized S&P 500 level. Under our null hypothesis, these CDF values are drawn from a uniform distribution with support $[0,1]$, and they are independent of each other. To test this hypothesis, we perform the Berkowitz test, two versions of the Knüppel test, and the Kolmogorov–Smirnov test.

7.4.1. Berkowitz test

To perform the test of Berkowitz (2001), we first transform our observed ("realized") CDF values using the inverse of the cumulative distribution function of the standard normal distribution. I.e., denoting our CDF value observations by x_t, we perform the transformation

$$z_t = \Phi^{-1}\left(x_t\right), \tag{31}$$

where Φ denotes the CDF of a standard normally distributed random variable. Then, under the null hypothesis, our observed z_t values follow standard normal distributions and they are independent of each other. This implies that in the AR(1) model

$$z_t - \mu = \rho\left(z_{t-1} - \mu\right) + \varepsilon_t, \tag{32}$$

under the null hypothesis we have $\mu = 0$, $\rho = 0$, and $Var\left(\varepsilon_t\right) = 1$. Using our z_t observations, we obtain the maximum likelihood estimates of μ, ρ, and $Var\left(\varepsilon_t\right)$ (denoted by $\hat{\mu}$, $\hat{\rho}$, and $\widehat{Var}\left(\varepsilon_t\right)$), and perform a likelihood ratio test where we compare the AR(1) model (32) with parameters $\hat{\mu}$, $\hat{\rho}$, and $\widehat{Var}\left(\varepsilon_t\right)$ to the same model with parameters $\mu = 0$, $\rho = 0$, and $Var\left(\varepsilon_t\right) = 1$. We find that the p-value of this test is 0.340.

[21] Our market-timing strategy invests in the stock market index heavily (taking highly leveraged positions) when the Sharpe ratio (implied by the Arbitrage-Based Recovery Theorem) is high, and less heavily (characterized by a balanced mix of the stock market index and the risk-free asset) during periods with a low Sharpe ratio. Consequently, during low-Sharpe-ratio periods, our wealth path closely tracks the S&P 500 total return index. On the other hand, during high-Sharpe-ratio periods, our wealth path exhibits higher volatility than the S&P 500 total return index, and these are the periods when the gap between the two wealth paths suddenly increases or decreases. This can be seen by comparing Figs. 7 and 9: the gap between the two wealth paths increases around 2002, 2012, and 2020, and it decreases around 2008, all of which periods are characterized by a high Sharpe ratio. Effectively, our featured market-timing strategy attempts to identify and reap the benefits of high-Sharpe-ratio periods.

[22] More precisely, since the options are AM-settled, we observe the closing level of the S&P 500 index on the previous trading day.

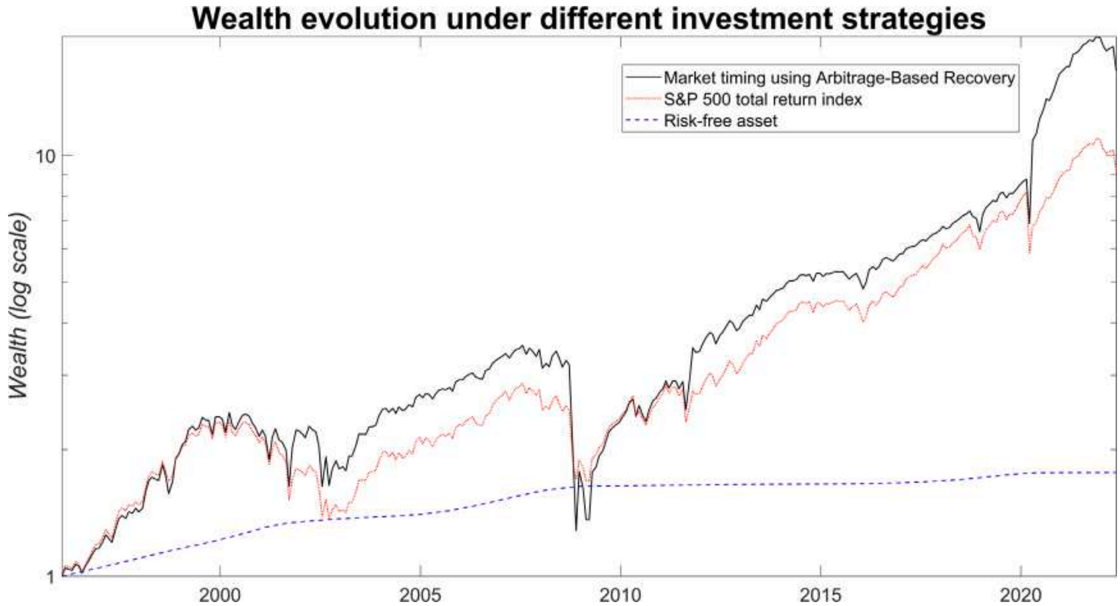

Fig. 9. Wealth evolution under three different investment strategies, starting with $1 on January 19, 1996. On the vertical axis, we show the wealth on a logarithmic scale. According to the first investment strategy (blue dashed curve), the total wealth is invested in the one-month risk-free asset and upon maturity it is always reinvested in the actual one-month risk-free asset. The second strategy (red dotted curve) corresponds to a buy-and-hold investment policy, where the total wealth is invested in the S&P 500 total return index on January 19, 1996, and it is kept there indefinitely. The third investment policy (black solid curve) is based on a market-timing strategy using Arbitrage-Based Recovery. Concretely, we invest in the one-month risk-free asset and in the S&P 500 total return index so that the portfolio weight of the S&P 500 total return index is proportional to the recovered Sharpe ratio of its one-month return. We choose the constant of proportionality such that the average (over the observation period) portfolio weight of the market index is equal to one. We re-evaluate the portfolio weights monthly (on each observation date), and re-balance our portfolio accordingly. The difference between the wealth paths of the market-timing strategy and the other two strategies can be attributed to the effect of market timing based on the Arbitrage-Based Recovery Theorem. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Hence, at the standard significance levels, we cannot reject the null hypothesis that the Arbitrage-Based Recovery Theorem recovers the true probability distributions.

