Generalized Recovery

Christian Skov Jensen, David Lando, and Lasse Heje Pedersen*

First version March, 2015. This version May 24, 2017

Abstract

We characterize when physical probabilities, marginal utilities, and the discount rate can be recovered from observed state prices for several future time periods. We make no assumptions of the probability distribution, thus generalizing the time-homogeneous stationary model of Ross (2015). Recovery is feasible when the number of maturities with observable prices is higher than the number of states of the economy (or the number of parameters characterizing the pricing kernel). When recovery is feasible, our model allows a closed-form linearized solution. We implement our model empirically, testing the predictive power of the recovered expected return and other recovered statistics.

*Jensen is at Copenhagen Business School; Lando is at Copenhagen Business School and CEPR; Pedersen is at Copenhagen Business School, NYU, AQR Capital Management, and CEPR; www.lhpedersen.com. We are grateful for helpful comments from Jaroslav Borovicka, Peter Christoffersen, Horatio Cuesdeanu, Darrell Duffie, Lars Peter Hansen, Jens Jackwerth, Pia Mølgaard, Stephen Ross, Jose Scheinkman, Paul Schneider, Paul Whelan as well as from seminar participants at University of Konstanz, University of Toronto, HEC-McGill winter finance workshop 2015, ATP, SED summer conference 2016, EFA 2016, AFA 2017, and Copenhagen Business School, and especially to Dan Petersen for guiding us through Sard's theorem. All three authors gratefully acknowledge support from the FRIC Center for Financial Frictions (grant no. DNRF102) and Jensen and Pedersen from the European Research Council (ERC grant no. 312417). Lando is a board member in the Danish FSA and in mutual funds under the administration of Nykredit. **This footnote contains all relevant disclosures.**

1 Introduction

The holy grail in financial economics is to decode probabilities and risk preferences from asset prices. This decoding has been viewed as impossible until Ross (2015) provided sufficient conditions for such a recovery in a time-homogeneous Markov economy (using the Perron-Frobenius Theorem). However, his recovery method has been criticized by Borovicka, Hansen, and Scheinkman (2016) (who also rely on Perron-Frobenius and results of Hansen and Scheinkman (2009)), arguing that Ross's assumptions rule out realistic models.

This paper sheds new light on this debate, both theoretically and empirically. Theoretically, we generalize the recovery theorem to handle a general probability distribution which makes no assumptions of time-homogeneity or Markovian behavior. We show when recovery is possible – and when it isn't – using a simple "counting" argument (formalized based on Sard's Theorem), which focuses the attention on the economics of the problem. When recovery is possible, we show that our recovery inversion from prices to probabilities and preferences can be implemented in closed form. We implement our method empirically using option data from 1996-2015 and study how the recovered expected returns predict future actual returns.

To understand our method, note first that Ross (2015) assumes that state prices are known not just in each final state, but also starting from each possible current state as illustrated in Figure 1, Panel A. Simply put, he assumes that we know all prices today and all prices in all "parallel universes" with different starting points. Since we clearly cannot observe such parallel universes, Ross (2015) proposes to implement his model based on prices for several future time periods, relying on the assumption that all time periods have identical structures for prices and probabilities (time-homogeneity), illustrated in Figure 1, Panel B. In other words, Ross assumes that, if S&P 500 is at the level 2000, then one-period option prices do not depend on the calendar time at which this level is observed.

We show that the recovery problem can be simplified by starting directly with the state prices for all future times given only the *current* state (Figure 1, Panel C). We impose no dynamic structure on the probabilities, allowing the probability distribution to be fully general at each future time, thus relaxing Ross's time-homogeneity assumption which is unlikely to be met empirically.

We first show that when the number of states S is no greater than the number of time periods T, then recovery is possible. To see the intuition, consider simply the number of equations and the number of unknowns: First, we have S equations at each time period, one for each Arrow-Debreu price, for a total of ST equations. Second, we have 1 unknown discount rate, $S-1$ unknown marginal utilities, and $S-1$ unknown probabilities for each future time period. In conclusion, we have ST equations with $1+(S-1)+(S-1)T = ST+S-T$ unknowns. These equations are not linear, but we provide a precise sense in which we can essentially just count equations. Hence, recovery is possible when $S \leq T$.

To understand the intuition behind this result, note that, for each time period, we have S equations and only $S-1$ probabilities. Hence, we have one extra equation that can help us recover the marginal utilities and discount rate — and the number of marginal utilities does not grow with the number of time periods.

By focusing on square matrices, Ross's model falls into the category $S=T$ so our counting argument explains why he finds recovery. However, our method applies under much more general conditions. We show that, when Ross's time-homogeneity conditions are met, then our solution is the same as his. The converse is not true: when Ross's conditions are not met, then our model can be solved while Ross's cannot. Further, we illustrate that our solution is far simpler and allows a closed-form solution that is accurate when the discount rate is close to 1.

To understand the economics of the condition $S \leq T$, consider what happens if the economy evolves in a standard multinomial tree with no upper or lower bound on the state space: For each extra time period, we get at least two new states since we can go up from highest state and down from the lowest state. Therefore, in this case $S > T$, so we see that recovery is impossible because of the number of states is higher than the number of time periods. Hence, achieving recovery without further assumptions is typically impossible in most standard models of finance where the state space grows in this way. In other words, our model provides a simple alternative way – via our counting argument – to understand the critique of Borovicka, Hansen, and Scheinkman (2016) that recovery is impossible in standard models.

Nevertheless, we show that recovery is possible even when $S > T$ under certain conditions. While maintaining that probabilities can be fully general (and, indeed, allow growth), we assume that the utility function is given via a limited number of parameters. Again, we simply need to make our counting argument work. To do this, we show that, if the marginal utilities can be written as functions of N parameters, then recovery is possible as long as $N + 1 < T$. This large state-space framework is what we use empirically as discussed further below.

We illustrate how our method works in the context of three specific models, namely Mehra and Prescott (1985), Black and Scholes (1973), and a simple non-Markovian economy. For each economy, we generate model-implied prices and seek to recover natural probabilities and preferences using our method. This provides an illustration of how our method works, its robustness, and its shortcomings. For Mehra and Prescott (1985), we show that $S > T$ so general recovery is impossible, but, when we restrict the class of utility functions, then we achieve recovery. For the binomial model in the spirit of Black and Scholes (1973), we show that recovery is impossible even under restrictive utility specifications because consumption growth is iid., which leads to a flat term structure, a pricing matrix of a lower rank, and a continuum of solutions for probabilities and preferences. While the former two models fall in the setting of Borovicka, Hansen, and Scheinkman (2016) (with a non-zero martingale component), we also show how recovery is possible in the non-Markovian setting, which falls outside the framework of Borovicka, Hansen, and Scheinkman (2016) and Ross (2015), illustrating the generality of our framework in terms of the allowed probabilities.

Finally, we implement our methodology empirically using a large data set of call and put options written on the S&P 500 stock market index over the time period

1996-2015. We estimate state price densities for multiple future horizons and recover probabilities and preferences each month. Based on the recovered probabilities, we derive the risk and expected return over the future month from the physical distribution of returns using four different methods. The recovered expected returns vary substantially across specifications, challenging the empirical robustness of the results. The recovered expected returns have weak predictive power for the future realized returns, but the predictability is stronger when we exclude the global financial crisis. We can also recover ex ante volatilities, which have much stronger predictive power for future realized volatility.

The literature on recovery theorems is quickly expanding. Bakshi, Chabi-Yo, and Gao (2015) and Audrino, Huitema, and Ludwig (2014) empirically test the restrictions of Ross's Recovery Theorem. Martin and Ross (2013) apply the recovery theorem in a term structure model in which the driving state variable is a stationary Markov chain and they show how recovery can be done using the (infinitely) long end of the yield curve. Several papers focus on generalizing the underlying Markov process to a continuous-time process with a continuum of values (Carr and Yu (2012), Linetsky and Qin (2016), Walden (2017)). All these papers impose time-homogeneity of the underlying Markov process.[1] Qin and Linetsky (2017) go beyond the Markov assumption, discussing factorization of stochastic discount factors and recovery in a general semimartingale setting. Their factorization requires an infinite time-horizon because it relies on limits of T-forward measures as T goes to infinity.

Prior to Ross (2015), the dynamics of the risk-neutral density and the physical density along with the pricing kernel has been extensively researched using historical option or equity market data (e.g., Jackwerth (2000), Jackwerth and Rubinstein (1996), Bollerslev and Todorov (2011), Ait-Sahalia and Lo (2000), Rosenberg and Engle (2002), Bliss and Panigirtzoglou (2004) and Christoffersen, Heston, and Jacobs

[1]See also Schneider and Trojani (2015) who focus on recovering moments of the physical distribution and Malamud (2016) who shows that knowledge of investor preferences is not necessarily enough to recover physical probabilities when option supply is noisy, but shows how recovery can may be feasible when the volatility of option supply shocks is also known.

(2013)).

Our paper contributes to the literature by characterizing recovery of any probability distributions, by proving a simple solution and its closed-form approximation, and by providing natural empirical tests of our generalized method. Rather than relying on specific probabilistic assumptions (Markov processes and ergodocity) as in Ross (2015) and Borovicka, Hansen, and Scheinkman (2016), we follow the tradition of general equilibrium (GE) theory, where Debreu (1970) pioneered the use of Sard's theorem and differential topology. Bringing Sard's theorem into the recovery debate provides new economic insight on when recovery is possible.[2] Indeed, the martingale decomposition applied by Borovicka, Hansen, and Scheinkman (2016) relies on knowing the infinite-time distribution of Markov processes, which imposes much more structure than needed and removes the focus from the essence of the recovery problem, namely the number of economic variables vs. economic restrictions.[3]

2 Ross's Recovery Theorem

This section briefly describes the mechanics of the recovery theorem of Ross (2015) as a background for understanding our generalized result in which we relax the assumption that transition probabilities are time-homogeneous.

The idea of the recovery theorem is most easily understood in a one-period setting. In each time period 0 and 1, the economy can be in a finite number of states which we label $1, \ldots, S$. Starting in any state i, there exists a full set of Arrow-Debreu securities, each of which pays 1 if the economy is in state j at date 1. The price of these securities is given by $\pi^{i,j}$.

The objective of the recovery theorem is to use information about these observed state prices to infer physical probabilities $p^{i,j}$ of transitioning from state i to j. We

[2]We thank Steve Ross for pointing out the historical role of Sard's theorem in general equilibrium theory.

[3]Said differently, if we observe a data from a finite number of time periods from an economy satisfying the conditions on Borovicka, Hansen, and Scheinkman (2016), then there is no unique Markov decomposition.

can express the connection between Arrow-Debreu prices and physical probabilities by introducing a pricing kernel m such that for any $i, j = 1, ..., S$

$$\pi^{i,j} = p^{i,j} m^{i,j} \tag{1}$$

It takes no more than a simple one-period binomial model to convince oneself, that if we know the Arrow-Debreu prices in one and only one state at date 0, then there is in general no hope of recovering physical probabilities. In short, we cannot separate the contribution to the observed Arrow-Debreu prices from the physical probabilities and the pricing kernel.

The key insight of the recovery theorem is that by assuming that we know the Arrow-Debreu prices for *all* the possible starting states, then with additional structure on the pricing kernel, we can recover physical probabilities. We note that knowing the prices in states we are not currently in ("parallel universes") is a strong assumption.

In any event, under this assumption, Ross's result is that there exists a unique set of physical probabilities $p^{i,j}$ for all $i, j = 1, \ldots, S$ such that (1) holds if the matrix of Arrow-Debreu prices is irreducible and if the pricing kernel m has the form known from the standard representative agent models:

$$m^{i,j} = \delta \frac{u^j}{u^i} \tag{2}$$

where $\delta > 0$ is the discount rate and $u = (u^1, \ldots, u^S)$ is a vector with strictly positive elements representing marginal utilities.

The proof can be found in Ross (2015), but here we note that counting equations and unknowns certainly makes it plausible that the theorem is true: There are S^2 observed Arrow-Debreu prices and hence S^2 equations. Because probabilities from a fixed starting state sum to one, there are $S(S-1)$ physical probabilities. It is clear that scaling the vector u by a constant does not change the equations, and thus we can assume that $u^1 = 1$ so that u contributes with an additional $S - 1$ unknowns. Adding to this the unknown δ leaves us exactly with a total of S^2 unknowns. The fact that there is a unique strictly positive solution hinges on the Frobenius theorem for positive matrices.

It is important in Ross's setting as it will be in ours, that a state corresponds to a particular level of the marginal utility of consumption. This level does not depend on calendar time. In our empirical implementation, a state will correspond to a particular level of the S&P500 index.

The most troubling assumption, however, in the theorem above is that we must know state prices also from starting states that we are currently not in. It is hard to imagine data that would allow us to know these in practice. Ross's way around this assumption is to leave the one-period setting and assume that we have information about Arrow-Debreu prices from several future periods and then use a time-homogeneity assumption to recover the same information that we would be able to obtain from the equations above.