7.4.2. Knüppel test

Next, we perform two versions of the test of Knüppel (2015). To this end, we transform our "realized" CDF values (denoted by x_t) according to

$$y_t = \sqrt{12} \times (x_t - 0.5)\,. \tag{33}$$

Under the null hypothesis, y_t follows a uniform distribution with support $\left[-\sqrt{3}, \sqrt{3}\right]$. We calculate the first three empirical raw moments of our observed y_t values, and form the 3×1 column vector $\boldsymbol{D}_3$ of differences between the empirical and theoretical moments. We also estimate the covariance matrix $\boldsymbol{\Omega}_3$ of $y_t - m_1$, $y_t^2 - m_2$, and $y_t^3 - m_3$, where m_i denotes the ith theoretical raw moment. Under the null hypothesis, the test statistic

$$\alpha_3 = T \times \boldsymbol{D}_3' \boldsymbol{\Omega}_3^{-1} \boldsymbol{D}_3 \tag{34}$$

asymptotically follows a χ^2 distribution with three degrees of freedom. Performing the test, we find that the p-value is 0.284. Hence, at the usual significance levels the Knüppel test with three moments does not reject the null hypothesis that our recovery approach indeed recovers the true probability distributions. Repeating the Knüppel test with the first four moments (instead of the first three), we find that the p-value is 0.080, which is just below the ten percent significance threshold.

7.4.3. Kolmogorov–Smirnov test

To perform the Kolmogorov–Smirnov test, we determine the maximum difference between the empirical cumulative distribution function implied by our "realized" physical CDF values and the theoretical cumulative distribution function of the uniform distribution with support $[0, 1]$. Under the null hypothesis, this maximum difference multiplied by the square root of the number of observations asymptotically follows the Kolmogorov distribution. Performing the test, we find that the p-value is 0.179. This, again, suggests that at the standard significance levels we cannot reject the null hypothesis that the Arbitrage-Based Recovery Theorem recovers the true probability distributions.

7.5. Mean prediction tests

Besides carrying out density tests, we also test whether our model-implied expected returns indeed predict the realized returns. Concretely, we run the regression

$$R_t = a + b\mu_t + \epsilon_t, \tag{35}$$

where R_t is the realized return at time t (not accounting for dividends), and μ_t is our model-implied expected return (again, without dividends). Under the null hypothesis, the intercept is $a = 0$ and the slope is $b = 1$. Assuming $b = 1$ and testing whether the intercept is statistically significantly different from zero, we find a p-value of 0.2920. Assuming a zero intercept and testing whether the slope is statistically significantly different from one, our p-value is 0.2915. Finally, without any restriction on our regression model, testing the joint hypothesis that $a = 0$ and $b = 1$, we find a p-value of 0.5593. Thus, neither of the above three mean prediction tests rejects the null hypothesis that the expected returns implied by the Arbitrage-Based Recovery Theorem are indeed the true expected returns.

Our results confirm the findings of Martin (2017), who argues empirically that his *negative correlation condition* is tight, i.e., that it is close to zero. Our recovered probabilities correspond to the case when Martin's correlation is exactly zero. And just like Martin, we cannot reject the hypothesis that the expected returns corresponding to Martin's zero correlation coincide with the true expected returns.

7.6. Empirical tests: summary and robustness checks

We summarize the results of our empirical tests in Table 5. For comparison purposes, we also report the results of Jackwerth and

Table 5
Empirical test p-values. From January 19, 1996 to May 19, 2022, we obtain the physical probability distributions of the 1-month return on the S&P 500 market index with monthly frequency by applying the arbitrage-based recovery theorem, as shown in Fig. 5. We also collect the realized returns on the S&P 500 market index one month after each observation date. We test the null hypothesis that the realized returns on the S&P 500 index are realizations of random variables with probability density functions implied by the arbitrage-based recovery theorem. We perform four different density tests (Berkowitz test, Knüppel test with three and four moments, Kolmogorov–Smirnov test) and three different mean prediction tests on the regression equation (35). For comparison purposes, we also report the results of Jackwerth and Menner (2020), who analyze alternative recovery theorems performing the same tests and using very similar data to ours (concretely, their sample period is from April 1, 1986 to December 31, 2017). Furthermore, we also perform the tests on the risk-neutral probabilities.

	Ross basic	Ross bounded	Ross unimodal	Ross stable	*Arbitrage-based recovery*	Risk-neutral probabilities
	Jackwerth and Menner (2020)				This paper	
Berkowitz test	0.001	0.000	0.000	0.002	***0.340***	0.070
Knüppel test (3 moments)	0.000	0.012	0.000	0.028	***0.284***	0.070
Knüppel test (4 moments)	0.000	0.000	0.000	0.002	***0.080***	0.020
Kolmogorov–Smirnov test	0.020	0.049	0.044	0.045	***0.179***	0.022
Intercept test	0.652	0.001	0.001	0.002	***0.292***	0.022
Slope test	0.000	0.055	0.022	0.219	***0.292***	0.023
Joint intercept and slope test	0.000	0.000	0.000	0.000	***0.559***	0.011

Menner (2020).[23] Furthermore, we also perform the tests on the risk-neutral probabilities. Apart from the Knüppel test with four moments (which has a p-value of 0.080), at the usual significance levels none of our empirical tests reject the hypothesis that realized returns are drawn from the probability distributions implied by the Arbitrage-Based Recovery Theorem. And although failing to reject the null hypothesis does not translate into accepting the alternative hypothesis, comparing the p-values of our recovery methodology to those of other recovery methods (and also to the p-values of the risk-neutral probabilities), our results are promising regarding the empirical plausibility and usefulness of the Arbitrage-Based Recovery Theorem.