We therefore consider a discrete-time economy with time indexed by t, states indexed by $s = 1, ..., S$, and $\pi^{i,j}_{t,t+\tau}$ denoting the time-t price in state i of an Arrow-Debreu security that pays 1 in state j at date $t + \tau$. The multi-period analogue of Eqn. (1) becomes

$$\pi^{i,j}_{t,t+\tau} = p^{i,j}_{t,t+\tau}\, m^{i,j}_{t,t+\tau} \tag{3}$$

Similarly, the multi-period analogue to equation (2) is the following assumption, which again follows from the existence of a representative agent with time-separable utility:

Assumption 1 (Time-separable utility) *There exists a $\delta \in (0, 1]$ and marginal utilities $u^j > 0$ for each state j such that, for all times τ, the pricing kernel can be written as*

$$m^{i,j}_{t,t+\tau} = \delta^\tau \frac{u^j}{u^i} \tag{4}$$

Critically, to move to a multi-period setting, Ross makes the following additional assumption of time-homogeneity in order to implement his approach empirically:

Assumption 2 (Time-homogeneous probabilities) *For all states i, j and time horizons $\tau > 0$, $p^{i,j}_{t,t+\tau}$ does not depend on t.*

This assumption is strong and not likely to be satisfied empirically. We note that Assumptions 1 and 2 together imply that risk neutral probabilities are also time-homogeneous, a prediction that can also be rejected in the data.

In this paper, we dispense with the time-homogeneity Assumption 2. We start by maintaining Assumption 1, but later consider a broader assumption that can be used in a large state space.

3 A Generalized Recovery Theorem

The assumption of time-separable utility is consistent with many standard models of asset pricing, but the assumption of time-homogeneity is much more troubling. It restricts us from working with a growing state space (as in standard binomial models) and it makes numerical implementation extremely hard and non-robust, because trying to fit observed state prices to a time-homogeneous model is extremely difficult. Furthermore, the main goal of the recovery exercise is to recover physical transition probabilities from the current states to all future states over different time horizons. Insisting that these transition probabilities arise from constant one-period transition probabilities is a strong restriction. We show in this section that by relaxing the assumption of time-homogeneity of physical transition probabilities, we can obtain a problem which is easier to solve numerically and which allows for a much richer modeling structure. We show that our extension contains the time-homogeneous case as a special case, and therefore ultimately should allow us to test whether the time-homogeneity assumption can be defended empirically.

3.1 A Noah's Arc Example: Two States and Two Dates

To get the intuition of our approach, we start by considering the simplest possible case with two states and two time-periods. Consider the simple case in which the economy has two possible states $(1,2)$ and two time periods starting at time t and ending on dates $t+1$ and $t+2$. If the current state of the world is state 1, then equation (3) consists of four equations:

$$\begin{aligned} \pi_{t,t+1}^{1,1} &= & p_{t,t+1}^{1,1} & \quad m_{t,t+1}^{1,1} \\ \pi_{t,t+1}^{1,2} &= & (1-p_{t,t+1}^{1,1}) & \quad m_{t,t+1}^{1,2} \\ \pi_{t,t+2}^{1,1} &= & p_{t,t+2}^{1,1} & \quad m_{t,t+2}^{1,1} \\ \pi_{t,t+2}^{1,2} &= & \underbrace{(1-p_{t,t+2}^{1,1})}_{\text{2 unknowns}} & \quad \underbrace{m_{t,t+2}^{1,2}}_{\text{4 unknowns}} \end{aligned} \tag{5}$$

We see that we have 4 equations with 6 unknowns so this system cannot be solved in full generality. However, the number of unknowns is reduced under the assumption of time-separable utility (Assumption 1). To see that most simply, we introduce the notation h for the normalized vector of of marginal utilities:

$$h = \left(1, \frac{u^2}{u^1}, \ldots, \frac{u^S}{u^1}\right)' \equiv (1, h_2, \ldots, h_S)'. \tag{6}$$

where we normalize by u^1. With this notation and the assumption of time-separable utility, we can rewrite the system (5) as follows:

$$\begin{aligned} \pi_{t,t+1}^{1,1} &= p_{t,t+1}^{1,1}\delta \\ \pi_{t,t+1}^{1,2} &= (1-p_{t,t+1}^{1,1})\delta h_2 \\ \pi_{t,t+2}^{1,1} &= p_{t,t+2}^{1,1}\delta^2 \\ \pi_{t,t+2}^{1,2} &= (1-p_{t,t+2}^{1,1})\delta^2 h_2 \end{aligned} \tag{7}$$

This system now has 4 equations with 4 unknowns, so there is hope to recover the physical probabilities p, the discount rate δ, and the ratio of marginal utilities h.

Before we proceed to the general case, it is useful to see how the problem is solved in this case. Moving h_2 to the left side and adding the first two and the last two equations gives us two new equation

$$\pi^{1,1}_{t,t+1} + \pi^{1,2}_{t,t+1}\frac{1}{h_2} - \delta = 0 \tag{8}$$

$$\pi^{1,1}_{t,t+2} + \pi^{1,2}_{t,t+2}\frac{1}{h_2} - \delta^2 = 0$$

Solving equation (8) for h_2 yields $\frac{1}{h_2} = (\delta - \pi^{1,1}_{t,t+1})/\pi^{1,2}_{t,t+1}$ and we can further arrive at

$$\pi^{1,1}_{t,t+2} - \frac{\pi^{1,2}_{t,t+2}\pi^{1,1}_{t,t+1}}{\pi^{1,2}_{t,t+1}} + \frac{\pi^{1,2}_{t,t+2}}{\pi^{1,2}_{t,t+1}}\delta - \delta^2 = 0 \tag{9}$$

Hence, we can solve the 2-state model by (i) finding δ as a root of the 2nd degree polynomial (9); (ii) computing the marginal utility ratio h_2 from (8); and (iii) computing the physical probabilities by rearranging (7).

3.2 General Case: Notation

Turning to the general case, recall that there are S states and T time periods. Without loss of generality, we assume that the economy starts at date 0 in state 1. This allows us to introduce some simplifying notation since we do not need to keep track of the starting time or the starting state — we only need to indicate the final state and the time horizon over which we are considering a specific transition.

Accordingly, let $\pi_{\tau s}$ denote the price of receiving 1 at date τ if the realized state is s and collect the set of observed state prices in a $T \times S$ matrix Π defined as

$$\Pi = \begin{bmatrix} \pi_{11} & \dots & \pi_{1S} \\ \vdots & & \vdots \\ \pi_{T1} & \dots & \pi_{TS} \end{bmatrix} \tag{10}$$

Similarly, letting $p_{\tau s}$ denote the physical transition probabilities of going from the current state 1 to state s in τ periods, we define a $T \times S$ matrix P of physical probabilities. Note that $p_{\tau s}$ is *not* the probability of going from state τ to s (as in the setting of Ross (2015)), but, rather, the first index denotes time for the purpose of the derivation of our theorem.

From the vector of normalized marginal utilities h defined as in (6) we define the $S-$dimensional diagonal matrix $H = \text{diag}(h)$. Further, we construct a $T-$dimensional diagonal matrix of discount factors as $D = \text{diag}(\delta, \delta^2, \ldots, \delta^T)$.

3.3 Generalized Recovery

With this notation in place, the fundamental TS equations linking state prices and physical probabilities, assuming utilities depend on current state only, can be expressed in matrix form as

$$\Pi = DPH \tag{11}$$

Note that the (invertible) diagonal matrices H and D depend only on the vector h and the constant δ so, if we can determine these, we can find the matrix of physical transition probabilities from the observed state prices in Π:

$$P = D^{-1}\Pi H^{-1} \tag{12}$$

Since probabilities add up to 1, we can write $Pe = e$, where $e = (1, \ldots, 1)'$ is a vector of ones. Using this identity, we can simplify (12) such that it only depends on δ and h:

$$\Pi H^{-1}e = DPe = De = (\delta, \delta^2, \ldots, \delta^T)' \tag{13}$$

To further manipulate this equation it will be convenient to work with a division of Π into block matrices:

$$\Pi = \begin{bmatrix} \Pi_1 & \Pi_2 \end{bmatrix} = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} \tag{14}$$

Here, Π_1 is a column vector of dimension T, where the first $S-1$ elements are denoted by Π_{11} and the rest of the vector is denoted Π_{21}. Similarly, Π_2 is a $T\times(S-1)$ matrix, where the first $S-1$ rows are called Π_{12} and the last rows are called Π_{22}. With this notation and the fact that $H(1,1) = h(1) = 1$, we can write (13) as

$$\Pi_1 + \Pi_2 \begin{bmatrix} h_2^{-1} \\ \vdots \\ h_S^{-1} \end{bmatrix} = \begin{bmatrix} \delta \\ \vdots \\ \delta^T \end{bmatrix} \tag{15}$$

where of course $h_s^{-1} = \frac{1}{h_s}$. Given that these equations are linear in the inverse marginal utilities h_s^{-1}, it is tempting to solve for these. To solve for these $S-1$ marginal utilities, we consider the first $S-1$ equations

$$\Pi_{11} + \Pi_{12} \begin{bmatrix} h_2^{-1} \\ \vdots \\ h_S^{-1} \end{bmatrix} = \begin{bmatrix} \delta \\ \vdots \\ \delta^{S-1} \end{bmatrix} \tag{16}$$

with solution[4]

$$\begin{bmatrix} h_2^{-1} \\ \vdots \\ h_S^{-1} \end{bmatrix} = \Pi_{12}^{-1} \left(\begin{bmatrix} \delta \\ \vdots \\ \delta^{S-1} \end{bmatrix} - \begin{bmatrix} \pi_{11} \\ \vdots \\ \pi_{S-1,1} \end{bmatrix} \right) \tag{17}$$

Hence, if δ were known, we would be done. Since δ is a discount rate, it is reasonable to assume that it is close to one over short time periods. We later use this insight to derive a closed-form approximation which is accurate as long as we have a reasonable sense of the size of δ. For now, we proceed for general unknown δ.

We thus have the utility ratios given as a linear function of powers of δ. The

[4]Of course, to invert Π_{12} it must have full rank. As long as Π_2 has full rank, we can re-order the rows to ensure that Π_{12} also has full rank.

remaining $T - S + 1$ equations give us

$$\Pi_{21} + \Pi_{22} \begin{bmatrix} h_2^{-1} \\ \vdots \\ h_S^{-1} \end{bmatrix} = \begin{bmatrix} \delta^S \\ \vdots \\ \delta^T \end{bmatrix} \tag{18}$$

and from this we see that if we plug in the expression for the utility ratios found above, we end up with $T - S + 1$ equations, each of which involves a polynomium in δ of degree a most T. If $T = S$, then δ is a root to a single polynomium so at most a finite number of solutions exist. If $T > S$, then typically no solution exists for general Arrow-Debreu prices Π since δ must simultaneously solve several polynomial equations. However, if the prices are generated by the model, then a solution exists and it will almost surely be unique. To be precise, we say that Π has been "generated by the model" if there exist δ, P, and H such that Π can be found from the right-hand side of (11). The following theorem formalizes these insights (using Sard's Theorem):

Proposition 1 (Generalized Recovery) *Consider an economy satisfying Assumption 1 with Arrow-Debreu prices for each of the T time periods and S states. The recovery problem has*

1. *a continuum of solutions if $S > T$;*
2. *at most S solutions if the submatrix Π_2 has full rank and $S = T$;*
3. *no solution generically in terms of an arbitrary positive matrix Π and $S < T$;*
4. *a unique solution generically if Π has been generated by the model and $S < T$.*

Proof. We have already provided a proof for 1 and 2 in the body of the text. Turning to 3, we note that the set X of all (δ, h, P) is a manifold-with-boundary of dimension $S \cdot T - T + S$. The discount rate, probabilities and marginal utilities map into prices, which we denote by $F(\delta, h, P) = DPH = \Pi$, where, as before, $D = \text{diag}(\delta, ..., \delta^T)$ and $H = \text{diag}(1, h_2, ..., h_S))$, and F is C^∞. If $S < T$, the image $F(X)$ has Lebesgue measure zero in $\mathbb{R}^{T\times S}$ by Sard's theorem, proving 3. Indeed, this means that the prices that are generated by the model $F(X)$ have measure zero relative to all prices Π.

Turning to 4, we first note that P and H can be uniquely recovered from (δ, Π) (given that Π is generically full rank). Indeed, H is recovered from (17) and P is recovered from (12). Therefore, we can focus on (δ, Π).