Since our recovery approach infers the Perron–Frobenius eigenvalue and eigenvector of the Arrow–Debreu price matrix from a theoretical recursive no-arbitrage restriction instead of reconstructing the entire matrix and then solving an eigenvalue–eigenvector problem, most of the robustness concerns of other recovery approaches do not apply to Arbitrage-Based Recovery. For example, when implementing the original form of Ross's recovery theorem, one might be concerned about how the state space is formed: should moneyness or log-moneyness be used, should the grid be equidistant or not, how many states should there be, what should be the "span" of the grid (i.e., what should the boundary points of the grid be), etc. By contrast, Arbitrage-Based Recovery is unaffected by how the discretized state space is formed or what its boundaries are. Therefore, our robustness check focuses on whether potentially relevant information is lost by excluding very cheap options from our sample (as it is usually done in the literature), and on whether the fineness of our grid has any significant effect on the smooth state price density curve which might carry over to our results.

During our data cleaning procedure, we exclude options with lower than \$0.50 bid price. To check that this practice does not lead to losing significant information (especially about the tail probabilities, since we use only out-of-the-money options), we repeat our analysis by excluding only those options the bid price of which is lower than \$0.10 (instead of \$0.50). As we see in Table 6, this does not change our results significantly.

In our main analysis, we use a fine grid of $\Delta = \$0.1$. To ensure that the fineness of our grid does not have a significant effect on our state price density curve (and therefore on our empirical tests), we repeat our analysis using an even finer grid of $\Delta = \$0.033$ and a much coarser grid of $\Delta = \$5$. As we show in Table 7, our p-values do not change significantly.

[23] Jackwerth and Menner (2020) perform the same tests as we do, using very similar data. Using data from April 1, 1986 to December 31, 2017, they have 380 observations, while our data is from January 19, 1996 to May 19, 2022 and we have 317 observations.

7.7. Effects of problem regularization

We find (Table 5) that the probabilities implied by Arbitrage-Based Recovery seem to be substantially closer to the true probabilities than the probabilities implied by alternative recovery approaches. Since each of the listed recovery approaches attempt to extract the transitory SDF component in a framework with a closed and bounded price process as the only state variable, each approach should recover the same SDF as long as the model environment is specified correctly. Under the maintained assumptions, this recovered SDF should be the reciprocal of the gross return. It is not possible to extract a "flexible" form of the SDF (which is constrained only by the transition independence assumption), and any attempt to do so should be futile.

Confronting this consideration with the results in Table 5, it is very likely that the empirical implementation of the listed alternative recovery approaches suffers from some form of model misspecification. This is further confirmed by Jackwerth and Menner (2020), who in their Fig. 5 show the SDFs on February 17, 2010 implied by the different recovery approaches they study. While in absence of model misspecification these implied SDFs should each be equivalent to the reciprocal of the gross stock market return, in reality this is far from the case. The SDF implied by the *Ross basic* approach features an inverted U-shape, the SDF of the *Ross bounded* approach has a U-shape and then an inverted U-shape at high moneyness levels, the *Ross unimodal* approach features alternating U-shapes and inverted U-shapes along the moneyness level in the implied SDF, while the SDF of the *Ross stable* approach is mostly flat.

In the case of the *Ross basic*, *Ross bounded*, and *Ross unimodal* approaches, the entire Arrow–Debreu price matrix needs to be constructed and, in absence of arbitrage opportunities, they should satisfy the $\boldsymbol{A}^{Ð}\boldsymbol{S} = \boldsymbol{S}$ identity. Again, the shapes of the recovered SDFs in Jackwerth and Menner (2020) strongly suggest that this identity is violated. Possible reasons of this violation can be the effects of discretization (i.e., the price is allowed to take only discrete values), truncation (the price process is assumed to be bounded), or failure of the Markov property (i.e., the price process in reality is unlikely to be Markovian). In the case of the *Ross stable* approach, although the entire Arrow–Debreu price matrix does not need to be constructed per se, one row of the matrix needs to be constructed for several different maturities and as long as these rows are consistent with the assumed model environment (e.g., time homogeneity and a Markovian price process), the implied (one-period) SDF should be the reciprocal of the gross stock market return. Besides Jackwerth and Menner (2020), further discussion on the detrimental effects of the enforced regularization of the problem (e.g., discretization and truncation) on the recovery results can be found in Walden (2017) and Tran and Xia (2018).

Arbitrage-Based Recovery, unlike the other listed recovery approaches, does not require constructing the entire Arrow–Debreu price matrix (or one row of the matrix for different maturities) and it also circumvents many of the regularization requirements of the

Table 6
Empirical test p-values, excluding options with a bid price lower than \$0.10 or excluding options with a bid price lower than \$0.50. From January 19, 1996 to May 19, 2022, we obtain the physical probability distributions of the 1-month return on the S&P 500 market index with monthly frequency by applying the arbitrage-based recovery theorem, as explained in Section 7. We also collect the realized returns on the S&P 500 market index one month after each observation date. We test the null hypothesis that the realized returns on the S&P 500 index are realizations of random variables with probability density functions implied by the arbitrage-based recovery theorem. We perform four different density tests (Berkowitz test, Knüppel test with three and four moments, Kolmogorov–Smirnov test) and three different mean prediction tests on the regression equation (35). We also perform the tests on the risk-neutral probabilities.