For two different choices of the discount rate (δ_a, δ_b) and a single set of prices Π, we consider the triplet $(\delta_a, \delta_b, \Pi)$. We are interested in showing that the different discount rates cannot both be consistent with the same prices, generically. To show this, we consider the space M where the reverse is true, hoping to show that M is "small." Specifically, M is the set of triplets where Π is of full rank and both discount rates are consistent with the prices, that is, there exists (unique) P_i and H_i $(i = a, b)$ such that $D_a P_a H_a = D_b P_b H_b = \Pi$.

Given that probabilities and marginal utilities can be uniquely recovered from prices and a discount rate (as explained above), we have a smooth map G from M to X by mapping any triplet $(\delta_a, \delta_b, \Pi)$ to (δ_a, h_a, P_a), where (h_a, P_a) are the recovered marginal utility and probabilities. The image of this map consists exactly of those elements of X for which F is not injective. The proof is complete if we can show that this image has Lebesgue measure zero, which follows again by Sard's theorem if we can show that the dimension of M is strictly smaller than $ST - T + S$.

To study the dimension of M, we note that we can think of M as the space of triplets such that the span of Π contains both the points $(\delta_a, \delta_a^2, ..., \delta_a^T)'$ and $(\delta_b, \delta_b^2, ..., \delta_b^T)'$. The span of Π is given by $V_\Pi := \{\Pi \cdot (1, h_2, h_3, ..., h_S)' | h_s > 0\}$, which is an affine $(S-1)$-dimensional subspace of $\mathbb{R}^T$ for Π of full rank. The set of all those $\Pi \in \mathbb{R}^{T\times S}$ such that V_Π passes through two given points of $\mathbb{R}^T$ (in general position with respect to each other) form a subspace of dimension $ST - 2(T - S + 1)$ since each point imposes $T - S + 1$ equations (and saying that the points are in general position means that all these equations are independent). Therefore, M is a manifold of dimension $ST - 2T + 2S$ since the pair (δ_a, δ_b) depends on two parameters, and, for a given pair, there is a $(ST - 2T + 2S - 2)$-dimensional subspace of possible Π (any two distinct points are always in general position). Hence, we see that $\dim(M) = ST - 2T + 2S < ST - T + S = \dim(X)$ since $S < T$, which implies that $G(M)$ has measure zero in X. Further, the prices where recovery is impossible, $F(G(M))$, have measure zero in the space of all prices generated by the model $F(X)$ where we use the Lebesgue measure on X to define a measure[5] on $F(X)$. ■

Proposition 1 provides a simple way to understand when recovery is possible, namely, essentially when the number of time periods T is at least as large as the number of states S.

Proposition 1 also sheds an alternative light on the critique of Borovicka, Hansen, and Scheinkman (2016) that recovery is infeasible in standard models. Indeed, we provide a simple counting argument: Suppose that the economy has growth such that, for each extra time period, the economy can increase from the previously highest state and go down from the previously lowest state. Then we get two new states for each new time period, which implies that $S > T$ such that recovery is impossible. Nevertheless, we can still achieve recovery in such a large state space if we consider a class of pricing kernels that is sufficiently low-dimensional as we discuss below in Section 5.

3.4 Further Results

We next show that our problem is indeed a generalized problem in the sense that if a solution exists satisfying the more restrictive assumptions in Ross (2015), then it is also a solution to our problem. The reverse is *not* true: a solution to the generalized recovery problem cannot be achieved in Ross's framework if the world is not time-homogeneous.

Proposition 2 (Strictly More General Method) *Suppose that we observe T pe-*

[5] We can define a measure on $F(X)$ by $\mu^*(A) := \mu(F^{-1}(A))$ for any set A, where μ is the Lebesgue measure on X.

riods of state prices given the current state at date 0 and Assumption 1 applies (time-separable utility).

1. *If Assumption 2 also applies (time-homogeneity) then a solution to Ross's Recovery problem produces a solution to our generalized recovery problem as well. Generically among price matrices for Ross's problem, the corresponding price matrix* Π *for the generalized recovery problem is full rank.*

2. *A solution to the generalized recovery problem is not in general a solution to Ross's recovery problem without Assumption 2. With* $S = T$*, there exists set of parameters with positive Lebesgue measure for the generalized recovery problem where no solution exists for Ross's recovery problem. With* $S > T$*, generically among price matrices for the the generalized recovery problem, there exists no solution to Ross's recovery problem.*

Proof. For part 1, let $\bar{\Pi}$ denote an $S \times S$ matrix of one-period state prices as considered in Ross (2015), i.e., $\bar{\pi}_{ij}$ is the value in state i at date 0 of receiving 1 in the next period if the state is j. Let F denote the corresponding matrix of one-period physical transition probabilities. A solution to Ross' problem satisfies

$$\bar{\Pi} = \delta H^{-1} F H \tag{19}$$

and therefore also by time-homogeneity for all $k = 1, \ldots, T$

$$\bar{\Pi}^k = \delta^k H^{-1} F^k H \tag{20}$$

If the starting state is 1 (without loss of generality) then the equations of our generalized recovery problem are the subset obtained by considering the first row of each equation obtained by varying k above. The equations above show that by setting the k'th row of our matrix of physical transition probabilities P equal the first row of F^k, we have a solution to the equations for our generalized recovery problem.

To see that Π is full rank, we first diagonalize Ross's price matrix as $\bar{\Pi} = VZV'$, where $Z = \text{diag}(z_1, ..., z_S)$ is the matrix of eigenvalues and V is the matrix of eigenvectors. The k'th row in the generalized-recovery pricing matrix is the first row (still assuming that the starting state is 1) of $\bar{\Pi}^k = VZ^kV'$. Letting v denote the first row in V, we see that the k'th row of Π is $vZ^kV' = (v_1 z_1^k, ..., v_S z_S^k)V'$ so

$$\Pi = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ z_1^{T-1} & \dots & z_S^{T-1} \end{bmatrix} \begin{bmatrix} v_1 z_1 & & 0 \\ & \ddots & \\ 0 & & v_S z_S \end{bmatrix} V' \tag{21}$$

Therefore, Π is full rank generically because it is the product of three full-rank matrices. Indeed, the first matrix is a Vandermonde matrix, which is full rank when the z's are non-zero and different, which is true generically. The second matrix is clearly also full-rank since the v's are also non-zero generically, and the third matrix is full rank by construction.

For part 2, consider first the case where $S < T$. The dimension of the parameter set (transition probabilities + utility parameters) generating the generalized-recovery price matrix Π is $ST - T + S$, which is strictly greater than the dimension S^2 of the parameter space generating price matrices in Ross's homogeneous case. Hence, generically no time-homogeneous solution can generate a generalized recovery price Π.

Our framework is also more general in the the case $S = T$. Recalling that $p_{\tau i}$ denotes the probability of going from the current state 1 to state i in τ periods, it is clear that in a time-homogeneous setting we must have $p_{22} \geq p_{11} p_{12}$, i.e., the probability of going from state 1 to state 2 in two periods is (conservatively) bounded below by the probability obtained by considering the particular path that stays in state 1 in the first time period and then jumps to state 2 in the second. However, such a bound need not apply for the true probabilities if the transition probabilities are not time-homogeneous. The set of parameters that can generate Π matrices that are not attainable from homogeneous transition probabilities is clearly of Lebesgue measure greater than zero in the S^2-dimensional parameter space. ■

Part 1 of the proposition shows that, when Ross's assumptions are met, a solution to his problem is also a solution to our generalized problem. Further, our method can also recover the underlying parameters (as per Proposition 1) since the price matrix Π is full rank. Part 2 of the proposition shows that for many "typical" price matrices (e.g., those observed in the data), no solution exists for Ross's recovery problem even though a solution exists for the generalized recovery problem.

We finally note that the very special case of an observed flat term structure of interest rates has some special properties. In particular, with a flat term structure there exists a solution to the problem in which the representative agent is risk neutral, echoing an analogous result by Ross.

To see this result, note that the price of a zero-coupon bond with maturity τ is equal to the sum of the τ'th row of Π, which we write as $(\Pi e)_\tau$. Having a flat term structure means that the yield on the zero-coupon bonds does not depend on maturity, i.e., that there exists a constant r such that

$$\frac{1}{(1+r)^\tau} = (\Pi e)_\tau \tag{22}$$

Let the $T \times S$ matrix Q contain the risk-neutral transition probabilities seen from the starting state, i.e., the k'th row of Q gives us the risk-neutral probabilities of ending in the different states at date k.

Proposition 3 (Flat Term Structure) *Suppose that the term structure of interest rates is flat, i.e., there exists $r > 0$ such that $\frac{1}{(1+r)^\tau} = (\Pi e)_\tau$ for all $\tau = 1, \dots, T$. Then the recovery problem is solved with equal physical and risk-neutral probabilities, $P = Q$. This means that either the representative agent is risk neutral or the recovery problem has multiple solutions.*

Proof. Let R denote the diagonal matrix whose k'th diagonal element is $\frac{1}{(1+r)^k}$. Having a flat term structure means that the matrix Π of state prices as seen from a particular starting state can be written as

$$\Pi = RQ$$

which defines Q as a stochastic matrix (i.e., with rows that sum to 1). Clearly, by letting $\delta = 1/(1+r)$ and having risk-neutrality, i.e. $H = I_S$ (the identity matrix of dimension S), we obtain a solution to our recovery problem

$$\Pi = RQ = DPH = RPI_S = RP$$

by setting $P = Q$. ■

We note that this result should be interpreted with caution. The knife-edge (i.e., measure zero) case of a flat term structure may well be generated by the knife-edge case of a price matrix Π with low rank, which implies that a continuum of solutions may exists and the representative agent may well be risk averse (as one would expect). Intuitively, a flat term structure may be generated by a Π with so much symmetry that it has a low rank.

4 Closed-Form Recovery

The recovery problem is almost linear, except for the powers of the discount rate δ which enter into the problem as a polynomial. In practical implementations over the time horizons where options are liquid, a linear approximation provides an accurate approximation given that δ is close to one. For instance, we know from the literature that δ is close to 0.97 at an annual horizon.

The linear approximation is straightforward. To linearize the discounting of δ^τ around a point δ_0 (say, $\delta_0 = 0.97$), we write $\delta^\tau \approx a_\tau + b_\tau \delta$ for known constants a_τ and b_τ. Based on the Taylor expansion $\delta^\tau \approx \delta_0^\tau + \tau\delta_0^{\tau-1}(\delta - \delta_0)$, we have $a_\tau = -(\tau - 1)\delta_0^\tau$ and $b_\tau = \tau\delta_0^{\tau-1}$. As seen in Figure 2, the approximation is accurate for $\delta \in [0.94, 1]$ for time horizons less than 2 years.

With the linearization of the polynomials in δ, the equations for the recovery problem (13) become the following:

$$\begin{pmatrix} \pi_{11} \\ \vdots \\ \pi_{T1} \end{pmatrix} + \begin{pmatrix} \pi_{12} & \dots & \pi_{1S} \\ \vdots & & \vdots \\ \pi_{T2} & \dots & \pi_{TS} \end{pmatrix} \begin{pmatrix} h_2^{-1} \\ \vdots \\ h_S^{-1} \end{pmatrix} = \begin{pmatrix} a_1 + b_1\delta \\ \vdots \\ a_T + b_T\delta \end{pmatrix} \tag{23}$$

which we can rewrite as a system of T equations in S unknowns as

$$\begin{pmatrix} -b_1 & \pi_{12} & \dots & \pi_{1S} \\ \vdots & \vdots & & \vdots \\ -b_T & \pi_{T2} & \dots & \pi_{TS} \end{pmatrix} \begin{pmatrix} \delta \\ h_2^{-1} \\ \vdots \\ h_S^{-1} \end{pmatrix} = \begin{pmatrix} a_1 - \pi_{11} \\ \vdots \\ a_T - \pi_{T1} \end{pmatrix} \tag{24}$$

Rewriting this equation in matrix form as

$$Bh_\delta = a - \pi_1 \tag{25}$$

we immediately see the closed-form solution

$$h_\delta = \begin{cases} B^{-1}(a - \pi_1) & \text{for } S = T \\ (B'B)^{-1}B'(a - \pi_1) & \text{for } S < T \end{cases} \tag{26}$$

We see that, when $S = T$, we simply need to solve S linear equations with S unknowns. When $S < T$, we could simply just consider S equations and ignore the remaining $T - S$ equations.

More broadly, if $S < T$ and we start with prices Π that are not exactly generated by the model (e.g., because of noise in the data), then (26) provides the values of δ and the vector h that best approximate a solution in the sense of least squares.

The following theorem shows that the closed-form solution is accurate as long as the value of δ_0 is close to the true discount rate:

Proposition 4 (Closed-Form Solution) *If prices are generated by the model and B has full rank $S \leq T$ then the closed-form solution (26) approximates the true solution in the following sense: The distance between the true solution $(\bar{\delta}, \bar{h}, \bar{P})$ and the approximate solution (δ, h, P) approaches 0 faster than $(\delta_0 - \bar{\delta})$ as δ_0 approaches $\bar{\delta}$.*

Proof. The approximation result follows from Lemma 1 in the appendix. ■

5 Recovery in a Large State Space

A challenge in implementing the Ross Recovery Theorem is that it does not allow for an expanding set of states as we know it, for example, from binomial models and multinomial models of option pricing. Simply stated, the expanding state space in a binomial model adds more unknowns for each time period than equations even under the assumption of utility functions that depend on the current state only. We next show how we handle an expanding state space in our model.

We have in mind a case where the number of states S is larger than the number of time periods T. In a standard binomial model, for example, with two time periods we need five states corresponding to the different values that the stock can take over its path. The key to solving this problem is to reduce the dimensionality of the utility ratios captured in the vector h. To do that, we replace Assumption 1 with the following assumption that the pricing kernels belong to a parametric family with limited dimensionality.

Assumption 1* (General utility with N parameters) *The pricing kernel at time τ in state s (given the initial state 1 at time 0) can be written as*

$$m_{0,\tau}^{1,s} = \delta^{\tau} h_s(\theta) \tag{27}$$

where $\delta \in (0, 1]$ and $h(\cdot) > 0$ is a one-to-one C^{∞} smooth function of the parameter

$\theta \in \Theta$, an embedding from $\Theta \subset \mathbb{R}^N$ to $\mathbb{R}^S$, and Θ has a non-empty interior.

With a large number of unknowns compared to the number of equations, we need to restrict the set of unknowns, and this is done by assuming that the utilities are parameterized by a lower-dimensional set Θ.

5.1 A Large Discrete State Space

Let us first consider two simple examples of how we can parameterize marginal utilities with a low-dimensional set of parameters. First, we consider a simple linear expression for the marginal utilities and then we discuss the case of constant relative risk aversion (a non-linear mapping from risk aversion parameters Θ to marginal utilities).

We start with a simple linear example of how the parametrization works. We consider a matrix B of full rank and dimension $(S-1) \times N$ such that

$$\begin{bmatrix} h_2^{-1} \\ \vdots \\ h_S^{-1} \end{bmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_{S-1} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1N} \\ \vdots & & \vdots \\ b_{S-1,1} & \dots & b_{S-1,N} \end{pmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_N \end{bmatrix} = A + B\theta \tag{28}$$

Combining this equation with the recovery problem (15) gives

$$(\Pi_1 + \Pi_2 A) + \Pi_2 B \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_N \end{pmatrix} = \begin{pmatrix} \delta \\ \vdots \\ \delta^T \end{pmatrix} \tag{29}$$

This equation has *exactly* the same form as our original recovery problem (15), but now $\Pi_1 + \Pi_2 A$ plays the role of Π_1, similarly $\Pi_2 B$ plays the role of Π_2, and θ plays the role of $(h_2^{-1}, ..., h_S^{-1})'$. The only difference is that the dimension of the unknown parameter has been reduced from $S-1$ to N. Therefore, Proposition 1 holds as stated with S replaced by $N+1$.

Hence, while before we could achieve recovery if $S \leq T$, now we can achieve recovery as long as $N + 1 \leq T$. In other words, recovery is possible as long as the representative agent's utility function can be specified by a number of parameters that is small relative to the number of time periods for which we have price data.

Assumption 1* also allows for the marginal utilities to be non-linear function of the risk aversion parameters θ. This generality is useful because standard utility functions may give rise to such a non-linearity. As a simple example, consider an economy with a representative agent with CRRA preferences. In this economy, the pricing kernel in state s at time τ (given the current state 1 at time 0) is

$$m_{0,\tau}^{1,s} = \delta^{\tau} \left(\frac{c_s}{c_1} \right)^{-\theta} \tag{30}$$

where c_s is the known consumption in state s of the representative agent and θ is the unknown risk aversion parameter. Hence, Assumption 1* is clearly satisfied with $h_s^{-1}(\theta) = (\frac{c_s}{c_1})^{\theta}$. Our generalized recovery result extends to the large state space as stated in the following proposition.

Proposition 5 (Generalized Recovery in a Large State Space) *Consider an economy satisfying Assumption 1* with Arrow-Debreu prices for each of the T time periods and S states such that $N + 1 < T$. The recovery problem has*

1. *no solution generically in terms of an arbitrary Π matrix of positive elements;*

2. *a unique solution generically if Π has been generated by the model.*

Proof. Following the same logic as the proof of Proposition 1, we note that the set X of all (δ, θ, P) is a manifold-with-boundary of dimension $S \cdot T - T + N + 1$. The discount rate, marginal utility parameters, and probabilities map into prices, which we denote by $F(\delta, \theta, P) = DPH = \Pi$, where, as before, $D = \text{diag}(\delta, ..., \delta^T)$ and $H = \text{diag}(h_1(\theta), h_2(\theta), ..., h_S(\theta)))$, and F is C^∞. Since $N + 1 < T$, the image $F(X)$ has Lebesgue measure zero in $\mathbb{R}^{T \times S}$ by Sard's theorem, proving part 1.

Turning to part 2, we first note that P can be uniquely recovered from $(\bar{\theta}, \Pi)$ using equation (12), where $\bar{\theta} = (\delta, \theta)$. Therefore, we can focus on $(\bar{\theta}, \Pi)$, studying the solutions to $\Pi(h_1^{-1}(\theta), ..., h_S^{-1}(\theta))' = (\delta, ..., \delta^T)'$.

For two different choices of the parameters $(\bar{\theta}_a, \bar{\theta}_b)$ and a single set of prices Π, we consider the triplet $(\bar{\theta}_a, \bar{\theta}_b, \Pi)$. We are interested in showing that the different parameters cannot both be consistent with the same prices, generically. To show this, we consider the space M where the reverse is true, hoping to show that M is "small." Specifically, M is the set of triplets where Π is of full rank and both discount rates are consistent with the prices, that is, there exists (unique) P_i $(i = a, b)$ such that $D_a P_a H_a = D_b P_b H_b = \Pi$.

Given that probabilities can be uniquely recovered from prices and parameters, we have a smooth map G from M to X by mapping any triplet $(\bar{\theta}_a, \bar{\theta}_b, \Pi)$ to $(\delta_a, \theta_a, P_a)$. The image of this map consists exactly of those elements of X for which F is not injective. The proof is complete if we can show that this image has Lebesgue measure zero, which follows again by Sard's theorem if we can show that the dimension of M is strictly smaller than $S \cdot T - T + N + 1$.

To study the dimension of M, consider first $V_\Pi := \{\Pi(h_1^{-1}(\theta), ..., h_S^{-1}(\theta))' | \theta \in \Theta\}$, which is an N-dimensional submanifold of $\mathbb{R}^T$ for Π of full rank and given that h is a one-to-one embedding. We note that we can think of M as the space of triplets such that V_Π contains both the points $(\delta_a, \delta_a^2, ..., \delta_a^T)'$ and $(\delta_b, \delta_b^2, ..., \delta_b^T)'$, where the corresponding θ's are given uniquely from the definition of V_Π since Π is full rank and h is one-to-one. The set of all those $\Pi \in \mathbb{R}^{T \times S}$ such that V_Π passes through two given points of $\mathbb{R}^T$ form a subspace of dimension $ST - 2(T - N)$ since each point imposes $T - N$ equations. Therefore, M is a manifold of dimension $ST - 2T + 2N + 2$. Hence, we see that $G(X)$ has measure zero in X and $F(G(X))$ has measure zero in $F(X)$. ■

As one simple application of the proposition, we can recover preferences from state prices if we know that the pricing kernel is bounded and we have sufficiently many time periods as seen in the following corollary. Said differently, using a simplified or winsorized pricing kernel (or state space) is a special case of Proposition 5.

Corollary 6 (Generalized Recovery with Bounded Kernel) *Suppose that the* *pricing kernel is bounded in the sense that there exist states $\bar{s} > \underline{s}$ such that $h_s = h_{\bar{s}}$ for $s > \bar{s}$ and $h_s = h_{\underline{s}}$ for $s < \underline{s}$. Then the conclusion of Proposition 5 applies, where N is the number of states from $\underline{s}$ to $\bar{s}$.*

5.2 Continuous State Space

Finally, we note that our framework also easily extends to a continuous state space under Assumption 1*. We start with a continuous state-space density $\pi_\tau(s)$ at each time point $\tau = 1, \ldots, T$ (given the current state at time 0). As before, $\pi_\tau(s)$ represents Arrow-Debreu prices or, more precisely, $\pi_\tau(s)ds$ represents the current value of receiving 1 at time τ if the state is in a small interval ds around s. Similarly, we let $p_\tau(s)$ denote the physical probability density of transitioning to s in τ periods. The fundamental recovery equations now become

$$\pi_\tau(s) = \delta^\tau h_s(\theta) p_\tau(s) \tag{31}$$

By moving h to the left-hand side and integrating, we can eliminate the natural probabilities as before.

$$\int \pi_\tau(s) h_s^{-1}(\theta) ds = \delta^\tau \tag{32}$$

For each time period τ, this gives an equation to help us recover the $N+1$ unknowns, namely the discount rate δ and the parameters $\theta \in \mathbb{R}^N$. Hence, we are in the same situation as in the discrete-state model of Section 5.1, and we have recovery if there are enough time periods as stated in Proposition 5.

As before, the linear case is particularly simple. Suppose that the marginal utilities can be written as[6]

$$h_s^{-1}(\theta) = A(s) + B(s)\theta \tag{33}$$

[6] Note that $h_s^{-1}(\theta)$ denotes $\frac{1}{h_s(\theta)}$, i.e., it is *not* the inverse function of $h_s(\theta)$.

where, for each s, $A(s)$ is a known scalar and $B(s)$ is a known row-vector of dimension N. Using this expression, we can rewrite equation (32) as a simple equation of the same form as our original recovery problem (15):

$$\pi_\tau^A + \pi_\tau^B\ \theta = \delta^\tau \tag{34}$$

where $\pi_\tau^A = \int \pi_\tau(s)A(s)ds$ and $\pi_\tau^B = \int \pi_\tau(s)B(s)ds$. Hence, as before, we have T equations that are linear except for the powers of the discount rate.

6 Recovery in Specific Models: Examples

In this section we investigate recovery of specific models of interest. In a controlled environment, we show when, given state prices, our model recovers the true underlying risk-aversion parameter, time-preference parameter along with the true multiperiod physical probabilities.

6.1 Recovery in the Mehra and Prescott (1985) model

The Mehra and Prescott (1985) model works as follows. The aggregate consumption either grows at rate $u = 1.054$ or shrinks at rate $d = 0.982$ over the next period. This consumption growth between time $t-1$ and t is captured by a process X_t. The aggregate consumption process can be written as

$$Y_t = \prod_{s=1}^{t} X_s \tag{35}$$

where the initial consumption is normalized as $Y_0 = 1$.

Consumption growth X_t is a Markov process with two states, up and down. The probability of having an up state after an up state is $\phi_{uu}; = Pr(X_t = u|X_{t-1} = u) = 0.43$ and, equally, the probability of staying in the down state is $\phi_{dd} = 0.43$. Hence, the probability of switching state is $\phi_{ud} = \phi_{du} = 0.57$.

The Arrow-Debreu price of receiving 1 at time t in a state $s_t = (y_t, x_t)$ is computed based on the CRRA preferences for the representative agent with risk aversion $\gamma = 4$ as

$$\pi_{0,t}^{1,s_t} = \delta^t y_t^{-\gamma} Pr(X_t = x_t, Y_t = y_t) \tag{36}$$

where the time-preference parameter is $\delta = 0.98$ and the physical probabilities $Pr(X_t = x_t, Y_t = y_t)$ of each state are computed based on the Markov probabilities above.[7]

Based on this model of Mehra and Prescott (1985), we compute Arrow-Debreu prices in each state over $T = 20$ time periods and examine whether we can recover probabilities and preferences based on knowing only these prices (we have also performed the recovery for other values of T).

Impossibility of general recovery. We first notice from equation (35) that consumption has growth, which immediately implies that $S > T$. This means that recovery is impossible without further assumptions. Hence, we proceed using the method concerning a large state space of Section 5.

Recovery under CRRA. The simplest way to proceed is to assume that we know the form of the pricing kernel (36), but we don't know the risk aversion γ, the discount rate δ, or the probabilities. We can then write the Generalized Recovery equation set on the form

$$\Pi h^{-1}(\gamma) = \begin{bmatrix} \delta & \delta^2 & \dots & \delta^T \end{bmatrix}' \tag{37}$$

where h is a one-to-one C^∞ smooth function of the parameter γ based on (36), see Appendix B for details.[8] Therefore, we are in the domain of Assumption 1* and, as

[7]We note that prices of long-lived assets, for example the overall stock market, depends on both X_t and Y_t (even if the aggregate consumption Y_t is the aggregate dividend). Therefore, stock index options would provide information on Arrow-Debreu prices on each state $s_t = (y_t, x_t)$. Alternatively, we could consider recovery based only on Arrow-Debreu securities that depend on y_t. This would correspond to observing options on "dividend strips." Either way, we get the same recovery results in the Mehra and Prescott (1985) model.