	Arbitrage-based recovery		Risk-neutral probabilities	
	Excl. if P_b < \$0.10	Excl. if P_b < \$0.50	Excl. if P_b < \$0.10	Excl. if P_b < \$0.50
Berkowitz test	0.330	***0.340***	0.064	0.070
Knüppel test (3 moments)	0.249	***0.284***	0.060	0.070
Knüppel test (4 moments)	0.062	***0.080***	0.016	0.020
Kolmogorov–Smirnov test	0.117	***0.179***	0.018	0.022
Intercept test	0.283	***0.292***	0.022	0.023
Slope test	0.282	***0.292***	0.023	0.023
Joint intercept and slope test	0.543	***0.559***	0.011	0.011

Table 7
Empirical test p-values, using grids with different levels of coarseness. From January 19, 1996 to May 19, 2022, we obtain the physical probability distributions of the 1-month return on the S&P 500 market index with monthly frequency by applying the arbitrage-based recovery theorem, as explained in Section 7. We also collect the realized returns on the S&P 500 market index one month after each observation date. We test the null hypothesis that the realized returns on the S&P 500 index are realizations of random variables with probability density functions implied by the arbitrage-based recovery theorem. We perform four different density tests (Berkowitz test, Knüppel test with three and four moments, Kolmogorov–Smirnov test) and three different mean prediction tests on the regression equation (35). We also perform the tests on the risk-neutral probabilities.

	Arbitrage-based recovery			Risk-neutral probabilities		
	Δ = \$0.033	***Δ = \$0.1***	Δ = \$5	Δ = \$0.033	Δ = \$0.1	Δ = \$5
Berkowitz test	0.339	***0.340***	0.321	0.070	0.070	0.066
Knüppel test (3 moments)	0.284	***0.284***	0.261	0.069	0.070	0.064
Knüppel test (4 moments)	0.080	***0.080***	0.069	0.020	0.020	0.018
Kolmogorov–Smirnov test	0.177	***0.179***	0.178	0.022	0.022	0.022
Intercept test	0.292	***0.292***	0.293	0.022	0.022	0.022
Slope test	0.292	***0.292***	0.293	0.023	0.023	0.023
Joint intercept and slope test	0.559	***0.559***	0.561	0.011	0.011	0.011

other approaches. Instead, it directly infers the SDF from a theoretical recursive no-arbitrage restriction which the asset price must satisfy, and it requires only the one-period Arrow–Debreu prices (or state price densities) as an input, which can be obtained from observed one-period option prices readily along the lines of Breeden and Litzenberger (1978).

8. SDF decomposition and arbitrage-based recovery

The empirical literature (Alvarez and Jermann, 2005 and Bakshi et al., 2018, among others) documents that the permanent component of the stochastic discount factor is very volatile and it is considerably more important (from an asset pricing perspective) than the transitory component. At first sight, our results in Section 7 might seem to contradict these findings. In this section, we show that there is no contradiction and our empirical results can be reconciled with the well-documented importance of the permanent SDF component.

8.1. Empirical significance of the permanent SDF component

As Alvarez and Jermann (2005) show, the unconditional volatility of the permanent SDF component relative to the unconditional volatility of the SDF itself is at least as large as

$$\frac{\mathrm{E}\left[\log\left(\frac{R_{t+1}}{R_{t+1,\infty}}\right)\right]}{\mathrm{E}\left[\log\left(\frac{R_{t+1}}{R_{t+1,1}}\right)\right]+L\left(\frac{1}{R_{t+1,1}}\right)}, \tag{36}$$

where the unconditional volatility of a random variable x_{t+1} is defined as $L\left(x_{t+1}\right) \triangleq \log\left[\mathrm{E}\left(x_{t+1}\right)\right] - \mathrm{E}\left[\log\left(x_{t+1}\right)\right]$, R_{t+1} is the gross one-period return on any traded asset, $R_{t+1,\infty}$ is the gross one-period return on the long bond (as defined in Section 4), and $R_{t+1,1}$ is the gross one-period risk-free rate corresponding to the period from t to $t+1$. Now, we estimate this lower bound using our data. We observe the monthly gross returns (including dividends) on a value-weighted portfolio of all stocks listed on the NYSE, AMEX, or NASDAQ from 31 January, 1947 to 29 December, 2023, with monthly frequency. We also observe the one-month risk-free rates and the monthly realized returns on the long bond.[24] We find that the lower bound is 0.7352, which is very close to the values of 0.7673 and 0.7755 obtained by Alvarez and Jermann (2005) using data from December 1946 to December 1999. Hence, the permanent SDF component is very volatile and it plays a significant role in pricing assets.