[8]Matlab code is available from the authors upon request.

long as $T > 2$ (since $N = 1$ is the number of risk aversion parameters and 2 is the total number of variables, δ and γ) then by Proposition 5 we know that the Generalized Recovery equation set generically has a unique solution.

We first seek to recover γ and δ by minimizing the pricing errors (again, see Appendix B for details). Panel A of Figure 3 shows the objective function for this minimization problem. As seen from the figure, there is a unique solution to the problem, which naturally equals the true parameters $\hat{\delta} = 0.98, \hat{\gamma} = 4$.

Finally, we turn to the recovery of natural probabilities. It is worth noticing that we do *not* recover the Markov switching probabilities $\phi_{uu}, \phi_{dd}, \phi_{ud}$ or ϕ_{du}. Rather, what is recovered is the multi-period probabilities $p_{0,t}^{1,s_t}$ of transitioning from the initial state to each future state (consistent with the intuition conveyed in Figure 1).[9] The probabilities $p_{0,t}^{1,s_t}$ are recovered exactly. Fortunately, these multi-period probabilities are all we need for making predictions about such statistics as expected returns, variances, and quantiles across different time horizons.

6.2 Black-Scholes-Merton and iid. consumption growth

We can capture a binomial model in the spirit of Black-Scholes-Merton and Cox, Ross, and Rubinstein (1979) as follows. We consider the same model for aggregate consumption Y_t, but now X_t is iid. (corresponding to $\phi_{uu} = \phi_{du}$ and $\phi_{dd} = \phi_{ud}$). In other words, the standard binomial Black-Scholes-Merton model has iid consumption growth. Specifically, we assume that up and down probabilities are always 50% ($\phi_{uu} = \phi_{du} = \phi_{dd} = \phi_{ud} = 0.5$).

This binomial model implies a flat term structure which puts us in the case of Proposition 3, where recovery is impossible.[10] Concretely, the problem is that the price matrix Π from (37) is not full rank. Hence, as seen in Figure 3 Panel B, the

[9] Recovery of the underlying path-dependent probabilities is possible if we have access to Arrow-Debreu prices for all paths or if we assume that we know the structure of the underlying tree.

[10] Iid. consumption growth and standard utility functions generally lead to a flat term structure because the price of a bond with τ periods to maturity can be written as $E_t(\delta^\tau \frac{u_{t+\tau}}{u_t}) = E_t(\prod_{s=1,\ldots,\tau} \delta \frac{u_{t+s}}{u_{t+s-1}}) =: (\frac{1}{1+r})^\tau$, where the expected utility increments are the same for all s because they depend on consumption growth $\frac{c_{t+s}}{c_{t+s-1}}$, which has constant expected value when it is iid.

objective of minimizing pricing errors has a continuum of solutions. In other words, recovery is not feasible.

6.3 A non-stationary model without Markov structure

Lastly, we consider a model where the consumption growth X_t is not Markov. Specifically, we still consider the binomial tree described above in Sections 6.1–6.2, but now we let the probability of transitioning up/down from any state s at any time t depend on the path taken from time 0 to time t. At each node at each path, we draw a random uniformly distributed probability for an "up" move, and, of course, assign one minus this probability to the next "down" node.

We now seek to recover δ and γ. As seen in Figure 3 Panel C, the objective function has a unique solution which again equals the true parameters $\hat{\delta} = 0.98$ and $\hat{\gamma} = 4$. Hence, recovery can be possible even when the driving process is non-stationary and non-Markovian, again under parametric assumptions about the utility function (i.e., a model outside the scope of Ross (2015) and Borovicka, Hansen, and Scheinkman (2016)).

7 Data and Empirical Methodology

This section describes our data and empirical methodology.

7.1 Data and Sample Selection

We use the Ivy DB database from OptionMetrics to extract information on standard call and put options written on the S&P 500 index for every last trading day of the month from January 1996 to December 2015. We obtain implied volatilities, strikes, and maturities, allowing us to back out market prices. As a proxy for the risk-free rate, we use the zero-coupon yield curve of the Ivy DB database, which is derived from LIBOR rates and settlement prices of CME Eurodollar futures. We also obtain expected dividend payments, calculated under the assumption of a constant dividend yield over the life time of the option. We consider options with time to maturity between 10 and 360 days and apply standard filters, excluding contracts with zero open interest, zero trading volume, and quotes with best bid below $0.50, and options with implied volatility higher than 100%.

7.2 Recovery Methodology

The Generalized Recovery Theorem relies on the knowledge of Arrow-Debreu state prices from the current initial state to all possible future states for several future time periods. Unfortunately, there is currently no market trading pure Arrow-Debreu securities. Therefore, we use options to back out Arrow-Debreu prices. Further, given the large number of states, we use the parametric kernel method from Section 5.

To study the robustness of recovery, we consider two different methods for backing out Arrow-Debreu prices and two different specifications of the pricing kernel, for a total of four different recovered distributions and preferences.

More specifically, we apply the following two methods of extracting Arrow-Debreu prices from options: (i) the parametric model of Bates (2000) and (ii) the non-parametric method of Jackwerth (2004). Each of the methods yields Arrow-Debreu prices across multiple time horizons and mutliple index levels for each day t as described in detail in Appendix C. Given these observed Arrow-Debreu prices, we recover preferences and probabilities based on the following two specifications of the pricing kernel.

Piecewise linear. The inverse marginal utilities are piecewise linear over states. Given the initial state 1 at time 0 the τ-period inverse marginal utility ratio in state s is[11]:

$$(h_s^\tau(\theta))^{-1} = B_s\theta \tag{38}$$

[11]Notice again that $(h_s^\tau(\theta))^{-1} = \frac{1}{h_s^\tau(\theta)}$ and is *not* the inverse function.

Here θ is an N-dimensional column vector and B_s is the $s'th$ row of the known $S \times N$ "design matrix" B. In our empirical implementation N is 5[12]. Interpreting the parameters $\theta_1, ..., \theta_N$ we let the first parameter θ_1 determine the initial level of the inverse pricing kernel $H^{-1}e = B\theta$. The next parameter, θ_2, determines the initial slope of the first line segment. Similarly, θ_3 is the slope of the next line segment generated by $B\theta$.

We impose that $\theta_1, ..., \theta_N \geq 0$ which means that the inverse pricing kernel is monotonically increasing or, equivalently, that the pricing kernel is monotonically decreasing i.e., that marginal utility decreases at higher levels of wealth.

The design matrix is characterized by its "break points" that separate the state space into $N-2$ regions. These regions are chosen as follows. The lowest region ranges over states from $(1-2.5\text{VIX}_t)S_t$ to $(1-2\text{VIX}_t)S_t$ where S_t is the current (time t) level of the S&P 500 index. The highest region covers states ranging from $(1+2\text{VIX}_t)S_t$ to $(1+4\text{VIX}_t)S_t$. In between these extremes, we consider $N-3$ regions of equal size in the range $(1-2\text{VIX}_t)S_0$ to $(1+2\text{VIX}_t)S_t$. When using this specification of B and the estimated Arrow-Debreu prices, we obtain an $S \times N$ matrix ΠB with full rank for every last trading day of the month for the period 1/1996 to 12/2015.

With this in place we set up the following minimization problem

$$\begin{aligned} \min_{\theta,\delta} \ & \text{norm}\left(D^{-1}\Pi B\theta - 1\right) \\ s.t. \quad & \theta > 0 \\ & \delta \in (0,1] \end{aligned} \tag{39}$$

Given a state price matrix Π and a design matrix B we estimate the θ and δ that best fit the model in a squared error sense. Once the marginal utilities and discount

[12]The lowest number of maturities with observed option prices in our sample is 7. Therefore, we can impose a structure on the pricing kernel with at most 6 parameters and hence N can at most be 5 because of the sixth parameter δ.

rate have been recovered, we back out the multi-period physical probabilities as

$$P = D^{-1}\Pi \ \mathrm{diag}(B\theta) \tag{40}$$

where D is a diagonal matrix with elements $D_{ii} = \delta^i$ and $\mathrm{diag}(B\theta)$ is a diagonal matrix with elements $\mathrm{diag}(B\theta)_{jj} = B_j\theta$ where B_j is the j'th row of B. We normalize P to have row sums of one, which is necessary since θ and δ are found from the minimization problem in (39) and not solved perfectly.

Polynomial. The inverse marginal utility ratio is a polynomial in the return on the market and time horizon. Given the initial state 1 at time 0 the τ-period inverse marginal utility ratio in state s is:

$$(h_s^\tau(\theta))^{-1} = \beta_0 + \beta_1 r_s + \beta_2 r_s^2 + \beta_3 \tau r_s + \beta_4 \tau r_s^2 \tag{41}$$

Here $r_s = S_s/S_1 - 1$ is the return on the market in state s. The parameters of interest are $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$. In our implementation we impose three conditions on the parameters; (i) $\beta_0 > 0$, ensuring a positive pricing kernel when $r = 0$, (ii) the risk-premium is non-negative and, (iii) the inverse marginal utility ratios are always strictly positive (we set a lower bound on the inverse marginal utility ratio at 0.01.). This means that the parameters $\beta_1, \beta_2, \beta_3, \beta_4$ can move freely (within the space of the conditions) and are all allowed to be either positive or negative.

The polynomial specification of the inverse marginal utility ratios illustrates one possible way of imposing structure on the marginal utilities, not only in the state dimension, but also in the time horizon dimension. This specification allows marginal utilities in a given state, say s, to differ when considering different time horizons, that is, e.g. $h_s^\tau(\theta) \neq h_s^{\tau+1}(\theta)$. The polynomial specification nests the linear specification as a special case when $\beta_2, \beta_3, \beta_4$ are all zero.

The minimization procedure for the polynomial specification is:

$$\min_{\theta,\delta} \sum_{t=1}^{T} \left(\left(\sum_{s=1}^{S} \delta^{-t} \pi_{ts} (h_s^\tau(\theta))^{-1} \right) - 1 \right)^2 \tag{42}$$

$$\begin{aligned} s.t. \quad & \beta_0 > 0 \\ & E_0^P(r_t|\theta,\delta) - r_t^f \geq 0 \quad \text{for all } t \in (1, ..., T) \\ & (h_s^\tau(\theta))^{-1} > 0 \quad \text{for all } s \in (1, ..., S) \text{ and all } \tau \in (1, ..., T) \\ & \delta \in (0, 1] \end{aligned}$$

where π_{ts} is the state price in state s with time horizon t. Here $E_0^P(r_t|\theta,\delta) - r_t^f$ is the excess return given parameter values $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$, and δ.

Given estimates of $\delta, \beta_0, \beta_1, \beta_2, \beta_3$ and β_4 we can arrive at the t period physical probabilities as

$$P_t = \delta^{-t} \Pi_t \operatorname{diag}\left((h_s^\tau(\theta))^{-1}\right) \tag{43}$$

where Π_t is the t'th row of the state price matrix Π and r is an $S \times 1$-dimensional vector of returns over states. We normalize P to have row sums of one, this is necessary since θ and δ are found from the minimization problem in (42) and not solved perfectly.

7.3 Computing Statistics under the Physical Probability Distribution

Once we have recovered the probabilities of each state for each future time period, it is straightforward to compute any statistic under the physical probability distribution. If the level of the index at time t is S_t, then the state space consists of all integer values of the index between the minimum value $(1 - 2.5\text{VIX}_t)S_t$ and $(1 + 4\text{VIX}_t)S_t$. Let N_t denote the number of states as seen from time t and think of state 1 as the lowest state and N_t as the highest state. We compute the recovered expected return at time t by summing over the N_t possible states:

$$E_t^{\mathbb{P}}[r_{t,t+1}] = \sum_{\nu=1}^{N_t} p_{t+1,\nu} r_{t+1,\nu} \tag{44}$$

where $p_{t+1,\nu}$ is the recovered time-t conditional physical probability for the transition to state ν at time $t+1$, $r_{t+1,\nu} = \frac{S_{t+1}(\nu)}{S_t} - 1$ is the return in state ν, and $S_{t+1}(\nu)$ is the value of the index at time $t+1$ if state ν is realized.

The recovered excess return is naturally given by

$$\mu_t = E_t^{\mathbb{P}}[r_{t,t+1}] - r_{t,t+1}^f \tag{45}$$

where $r_{t,t+1}^f$ is the risk-free rate. We compute the contemporaneous unpredictable innovation in the conditional expected return as

$$\Delta\mu_{t+1} = \mu_{t+1} - E_t[\mu_{t+1}] \tag{46}$$

where we impose an AR(1)-process on the innovation to the risk premium $E_t[\mu_{t+1}] = \alpha_0 + \alpha_1 \mu_t$ based on the regression

$$\mu_{t+1} = \alpha_0 + \alpha_1 \mu_t + \varepsilon_{t+1} \tag{47}$$

The estimated persistence parameter α_1 is 0.3 at the monthly horizon.

We compute the recovered conditional variance, $\mathrm{VAR}_t^{\mathbb{P}}(r_{t,t+1})$, analogously to how we computed the expected return:

$$\mathrm{VAR}_t^{\mathbb{P}}(r_{t,t+1}) = \sum_{\nu=1}^{N_t} p_{t+1,\nu} \left(r_{t+1,\nu} - E_t^{\mathbb{P}}[r_{t,t+1}] \right)^2 \tag{48}$$

and we denote the recovered volatility by $\sigma_t = \sqrt{\mathrm{VAR}_t^{\mathbb{P}}(r_{t,t+1})}$.

8 Empirical Results

We next investigate the properties of the recovered probabilities based on each of our four methods. We first consider the recovered expected return. Table 1 shows the correlation matrix for the recovered expected returns based on each of our four methodologies as well as the VIX volatility index and the SVIX variable of Martin (2017). The good news is that all variables are positively correlated, as we would expect. The less good news is that the correlations between the different recovered expected returns are modest in magnitude, with an average pairwise correlation of only 0.5. This modest correlation is concerning because all these recovered expected returns should be measures of the same thing, namely the market's expected return at any given time.

Figure 4 shows the time series variation of the recovered expected return based on one of the methodologies (we plot just one time series since it is difficult to look at all four together). These recovered expected returns do not look unreasonable, but we next try to test their predictability of actual realized returns. Specifically, we regress the ex post realized excess return on the ex ante recovered expected excess return, μ_t, and the ex post innovation in expected return, $\Delta\mu_{t+1}$:

$$r_{t,t+1} = \beta_0 + \beta_1\mu_t + \beta_2\Delta\mu_{t+1} + \epsilon_{t,t+1} \tag{49}$$

where ϵ_{t+1} is a noise term. To understand this regression, note that we are interested in testing whether the recovered probabilities give rise to reasonable expected returns, that is, time-varying risk premia. For this, we want to test whether a higher ex ante expected return is associated with a higher ex post realized return ($\beta_1 > 0$), whether an increase in the risk premium is associated with a contemporaneous drop in the price ($\beta_2 < 0$), and whether the intercept as zero ($\beta_0 = 0$).

Table 2 reports the results of this regression for each of our four recovery methodologies as well as using VIX and SVIX as the expected return over the full sample from 1997 to 2015. First, the intercept β_0 is insignificantly different from zero in most specifications, but significantly different from zero using method 2 and using VIX, providing evidence against these models. Second, β_1 is positive and marginally significant from 0 in model 1, but otherwise insignificantly different from zero, providing neither evidence in favor or against the models. The coefficient β_2 is highly significant and has the desired negative sign in all models. Further, as expected the absolute value of β_2 is greater than one since a shock to the discount rate leads to a larger shock to the price (cf. Gordon's growth model for the extreme example of a permanent shock).

Table 3 reports the result of regression (49) over the sub-sample that excludes the global financial crisis (9/2008–7/2009), a sub-sample that has been considered in the literature (e.g., Martin (2017)). The results here are stronger and more consistent with theory. All the key parameters have the expected sign, the estimated coefficient β_0 is small and insignificant in all models, the estimated coefficient β_1 is positive and marginally significant or insignificant, and β_2 is negative and significant.

The reason that the models work better when we exclude the crisis is intuitive: During the crisis, there were several months in which the ex ante recovered expected return was high, but, nevertheless, the ex post realized return was negative and large in magnitude. It seems plausible that investors were scared at that time, which means that it is plausible that the true required return was indeed high, which in turn implies that the negative realized return was a negative surprise. Hence, one could argue that the model gets this period wrong for the "right" reason, but we don't want to push this argument too far as the most compelling evidence is almost always that of using the full sample.

Finally, we consider the recovered physical volatility as plotted in Figure 5. This recovered volatility looks reasonable. Further, the recovered volatilities are similar across the different methodologies with an average pairwise correlation of 0.95 and an average correlation to VIX of 0.92. It is not that surprising that volatilities can be recovered, but studying volatility provides a simple and powerful reality check of our method since the true future volatility is known with much less error than the expected return. Hence, we regress the ex post realized volatility on the ex ante recovered conditional volatility, σ_t:

$$\sqrt{\text{VAR}(r_{t,t+1})} = \beta_0 + \beta_1 \sigma_t + \epsilon_{t,t+1} \tag{50}$$

where the realized volatility $\sqrt{\text{VAR}(r_{t,t+1})}$ is computed using close-to-close daily data over the 4 weeks from t to $t+1$ by OptionMetrics. We also run the same regression where we replace the recovered volatilities by the VIX volatility index. The theory predicts that $\beta_0 = 0$ and $\beta_1 = 1$.

Table 4 reports the results. As seen in Table 4, the estimated intercept coefficient β_0 is insignificant for models 1 and 2, but significant for models 3 and 4. However, for all models, the intercept is smaller than that of VIX, suggesting that the recovered volatilities are less biased than VIX.

The estimated slope coefficient β_1 is positive and highly significant for all models. Further, the estimated slope is close to the predicted value of 1, in particular closer than the estimated value for VIX. Lastly, we see that VIX has a slightly higher R^2, which may reflect that the recovery method introduces some noise in the volatility measures.

In summary, we find substantial differences across the recovered probabilities based on different methodologies, and the predictive power for future returns appears weak in the full sample, but slightly stronger in the sample that excludes the global financial crisis. The recovered volatilities predict well the future volatility in a way that is less biased than VIX, but slightly lower R^2. We are able to reject that the recovered probabilities provide a perfect description of the future evolution of the market based on a Berkowitz (2001) test.[13]

[13] The details of this test are not reported for brievity. The idea is that, given the estimated distribution $\hat{F}_t$ of the excess return r_{t+1} at time t, the distribution of the transformed variable $u_{t+1} = \hat{F}_t(r_{t+1})$ should be uniform and the distribution of the further transformed variable $x_{t+1} = \Phi^{-1}(u_{t+1})$ should be standard normal, which is tested by estimated the coefficients in the model $x_{t+1} = c + \beta x_t + \epsilon_t$ and perform a likelihood ratio test of the joint hypothesis that $c = \beta = 0$ and $Var(\epsilon_t) = 1$.

9 Conclusion

We characterize when preferences and natural probabilities can be recovered from observed prices using a simple counting argument. We make no assumptions on the physical probability distribution, thus generalizing Ross (2015) who relies on strong time-homogeneity assumptions.

In economies with growth, our counting argument immediately shows that recovery is generally not feasible. While this finding parallels results by Borovicka, Hansen, and Scheinkman (2016), our intuitive counting argument does not rely on the assumptions of an infinite-period time-homogeneous Markov setting, but, rather, is based on the general methods pioneered by Debreu (1970) for general equilibrium.

To pursue recovery even in economies with growth, e.g., classical multinomial models, we show how our method can be used when the pricing kernel can be parameterized by a sufficiently low-dimensional parameter vector. When recovery is feasible, our model allows a closed-form linearized solution.

We implement our model empirically using several different specifications, testing the predictive power of the recovered statistics. Our empirical findings indicate that the ex ante expected returns based on our recovered physical probabilities vary substantially across specifications, highlighting the difficulty in applying the method in a robust way. Further, the overall evidence suggests that the recovered distribution's predictive power for future returns is weak and we reject that the full recovered probability distribution is a perfect description of the future market behavior. The recovered expected returns better predict the realized returns in the sub-sample that excludes the global financial crisis and the recovered volatilities predict well the future realized volatility. Hence, while our theory sheds new light on a fundamental asset pricing question, the empirically usefulness of the recovered probabilities remains an interesting challenge for future research.

A Appendix: Proofs

Lemma 1 *Suppose that $x^* \in \mathbb{R}^n$ is defined by $f(x^*) = 0$ for a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^n$ with full rank of the Jacobian df in the neighborhood of x^*, and x is defined as the solution to the equation, $f(\bar{x}) + df(\bar{x})(x - \bar{x}) = 0$, where f has been linearized around $\bar{x} = x^* + \Delta x\, \varepsilon$ for $\Delta x \in \mathbb{R}^n$ and $\varepsilon \in \mathbb{R}$. Then $x = x^* + o(\varepsilon)$ for $\varepsilon \to 0$.*

Proof. Since we have $x = \bar{x} - df^{-1} f(\bar{x})$ we see that, as $\varepsilon \to 0$,

$$\frac{x - x^*}{\varepsilon} = \frac{\bar{x} - x^*}{\varepsilon} - df^{-1} \frac{f(\bar{x}) - f(x^*)}{\varepsilon} \to \Delta x - df^{-1} df \Delta x = 0 \tag{A.1}$$

■

B Appendix: Details on Recovery in Mehra-Prescott

Let

$$\Pi = \begin{bmatrix} \pi_{0,1}^{0,d} & \pi_{0,1}^{1,u} & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \pi_{0,2}^{0,d} & \pi_{0,2}^{1,d} & \pi_{0,2}^{1,u} & \pi_{0,2}^{2,u} & 0 & \dots & 0 & 0 & 0 & & 0 \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \pi_{0,T}^{0,d} & \pi_{0,T}^{1,d} & \pi_{0,T}^{1,u} & \dots & \pi_{0,T}^{T,u} \end{bmatrix} \tag{B.1}$$

where $\pi_{0,t}^{k,u}$ is the state price of making a total of k "up" moves in t periods where the last move was "up," that is, the Arrow-Debreu price for the state $s_t = (y_t, x_t) = (u^k d^{t-k}, u)$. Similarly, $\pi_{0,t}^{k,d}$ is the state price of making a total of k "up" moves in t periods where the last move was "down".

Π has dimension $T \times (\sum_{t=1}^{T} 2t)$. This implies that the $h^{-1}(\gamma)$ vector of inverse marginal utility ratios must be $(\sum_{t=1}^{T} 2t)$-dimensional. We fix this in the following way. We let

$$h^{-1}(\gamma) = \begin{bmatrix} (y_1^0)^\gamma & (y_1^1)^\gamma & (y_2^0)^\gamma & (y_2^1)^\gamma & (y_2^1)^\gamma & (y_2^2)^\gamma & \ldots & (y_T^T)^\gamma \end{bmatrix}' \tag{B.2}$$

where $y_t^k = u^k d^{t-k}$ is the level of aggregate consumption when making a total of k "up" moves in t periods and γ is the risk-aversion parameter that we wish to recover.

There is no closed-form solution to the non-linear case of CRRA preferences. In order to obtain model estimates we sort to a numerical exercise, that is to minimize the objective function g:

$$\min_{\gamma,\delta} g(\gamma,\delta) := \text{norm}\left(\Pi h^{-1}(\gamma) - \begin{bmatrix} \delta \\ \delta^2 \\ \vdots \\ \delta^T \end{bmatrix} \right) \tag{B.3}$$

$$s.t. \quad \gamma \in \mathbb{R}_+$$

$$\delta \in (0,1]$$

Based on the recovered (γ,δ) that solve this minimizition problem, we can recover the natural probabilities from (36).

C Appendix: Computing State Prices Empirically

Before we can recover probabilities, we need to know the Arrow-Debreu prices or, said differently, characterize the risk-neutral distribution. There exist many ways to do this in practice based on observed option prices, including various interpolation methods. We implement two methods; (i) the parametric stochastic volatility model of Bates (2000) and (ii) the non-parametric "Fast and Stable" method of Jackwerth (2004).

C.1 The Bates (2000) stochastic volatility model with jumps

To ensure that we start with an arbitrage-free collection of Arrow-Debreu prices by strike and maturity, we use the model of Bates (2000) to derive state prices from observed option prices. This parametric approach puts structure on the tails of the risk-neutral density, which also allows us to extrapolate outside the range of observable option quotes. While the Bates (2000) model may not be the "true" specification of the economy, we simply use this framework as a standard method in the literature to compute state prices, and, consistent with this pragmatic view, we allow parameters to change over time (which also avoids look-ahead bias).

In this model, the risk-neutral process for the price of the underlying asset, S_t, and the instantaneous variance, V_t, are assumed to be of the form

$$dS_t/S_t = (r^f - d - \lambda\bar{k})dt + \sqrt{V_t}dZ_t + kdq_t \tag{C.1}$$

$$dV_t = (\alpha - \beta V_t)dt + \sigma_v\sqrt{V_t}dZ_{vt} \tag{C.2}$$

where Z_t and Z_{vt} are Brownian motions with correlation ρ, and q_t is a Poisson counting process that captures the risk of jumps in the price. The jumps occur with intensity λ and each jump causes the price to be multiplied by the factor $1 + k$, which is lognormally distributed, i.e., $\ln(1 + k) \sim N(\ln(1 + \bar{k})\frac{1}{2}\delta^2, \delta^2)$. Further, r^f is the risk-free rate and d is the dividend yield.