8.2. True transitory SDF component vs. recovered SDF

Arbitrage-Based Recovery – along with the empirical implementation of other recovery approaches – assumes a framework with a stationary, closed and bounded price process. The inferred SDF is equivalent to the transitory SDF component of this featured financial market. However, importantly, if the assumed model is not correctly specified

[24] For the stock market returns, the sources of our data are the CRSP Stock File Indexes. Our risk-free rates are from the CRSP Treasuries Monthly Riskfree Series. For the returns on the long bond from January 1947 to December 1985, we use the yield curve data in McCulloch and Kwon (1993). To obtain the monthly long bond returns, at the end of each month we calculate the price of a zero-coupon bond with maturity in τ years, then one month later we calculate the price of a zero-coupon bond with maturity in $\tau-1$ years and 11 months (by linearly interpolating between the τ-year and the τ-minus-one-year yields). For each monthly return, we use the longest possible maturity (i.e., τ) such that the necessary yield curve data is available to calculate the return via the above procedure. From January 1986 to December 2023, we use the yield curve data provided by the Board of Governors of the Federal Reserve System obtained using the methodology of Gurkaynak et al. (2007). We calculate the monthly returns on the long bond using the same procedure as above, now using a maturity of $\tau = 30$ years.

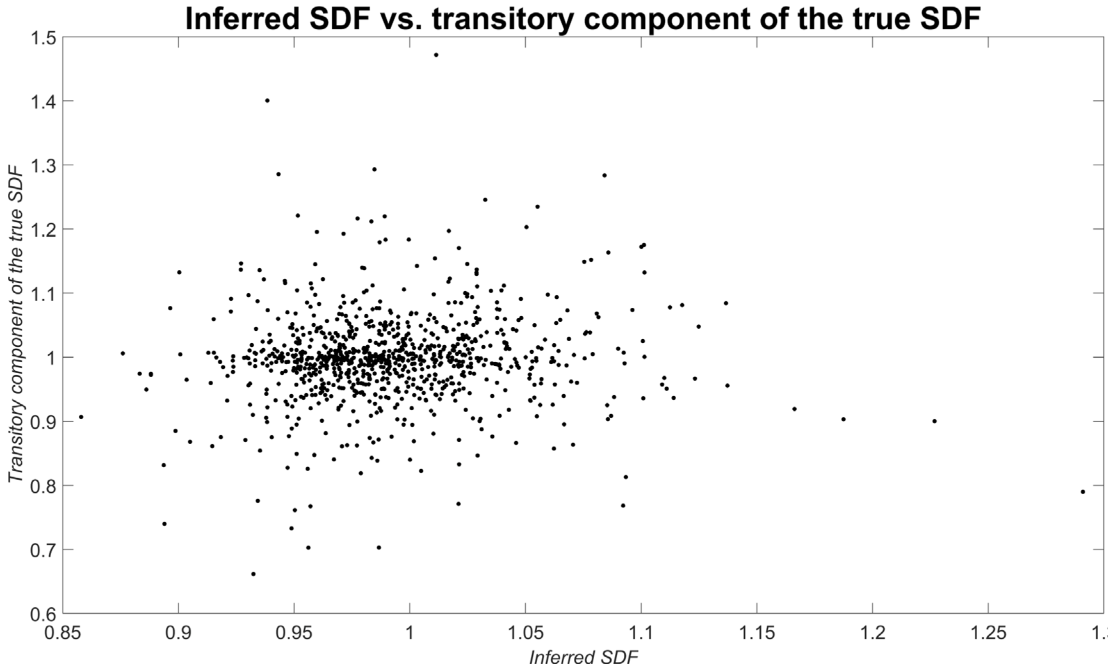

Fig. 10. Realized values of the inferred SDF and realized values of the transitory component of the true SDF. The SDFs are projected on the space spanned by the monthly non-annualized gross stock market index return. The inferred SDF is implied by Arbitrage-Based Recovery, and its realized values are equal to the reciprocals of the monthly non-annualized gross stock market index returns. The realized values of the transitory component of the true SDF are equal to the reciprocals of the monthly non-annualized gross returns on the long bond. The observation period is from January 1947 to December 2023, and the observation frequency is monthly.

(e.g., if the price process is not stationary in reality), then the inferred SDF might not be equivalent to the true transitory SDF component. The inferred SDF may partially absorb both the true transitory SDF component and the true permanent SDF component. Consequently, the inferred SDF might be very similar to the true SDF (and hence it might be hard to reject statistically using market data) even if the true SDF has a significant permanent component. This observation can help reconcile the empirical success (in the sense of Section 7) of our recovery approach with the empirical significance of the permanent SDF component (as we found in Section 8.1).

Example 1. For demonstration purposes, consider an economy with only two assets: a stock (paying no dividend) and a locally risk-free security. The stock price follows a geometric Brownian motion. The risk-free rate and the market price of risk are constant. In this economy, the SDF corresponding to a holding period τ is $\exp\left[-r_f\tau - \lambda^2\tau/2 - \lambda\left(W^{\mathbb{P}}_{t+\tau} - W^{\mathbb{P}}_t\right)\right]$, where r_f is the continuously-compounded risk-free rate, λ is the market price of risk, and $W^{\mathbb{P}}_t$ is a Wiener process under the physical probability measure $\mathbb{P}$. The transitory SDF component is equivalent to the risk-free discount factor (i.e., to $\exp\left(-r_f\tau\right)$), and this is also equivalent to the reciprocal of the return on the long bond (since the risk-free rate is constant). The permanent SDF component is equivalent to $\exp\left[-\lambda^2\tau/2 - \lambda\left(W^{\mathbb{P}}_{t+\tau} - W^{\mathbb{P}}_t\right)\right]$. The lower bound in (36) is equal to one, and the unconditional volatility of the SDF is equal to the unconditional volatility of the permanent SDF component. However, if we apply our Arbitrage-Based Recovery approach in this economy, our inferred SDF is the reciprocal of the gross stock return, i.e., $\exp\left[-\mu\tau + \sigma^2\tau/2 - \sigma\left(W^{\mathbb{P}}_{t+\tau} - W^{\mathbb{P}}_t\right)\right]$, where μ and σ are the drift and the volatility parameters of the stock return process, respectively. Clearly, our inferred SDF absorbs a part of the true permanent SDF component. For example, using reasonable parameter values of $\mu = 0.1$, $\sigma = 0.2$, and $\lambda = 0.4$, the SDF implied by Arbitrage-Based Recovery has the same geometric mean as the permanent component of the true SDF, while the former's geometric volatility is half of the latter's. Hence, in this economy, our inferred SDF can indeed be deemed to absorb a substantial part of the permanent component of the true SDF.

The standard assumption of recovery approaches that the price process is bounded, closed, and stationary, is not supported empirically. To show that this indeed causes the inferred SDF to be substantially different from the transitory component of the true SDF, we plot the realized values of the inferred SDF against the realized values of the transitory component of the true SDF (Fig. 10). The SDFs are projected on the space spanned by the monthly non-annualized gross stock market index return. The inferred SDF is implied by Arbitrage-Based Recovery, and its realized values are equal to the reciprocals of the monthly non-annualized gross stock market index returns. The realized values of the transitory component of the true SDF are equal to the reciprocals of the monthly non-annualized gross returns on the long bond. The observation period is from January 1947 to December 2023, and the observation frequency is monthly. Further details on our data and on how we construct the long bond returns can be found in Footnote 24.

If the inferred SDF were equivalent to the transitory component of the true SDF, the observations in Fig. 10 would all be on a line with unit slope. We find that in reality the case is vastly different. We also calculate the correlation between the inferred SDF and the transitory component of the true SDF, and we find that their correlation is only 0.0595. Based on the sample autocorrelation function, we find no evidence of autocorrelation in the differences between the realized reciprocals of the stock market index return and the long bond return, although we find some evidence of weak positive autocorrelation (with a coefficient around 0.2) in the absolute differences.

Since we find empirically that our inferred SDF (pricing the S&P 500 stock market index return) is very different from the transitory component of the true SDF, the fact that the permanent SDF component is very relevant for pricing assets (as we show in Section 8.1) does not contradict the empirical success of our Arbitrage-Based Recovery approach.

The fact that Arbitrage-Based Recovery implies an SDF which is not generally equivalent to the true transitory SDF component can also be argued at a more fundamental level. Consider the (true) SDF projected on the space spanned by any arbitrary return (i.e., not necessarily the stock market index return). Then, the transitory component of this SDF is the same, regardless of which return spans the space. To see this, note that the transitory SDF component is always equal to the reciprocal of the gross return on the long bond. And just like there is one unique one-period risk-free rate in the economy, there is one unique k-period risk-free rate in the economy for any arbitrary positive k integer as well. Hence, regardless of which return spans the space, the SDF processes corresponding to those spaces will imply the same k-period risk-free rate for any arbitrary positive k integer. This holds both at time t and at time $t+1$. Thus, observing a time series of realized transitory SDF components, we observe the same realizations, regardless of which return spans the space. And since the time series of realized returns on different stocks will obviously be different (and Arbitrage-Based Recovery always identifies the inferred SDF with the reciprocal of the spanning stock return) but the true transitory SDF components are the same, the inferred SDFs clearly cannot all be equivalent to the true transitory SDF component.

At this point it is worth referring back to Corollary 2, according to which the conditional probabilities implied by Arbitrage-Based Recovery are equivalent to the probabilities used by a pseudo-representative investor with log-utility over the terminal wealth, who fully invests her wealth in the market. Actually, the conditional probabilities implied by Arbitrage-Based Recovery could also be obtained in a standard representative agent framework, where the representative investor has logarithmic utility over terminal wealth, and the market constitutes all wealth. In such a model, the SDF is the reciprocal of the gross return on the stock market. The derivation of this SDF does not hinge upon the stock price being the only state variable, it does not assume a stationary environment, and it does not rely on our additional arbitrage condition. Importantly, it does not assume or imply that the permanent SDF component is constant, the framework is silent on this. Since the implied conditional probabilities of such a representative investor framework are the same as what our Arbitrage-Based Recovery implies, our results in Section 7 can also be interpreted as empirical tests of such a representative investor framework. This log-utility representative investor model is also consistent with the results of Martin (2017) in the sense that the implied expected stock market return is exactly the same as the lower bound of Martin (2017). Interpreting our results from the perspective of such a representative investor framework, there is no (seeming) tension between our empirical results and the literature's findings on the empirical relevance of the permanent SDF component.

9. Conclusion

We develop a novel recovery theorem based on no-arbitrage principles. Our Arbitrage-Based Recovery Theorem requires the observation of option prices only for one single maturity and only for the current state of the world. To test how well our theorem works empirically, we use more than 26 years of S&P 500 options data with monthly observation frequency, and perform several density tests and mean prediction tests. Using the same tests and very similar data to ours, the literature tends to reject the empirical validity of other, competing recovery theorems. In contrast with this, none of the commonly used density tests and mean prediction tests rejects the empirical validity of the Arbitrage-Based Recovery Theorem at the standard five percent significance level, and the vast majority of the tests cannot even reject it at a very conservative 15% significance level. After conducting our empirical exercise, we demonstrate that our results do not contradict the findings of Alvarez and Jermann (2005), among others, that the permanent component of the stochastic discount factor is empirically very volatile.

CRediT authorship contribution statement

Ferenc Horvath: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

In this appendix, we describe our procedure of attaching tails to the risk-neutral probability density functions in Section 7.