We calibrate these model parameters every fourth Wednesday as follows:[14] On each day, given the current level of the market S_t and the risk-free term structure $r^f_{t,t+\tau}$, we find the model parameters $(\alpha, \beta, \lambda, \bar{k}, \sigma_v, \delta)$ and state variable V_t that minimize the vega-weighted squared pricing errors for fifty call and put options, following the methodology of Trolle and Schwartz (2009). The fifty chosen call/put options are those with the highest volumes. We allow the model parameters to vary over time since we simply use the model to smooth observed option prices (that may be noisy)

[14] We use data for every fourth Wednesday as a compromise between (i) the tradition in the asset pricing literature on return predictability of focusing on monthly returns, and (ii) the tradition in the option literature of focusing on Wednesdays, where among other reasons option liquidity is high.

such that they are arbitrage-free.

Once we have obtained model estimates, we compute the risk-neutral density $f(\tau, S_\tau)$ for any time τ periods into the future and state S_τ given the current time state S_t as:

$$f(\tau; S_\tau) = \frac{1}{\pi} \int_0^\infty \left(\frac{S_\tau}{S_t} \right)^{-iu} \psi(\tau, u) du \tag{C.3}$$

that is, by integrating the characteristic function ψ numerically using the Gauss-Laguerre quadrature method. Knowing the risk-neutral density, the corresponding state price density $\pi(\tau; S_T)$ is the density discounted by the τ-period risk-free rate $r^f_{t,t+\tau}$:

$$\pi(\tau; S_\tau) = e^{-r^f_{t,t+\tau}} f(T; S_\tau) \tag{C.4}$$

This completes the computation of state prices. Indeed, we think of $\pi(\tau; S_\tau)$ as the Arrow-Debreu prices we need as starting point for our recovery for each index level. For example $\pi(1, 2000)$ is the Arrow-Debreu price of receiving \$1 in one year of the S&P500 is between 2000 and 2001. We consider the grid of maturities and index levels described in Section 7.2.

C.2 The Jackwerth (2004) "Fast and Stable" method

We are interested in converting a (noisy) sparse set of implied volatilities into a full risk-neutral distribution. In section C.1 we imposed a parametric form on the implied volatility surface through a stochastic volatility model with jumps. In this section we refrain from imposing any structure on implied volatilities, that is, we fit a non-parametric method to implied volatilities. The method we have chosen is the "Fast and Stable" method of Jackwerth (2004). This method has a single tuning parameter, λ, which simultaneously controls the smoothness of the function and the fit to observed implied volatilities. Clearly, there is a trade-off in choosing the value of the tuning parameter, which is: the smoother the function the worse the fit to observations. We therefore control the smoothness of the fit by imposing two conditions; (i) the estimated implied volatilities gives rise to a non-negative risk-neutral distribution, (ii) the risk-neutral distribution is unimodal in the range from 0.8 to 1.2 in moneyness (defined as S_t/S_0, the index level at time t relative to the current index level). Under these conditions we minimize the objective function:

$$\min_{\sigma_s} \frac{1}{2(S+1)} \sum_{s=1}^{S} \left(\sigma_s^{''}\right)^2 + \frac{\lambda}{2I} \sum_{i=1}^{I} (\sigma_i - \bar{\sigma}_i)^2 \tag{C.5}$$

Where S is the number of states. σ_s is the implied volatility associated with state s. $\sigma_s^{''}$ is the second derivative of the implied volatility function with respect to strike prices. $i = 1, ..., I$ is the index for the observed implied volatilities and $\bar{\sigma}_i$ is the i'th observed implied volatility. As seen from (C.5), if λ is high then the fit to observations will be good compared to when λ is low. We therefore choose the highest value of λ which satisfies our two conditions described above. See Jackwerth (2004) for further comments on the method.

Once a smooth function for the implied volatilities is obtained we can back out a risk-neutral distribution by evaluating the Black and Scholes (1973) formula in the estimated implied volatilities and then differentiate the resulting call function twice with respect to strike prices as explained in Breeden and Litzenberger (1978).

The Fast and Stable method estimates a single option maturity at a time. In the period from January 1996 until December 2015 we have at least 7 maturities on any given last trading day of the month. In the framework of Proposition 5 this allows us to parameterize the pricing kernel with up to 6 parameters and still obtain generalized recovery.

References

Ait-Sahalia, Yacine, and Andrew W. Lo, 2000, Nonparametric risk management and implied risk aversion, *Journal of Econometrics* 94, 9–51.

Audrino, Francesco, Robert Huitema, and Markus Ludwig, 2014, An empirical analysis of the Ross recovery theorem, *Working paper.*

Bakshi, Gurdip, Fousseni Chabi-Yo, and Xiaohui Gao, 2015, A recovery that we can trust? deducing and testing the restrictions of the recovery theorem, *Working paper.*

Bates, D. S., 2000, Post 87 crash fears in the S&P 500 futures options market, *Journal of Econometrics* 94, 181–238.

Berkowitz, Jeremy, 2001, Testing density forecats, with applications to risk management, *Journal of Business and Economic Statistics* 19, 465–474.

Black, F., and M. S. Scholes, 1973, The pricing of options and corporate liabilities, *The Journal of Political Economy* 81, 637–654.

Bliss, Robert R., and Nikolaos Panigirtzoglou, 2004, Option implied risk aversion estimates, *Journal of Finance* 1, 407–446.

Bollerslev, Tim, and Viktor Todorov, 2011, Tails, fears, and risk premia, *Journal of Finance* 66, 2165–2211.

Borovicka, Jaroslav, Lars Peter Hansen, and Jose A. Scheinkman, 2016, Misspecified recovery, *Journal of Finance* 71, 2493–2544.

Breeden, Douglas T, and Robert H Litzenberger, 1978, Prices of state-contingent claims implicit in option prices, *Journal of business* pp. 621–651.

Carr, Peter, and Jiming Yu, 2012, Risk, return, and Ross recovery, *Journal of Derivatives* 20, 38–59.

Christoffersen, Peter, Steven L. Heston, and Kris Jacobs, 2013, Capturing option anomalities with a variance-dependent pricing kernel, *Review of Financial Studies* 26, 1963–2006.

Cox, John C, Stephen A Ross, and Mark Rubinstein, 1979, Option pricing: A simplified approach, *Journal of Financial Economics* 7, 229–263.

Debreu, Gerard, 1970, Economies with a finite set of equilibria, *Econometrica* 38, 387–392.

Hansen, Lars Peter, and Jose A. Scheinkman, 2009, Long term risk: An operator approach, *Econometrica* 77, 177–234.

Jackwerth, J., 2000, Recovering risk aversion from option prices and realized returns, *Review of Financial Studies* 13, 433–451.

Jackwerth, Jens C., 2004, Option-implied risk-neutral distributions and risk aversion, *CFA Institute Research Foundation of AIMR Publications* pp. 1–86.

Jackwerth, Jens C., and Mark Rubinstein, 1996, Recovering probability distributions from option prices, *Journal of Finance* 51, 1611–1631.

Linetsky, Vadim, and Likuan Qin, 2016, Positive eigenfunctions of markovian pricing operators: Hansen-Scheinkman factorization, Ross recovery and long-term pricing, *Operations Research* 64, 99–117.

Malamud, Semyon, 2016, Noisy recovery, *Working paper.*

Martin, Ian, 2017, What is the expected return on the market?, *Quarterly Journal of Economics, Forthcoming.*

Martin, Ian, and Steve Ross, 2013, The long bond, *Working paper.*

Mehra, Rajnish, and Edward C Prescott, 1985, The equity premium: A puzzle, *Journal of monetary Economics* 15, 145–161.

Qin, Likuan, and Vadim Linetsky, 2017, Long term risk: A martingale approach, *Econometrica* 85, 299–312.

Rosenberg, Joshua V., and Robert F. Engle, 2002, Empirical pricing kernels, *Journal of Financial Economics* 64, 341–372.

Ross, Steve, 2015, The recovery theorem, *The Journal of Finance* 70, 615–648.

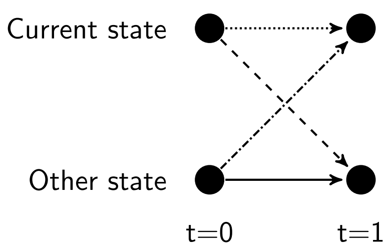
Schneider, Paul, and Fabio Trojani, 2015, (almost) model-free recovery, *Working paper.*

Trolle, Anders B., and Eduardo S. Schwartz, 2009, Unspanned stochastic volatility and the pricing of commodity derivatives, *Review of Financial Studies* 22, 4423–4461.

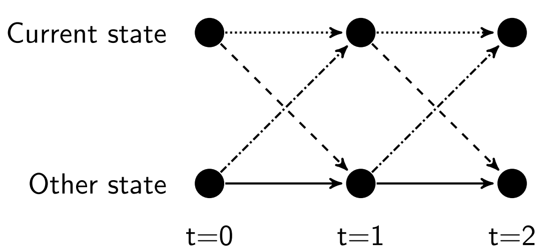
Walden, Johan, 2017, Recovery with unbounded diffusion processes, *Review of Finance, Forthcoming.*

D Tables and Figures

Panel A. Ross's Recovery Theorem: one period, two "parallel universes"



Panel B. Ross's Recovery Theorem: time-homogeneous dynamic setting



Panel C. Our Generalized Recovery: No assumptions about probabilities

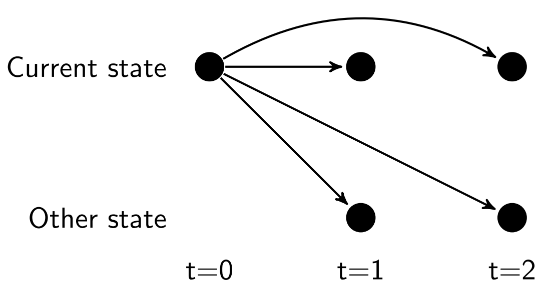
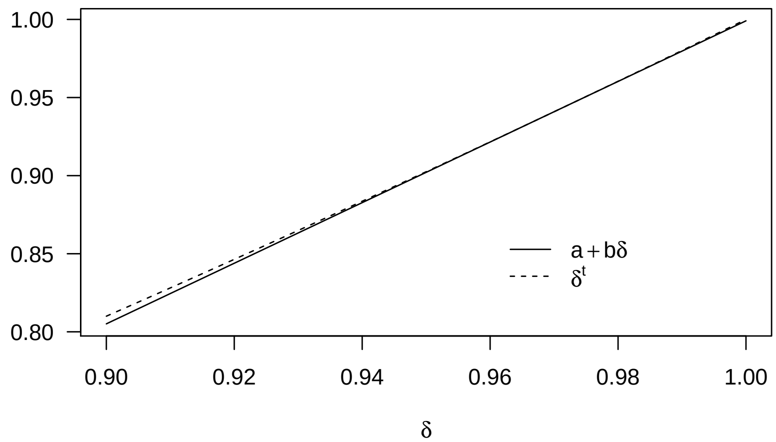
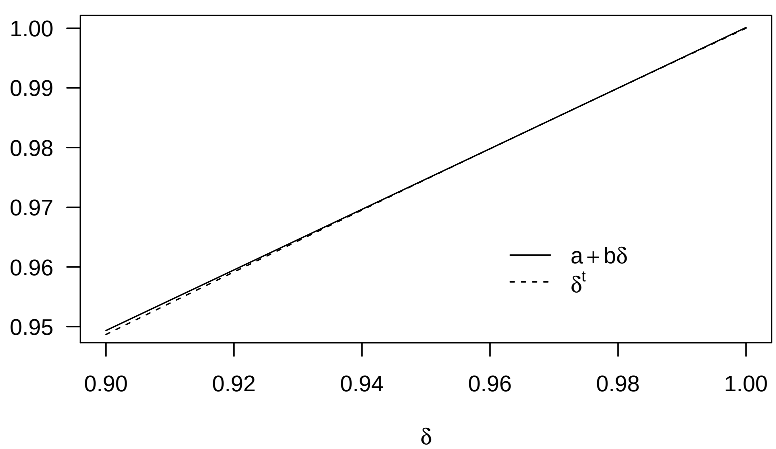


Figure 1: **Generalized Recovery Framework.** Panel A illustrates the idea behind Ross's Recovery Theorem, namely that we start with information about all Arrow-Debreu prices in *all* initial states (not just the state we are currently in, but also prices in "parallel universes" where today's state is different). Panel B shows how Ross moves to a dynamic setting by assuming time-homogeneity, that is, assuming that the prices and probabilities are the same for the two dotted lines, and so on for each of the other pairs of lines. Panel C illustrates our Generalized Recovery method, where we make no assumptions about the probabilities.