Attaching left tails

To attach left tails to the risk-neutral probability density functions below the lowest traded strike prices, we proceed as follows. Consider an observation date. First of all, note that between the lowest and the highest traded strike prices we have already obtained the risk-neutral pdf by calculating the second derivative of the call price with respect to the strike price, and then dividing this second derivative by risk-free discount factor. In a similar vein, we also obtain the risk-neutral cdf between the lowest and the highest traded strike prices by calculating the first derivative of the call price with respect to the strike price, dividing this first derivative by the risk-free discount factor, and adding one.

Let us denote the lowest traded strike price by S_l, and the strike price one grid point to the right from it by S_{l+1}. We calculate the risk-neutral pdf and cdf values at S_{l+1}. Then, we attach the "reflected" (in the sense explained below) right tail of a generalized extreme value (GEV) distribution pdf to the risk-neutral pdf at S_{l+1}, as the left tail of the risk-neutral pdf itself. Since the GEV distribution is uniquely characterized by three parameters, we choose the GEV distribution which satisfies the following three conditions:

- the GEV distribution pdf value at $-S_{l+1}$ is equal to the risk-neutral pdf value at S_{l+1};
- the GEV distribution cdf value at $-S_{l+1}$ is equal to one minus the risk-neutral cdf value at S_{l+1}; and
- the GEV distribution cdf value at $S = 0$ is equal to one.

Then, we reflect the obtained GEV distribution pdf on a vertical line at $-S_{l+1}$, and attach the reflected GEV pdf section below $-S_{l+1}$ to the previously obtained ("tailless") risk-neutral pdf at the strike price S_{l+1}.

Attaching right tails

After adding left tails to the risk-neutral pdfs, we now similarly attach right tails to them. Let us denote the highest traded strike price on a particular observation date by S_r, the strike price one grid point to its left by S_{r-1}, and the strike price one grid point to its right by S_{r+1}. We calculate the risk-neutral pdf and cdf values at S_{r-1}. Then, we attach the right tail of a GEV distribution pdf to the risk-neutral pdf at S_{r-1}, as the right tail of the risk-neutral pdf itself. Concretely, we choose the GEV distribution which satisfies the following three conditions:

- the GEV distribution pdf value at S_{r-1} is equal to the risk-neutral pdf value at S_{r-1};
- the GEV distribution cdf value at S_{r-1} is equal to the risk-neutral cdf value at S_{r-1}; and
- the GEV distribution cdf value at $2 \times S_{r+1}$ is equal to one.

Modifying the tails to enforce no-arbitrage

After adding left and right tails to the risk-neutral density functions, we still need to make sure that these density functions are arbitrage-free. In other words, we must check whether the price (implied by the risk-neutral density function) of the future stock cash flow (i.e., dividend plus post-dividend stock price) is equal to the current stock price. Since we did not enforce this no-arbitrage condition when attaching left and right tails, it is not necessarily the case that our risk-neutral pdf at this stage is free of arbitrage. Therefore, we slightly modify the attached left and right tails as follows.

First, we divide both tails into two equal sections. Then, on each of the four sections (two on the left tail and two on the right tail) we define a *modifier function* $g(S): \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ as

$$g(S) = \alpha \times \sin\left(\frac{S - S_a}{S_b - S_a} \times \pi\right) + 1, \tag{37}$$

where $\alpha \in \mathbb{R}$ is a scalar greater than minus one, and $S_a \in \mathbb{R}_{>0}$ and $S_b \in \mathbb{R}_{>0}$ are the lower and the upper end points of the section, respectively. The modifier parameters (denoted by α in (37)) of the four sections can be different. To obtain the new left and right tails of the risk-neutral density function, we multiply the original pdf values on the tails by the modifier function of the respective section. Finally, we choose the four modifier parameters so that neither the area under the left tail nor the area under the right tail changes, but the expected value implied by the new risk-neutral density function, multiplied by the risk-free discount factor, is exactly equal to the current stock price less the price of the dividend.[25] Throughout our sample, the magnitudes of the modifier parameters are small (close to zero), which suggests that the new ("modified" and arbitrage-free) risk-neutral density functions are very similar to their "non-modified" counterparts.

References

Alvarez, F., Jermann, U.J., 2005. Using asset prices to measure the persistence of the marginal utility of wealth. Econometrica 73 (6), 1977–2016.

Audrino, F., Huitema, R., Ludwig, M., 2019. An empirical implementation of the Ross recovery theorem as a prediction device. J. Financ. Econom. (nbz002).

Bachelier, L., 1900. Théorie de la spéculation. Ann. Sci. Ec. Norm. Supér. 17, 21–86.

Bakshi, G., Chabi-Yo, F., Gao, X., 2018. A recovery that we can trust? Deducing and testing the restrictions of the recovery theorem. Rev. Financ. Stud. 31 (2), 532–555.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. J. Finance 59 (4), 1481–1509.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. J. Bus. Econom. Statist. 19 (4), 465–474.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. J. Polit. Econ. 81 (3), 637–654.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Rev. Financ. Stud. 22 (11), 4463–4492.

Borovička, J., Hansen, L.P., Scheinkman, J.A., 2016. Misspecified recovery. J. Finance 71 (6), 2493–2544.

Borovička, J., Stachurski, J., 2021. Stability of equilibrium asset pricing models: A necessary and sufficient condition. J. Econom. Theory.

Breeden, D.T., Litzenberger, R.H., 1978. Prices of state-contingent claims implicit in option prices. J. Bus. 51 (4), 621–651.

Carr, P., Yu, J., 2012. Risk, return, and Ross recovery. J. Deriv. 20 (1), 38–59.