Panel A: $t = 2$ years

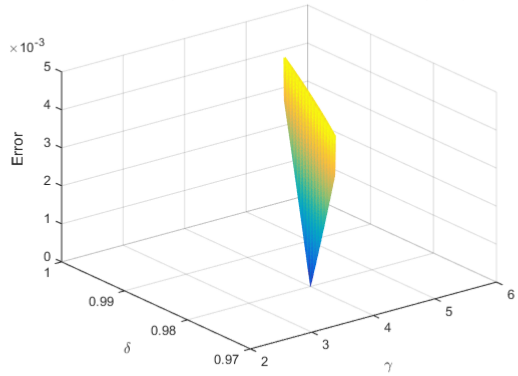


Panel B: $t = 0.5$ years

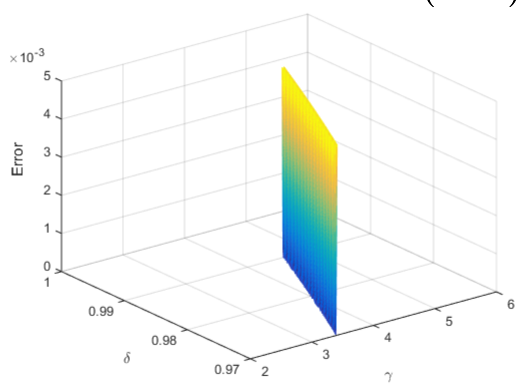
Figure 2: **Closed-Form Solution: Approximation Error.** The figure shows that the generalized recovery problem is very close to being linear. We show that the only non-linearity comes from the discount rate δ due to the powers of time, δ^t. However, the function $\delta \to \delta^t$ is very close to being linear for the relevant range of annual discount rates, say $\delta \in [0.94, 1]$, and the relevant time periods that we study. Panel A plots the discount function and the linear approximation around $\delta_0 = 0.97$ given a horizon of $t = 2$ years. Panel B plots the same for a horizon of a half year.

Table 1: **Correlation Matrix.** This tables shows the pairwise correlations between the recovered conditional expected excess return for different specifications of marginal utilities and method for estimating risk-neutral prices; (i) $\mu_{t,1}$: Bates and polynomial, (ii) $\mu_{t,2}$: Bates and piecewise linear, (iii) $\mu_{t,3}$: Jackwerth and polynomial, (iv) $\mu_{t,4}$: Jackwerth and piecewise linear. We augment the table with pairwise correlations with the VIX_t index and the lower boundary on the equity premium, SVIX_t, due to Martin (2017).

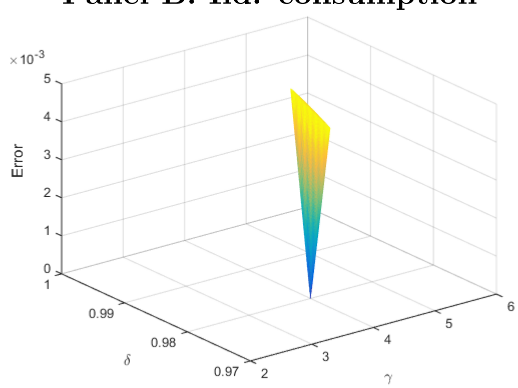
	$\mu_{t,1}$	$\mu_{t,2}$	$\mu_{t,3}$	$\mu_{t,4}$	VIX_t	SVIX_t
$\mu_{t,1}$	1	0.359	0.393	0.392	0.534	0.485
$\mu_{t,2}$		1	0.642	0.523	0.716	0.794
$\mu_{t,3}$			1	0.642	0.784	0.830
$\mu_{t,4}$				1	0.634	0.689
VIX_t					1	0.928
SVIX_t						1



Panel A: Mehra Prescott (1985)



Panel B: Iid. consumption



Panel C: Non-Markovian

Figure 3: **Generalized Recovery: Objective Function in Specific Economic Models.** This figure shows the objective function used for the generalized recovery method, the squared pricing errors in (B.3). Panel A shows that the objective function for the Mehra Prescott (1985) model has a unique minimum, making the generalized recovery feasible. Panel B shows that generalized recovery is not feasible in the Black-Scholes-Merton model with iid. consumption as the objective has a continuum of solutions. Panel C shows that generalized recovery is feasible in the non-Markovian model.

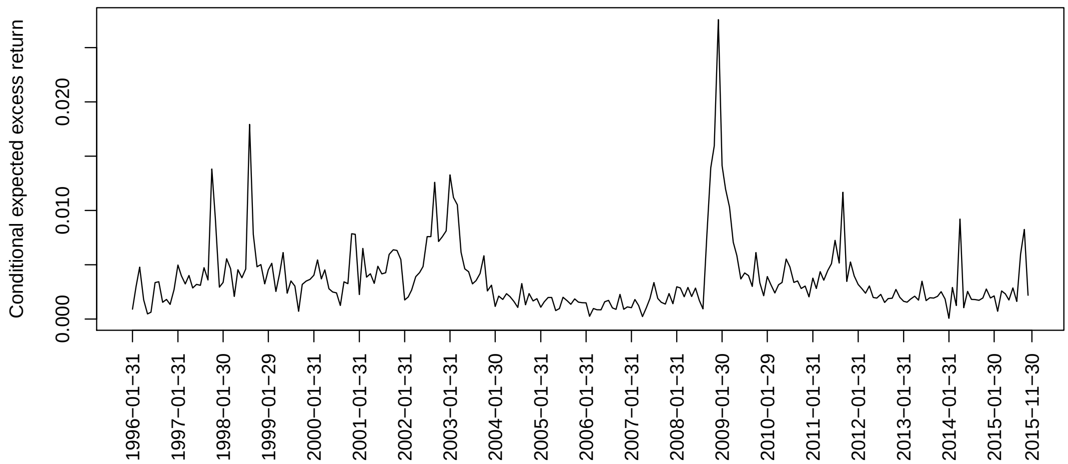


Figure 4: **Recovered conditional expected excess return.** The figure plots monthly conditional expected excess market returns, recovered last trading day of each month from 1/1996 to 12/2015. Marginal utilities are piecewise linear and risk-neutral prices are estimated using Jackwerth (2004).

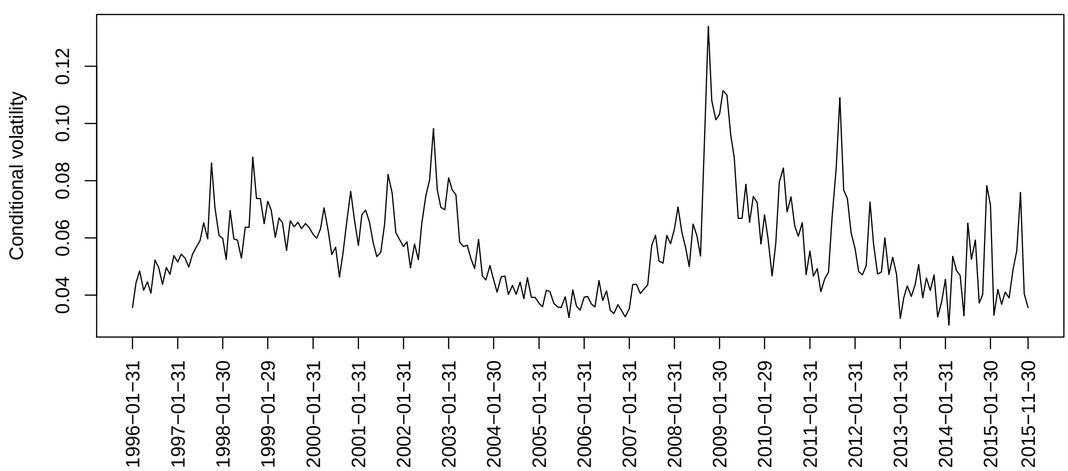


Figure 5: **Recovered conditional volatility of excess return.** The figure plots monthly conditional market volatility, recovered last trading day of each month from 1/1996 to 12/2015. Marginal utilities are piecewise linear and risk-neutral prices are estimated using Jackwerth (2004).

Table 2: **Does the Recovered Expected Return Predict the Future Return?** This table reports results of the regression of the ex post realized excess return r_{t+1} on the ex ante recovered expected excess return, μ_t, the ex post innovation in expected return, $\Delta\mu_{t+1}$, and ex ante the SVIX index of Martin (2017):

$$r_{t,t+1} = \beta_0 + \beta_1\mu_t + \beta_2\Delta\mu_{t+1} + \beta_3\text{SVIX}_t + \beta_4\text{VIX}_t + \epsilon_{t,t+1}$$

The regression uses monthly data over the full sample 1/1996–12/2015, t-statistics are reported in parentheses, and significance at a 10% level is indicated in bold.

Dependent variable	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Intercept	-0.00	**0.01**	0.01	0.00	**0.01**	0.01
	(-0.06)	(2.34)	(1.64)	(1.46)	(2.05)	(1.78)
μ_t	1.23	-2.95	-0.22	0.07	-0.00	0.18
	(1.55)	(-1.25)	(-0.18)	(0.09)	(-1.25)	(0.28)
$\Delta\mu_{t+1}$	**-3.66**	**-22.07**	**-14.00**	**-7.83**	**-0.55**	**-16.11**
	(-3.80)	(-7.65)	(-7.93)	(-7.32)	(-10.1)	(-16.01)
Adj. R^2 (%)	5.9	18.1	20.4	17.8	30.0	51.7
Method:						
Expected excess return (μ_t)	Recovered	Recovered	Recovered	Recovered	VIX	SVIX
Q-prices	Bates	Bates	Jackwerth	Jackwerth		
Pricing kernel	Polynomial	Piecewise linear	Polynomial	Piecewise linear		

Table 3: **Does the Recovered Expected Return Predict the Future Return - Excluding 8/2008-7/2009**
This table reports results of the regression of the ex post realized excess return r_{t+1} on the ex ante recovered expected excess return, μ_t, the ex post innovation in expected return, $\Delta\mu_{t+1}$, and ex ante the SVIX index of Martin (2017):

$$r_{t,t+1} = \beta_0 + \beta_1\mu_t + \beta_2\Delta\mu_{t+1} + \beta_3\text{SVIX}_t + \beta_4\text{VIX}_t + \epsilon_{t,t+1}$$

The regression uses monthly data over the full sample 1/1996–12/2015, t-statistics are reported in parentheses, and significance at a 10% level is indicated in bold.

Dependent variable	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$	$r_{t,t+1}$
Intercept	0.00	0.01	-0.00	0.00	0.00	0.00
	(0.06)	(0.98)	(-0.15)	(1.20)	(1.11)	(0.12)
μ_t	**1.37**	0.28	**3.06**	**1.65**	0.00	**1.71**
	(1.79)	(0.07)	(1.82)	(1.72)	(-0.25)	(1.99)
$\Delta\mu_{t+1}$	**-3.04**	**-26.98**	**-12.74**	**-9.00**	**-0.50**	**-17.69**
	(-3.30)	(-7.15)	(-6.29)	(-7.93)	(-8.75)	(-15.53)
Adj. R^2 (%)	5.0	18.1	16.4	23.1	24.6	52.5
Method:						
Expected excess return (μ_t)	Recovered	Recovered	Recovered	Recovered	VIX	SVIX
Q-prices	Bates	Bates	Jackwerth	Jackwerth		
Pricing kernel	Polynomial	Piecewise linear	Polynomial	Piecewise linear		

Table 4: **Does the Recovered Volatility Predict the Future Volatility?** This table reports results of a monthly regression of the ex post realized volatility on the ex ante recovered return volatility, σ_t, and the VIX volatility index:

$$\sqrt{var(r_{t,t+1})} = \beta_0 + \beta_1 \sigma_t + \beta_2 \mathrm{VIX}_t + \epsilon_{t,t+1}$$

The regression uses monthly data over the full sample 1/1996–12/2015, t-statistics are reported in parentheses, and significance at a 10% level is indicated in bold.

Dependent variable	$\sqrt{var(r_{t,t+1})}$	$\sqrt{var(r_{t,t+1})}$	$\sqrt{var(r_{t,t+1})}$	$\sqrt{var(r_{t,t+1})}$	$\sqrt{var(r_{t,t+1})}$
Intercept	-0.00	-0.00	**-0.01**	**-0.01**	**-0.05**
	(-1.45)	(-1.40)	(-2.41)	(-2.83)	(-9.63)
σ_t	**0.89**	**0.86**	**0.95**	**0.98**	**0.71**
	(16.7)	(16.8)	(14.67)	(14.63)	(17.19)
Adj. R^2 (%)	54.0	54.0	47.4	47.3	55.3
Method:					
Volatility (σ_t)	Recovered	Recovered	Recovered	Recovered	VIX
Q-prices	Bates	Bates	Jackwerth	Jackwerth	
Pricing kernel	Polynomial	Piecewise linear	Polynomial	Piecewise linear	