Chetty, R., Rosenberg, J., Saez, E., 2005. The Effects of Taxes on Market Responses to Dividend Announcements and Payments: What Can We Learn from The 2003 Dividend Tax Cut?. NBER Working Paper.

Cox, J.C., Ross, S.A., Rubinstein, M., 1979. Option pricing: A simplified approach. J. Financ. Econ. 7 (3), 229–263.

Dillschneider, Y., Maurer, R., 2019. Functional Ross recovery: Theoretical results and empirical tests. J. Econom. Dynam. Control 108.

Dubynskiy, S., Goldstein, R.S., 2013. Recovering Drifts and Preference Parameters from Financial Derivatives. Working Paper.

Ebert, S., 2013. Moment characterization of higher-order risk preferences. Theory and Decision 74, 267–284.

Figlewski, S., 2010. In: Bollerslev, T., Russell, J., Watson, M. (Eds.), Volatility and Time Series Econometrics: Essays in Honor of Robert Engle. Oxford Scholarship Online, chap. Estimating the Implied Risk Neutral Density for the U.S. Market Portfolio.

Ghosh, A., Roussellet, G., 2020. Identifying Beliefs from Asset Prices. Working Paper.

Gormsen, N.J., Jensen, C.S., 2020. Higher-Moment Risk. Working Paper.

Gurkaynak, R.S., Sack, B., Wright, J.H., 2007. The U.S. Treasury yield curve: 1961 to the present. J. Monetary Econ. 54 (8), 2291–2304.

Hansen, L.P., Scheinkman, J.A., 2009. Long-term risk: an operator approach. Econometrica 77 (1), 177–234.

Heston, S.L., 2021. Recovering the Variance Premium. Working Paper.

Huang, D., Shaliastovich, I., 2014. Risk Adjustment and the Temporal Resolution of Uncertainty: Evidence from Options Markets. Working Paper.

Jackwerth, J.C., 2004. Option-Implied Risk-Neutral Distributions and Risk Aversion. The Research Foundation of AIMR.

Jackwerth, J.C., Menner, M., 2020. Does the Ross recovery theorem work empirically? J. Financ. Econ. 137 (3), 723–739.

Jensen, C.S., 2021. The Ex Ante Physical Distributions of Individual Stock Returns. Working Paper.

Jensen, C.S., Lando, D., Pedersen, L.H., 2019. Generalized recovery. J. Financ. Econ. 133 (1), 154–174.

Kazemi, H.B., 1992. An Intemporal Model of Asset Prices in a Markov Economy with a Limiting Stationary Distribution. Rev. Financ. Stud. 5 (1), 85–104.

Knüppel, M., 2015. Evaluating the calibration of multi-step-ahead density forecasts using raw moments. J. Bus. Econom. Statist. 33 (2), 270–281.

Liu, F., 2014. Recovering Conditional Return Distributions by Regression: Estimation and Applications. Working Paper.

Markowitz, H., 1952. Portfolio selection. J. Finance 7 (1), 77–91.

Martin, I., 2017. What is the expected return on the market? Q. J. Econ. 132 (1), 367–433.

Martin, I.W.R., Ross, S., 2019. Notes on the yield curve. J. Financ. Econ. 134 (3), 689–702.

Massacci, F., Williams, J.M., Zhang, Y., 2016. Hansen-Scheinkman Factorization and Ross Recovery from Option Panels. Working Paper.

McCulloch, J., Kwon, H., 1993. U.S. Term Structure Data, 1947–1991. Ohio State University Working Paper 93-6.

Mehra, R., Prescott, E.C., 1985. The equity premium: A puzzle. J. Monetary Econ. 15 (2), 145–161.

Merton, R.C., 1973. Theory of rational option pricing. Bell J. Econ. Manag. Sci. 4 (1), 141–183.

Park, H., 2016. Ross recovery with recurrent and transient processes. Quant. Finance 16 (5), 667–676.

Pazarbasi, A., Schneider, P., Vilkov, G., 2021. Dispersion of Beliefs Bounds: Sentimental Recovery. Swiss Finance Institute Research Paper Series 19–57.

Qin, L., Linetsky, V., 2016. Positive Eigenfunctions of Markovian Pricing Operators: Hansen-Scheinkman Factorization, Ross Recovery, and Long-Term Pricing. Oper. Res. 64 (1), 99–117.

Qin, L., Linetsky, V., 2017. Long term risk: A martingale approach. Econometrica 85 (1), 299–312.

Qin, L., Linetsky, V., Nie, Y., 2018. Long forward probabilities, recovery, and the term structure of bond risk premiums. Rev. Financ. Stud. 31 (12), 4863–4883.

Ross, S.A., 1976. Options and efficiency. Q. J. Econ. 90 (1), 75–89.

Ross, S., 2015. The recovery theorem. J. Finance 70 (2), 615–648.

Schneider, P., Trojani, F., 2019. (Almost) model-free recovery. J. Finance 74 (1), 323–370.

Tran, N.-K., Xia, S., 2018. Specified Recovery. Working Paper.

Walden, J., 2017. Recovery with unbounded diffusion processes. Rev. Finance 21 (4), 1403–1444.

Zhan, X., 2013. Matrix Theory. American Mathematical Society.

Zhu, S., 2020. Why Does the Recovery Theorem Not Work Empirically? A Martingale-Based Explanation. Working Paper.

[25] In our empirical analysis in Section 7, we assume that the dividend to be received until the maturity of the option (i.e., in the next 28 days) is known at the time of observation. As noted by Chetty et al. (2005), firms usually announce their dividends four to six weeks before the dividend payment itself, so this assumption on our side is reasonable.