EXCESS VOLATILITY: BEYOND DISCOUNT RATES[*]

STEFANO GIGLIO AND BRYAN KELLY

We document a form of excess volatility that is difficult to reconcile with standard models of prices, even after accounting for variation in discount rates. We compare prices of claims on the same cash flow stream but with different maturities. Standard models impose precise internal consistency conditions on the joint behavior of long- and short-maturity claims and these are strongly rejected in the data. In particular, long-maturity prices are significantly more variable than justified by the behavior at short maturities. We reject internal consistency conditions in all term structures that we study, including equity options, currency options, credit default swaps, commodity futures, variance swaps, and inflation swaps. *JEL Codes:* G12, G40.

I. INTRODUCTION

Term structure analysis is a powerful setting for evaluating a model's ability to describe asset price data for two reasons. First, any model that satisfies a minimal requirement—that it rules out arbitrage opportunities—imposes strict testable restrictions on the joint behavior of prices along the term structure. Specifically, no-arbitrage prices must obey the law of iterated values, as the prices of long-maturity claims must reflect investors' expectations about the future value of short-maturity claims.[1] This places tight bounds on the extent of covariation between prices at different maturities that is admissible within a given model. Too

1. For seminal work on the role of cross-equation restrictions and the law of iterated values in rational models, see Samuelson (1965), Hansen and Sargent (1980), Hansen and Richard (1987), Anderson, Hansen, and Sargent (2003), Hansen and Scheinkman (2009), and Hansen (2012).

much (or too little) covariation between long- and short-maturity claim prices can rule out a model as a viable descriptor of the economy. Second, term structure data are unique in economics in how accurately they are described with parsimonious models[2] and are thus ideal proving grounds for discriminating between alternative models.

In this article, we document a form of excess volatility in prices along the term structure that is difficult to reconcile with "standard" asset-pricing models. Our central finding is that price fluctuations at different points in the term structure are internally inconsistent with each other—prices on the long end of the term structure are far more variable than justified by the behavior of short-end prices—given usual modeling assumptions. The consistency violations are highly significant statistically and economically. Perhaps most interesting, excess volatility of long-maturity prices is evident in a large number of asset classes, including claims to equity and currency volatility, sovereign and corporate credit risk, Treasury yields, commodities, and inflation.

We define as "standard" any model in which cash flows and asset prices are linear functions of common factors. This type of model is pervasive in financial economics because of its convenience in delivering closed-form pricing solutions in a wide range of valuation problems. This encompasses many leading asset-pricing paradigms, from structural equilibrium models[3] with long-run risks (Bansal and Yaron 2004) or variable rare disasters (Wachter 2013) to reduced-form models ubiquitous in fixed income and derivatives pricing (Duffie, Pan, and Singleton 2000).

We refer to this class of models as "affine-$\mathbb{Q}$" following terminology in the asset-pricing literature. In the canonical model, an asset's "physical," or "$\mathbb{P}$," distribution of payoffs is determined by factors with linear time series dynamics. Investor preferences can be represented as a subjective adjustment to the payoff distribution. This preference-adjusted payoff distribution is known variously as the "pricing," "risk-neutral," or "$\mathbb{Q}$" measure. It has the special property that prices are equal to $\mathbb{Q}$-expectations of cash flows discounted at the risk-free rate. Furthermore, any risk adjustment that investors apply when valuing a stream of cash flows operates through the $\mathbb{Q}$ measure. Affine-$\mathbb{Q}$ models choose

---

2. For example, a linear three-factor model explains the panel of Treasury yields for maturities of 1 year up to 30 years with an R^2 in excess of 99%.

3. We discuss affine structural models in Online Appendix A.

preferences so that payoffs retain their linearity in factors under $\mathbb{Q}$, and in turn equilibrium prices are also linear in the factors.

There is an important advantage in working directly with the $\mathbb{Q}$ representation of asset prices when studying the term structure. Because it integrates investor risk preferences into its description of the economy, a model's $\mathbb{Q}$ representation summarizes any variation in discount rates that may influence asset price behavior.[4] Therefore, any inferences regarding price volatility that are based on a model's $\mathbb{Q}$ representation take investors' discount-rate behavior into account.

This contrasts with the notion of excess volatility famously documented by Shiller (1979, 1981) in which price fluctuations are deemed excessive relative to predictions from a specific model—one with constant discount rates. A potential resolution of the Shiller puzzle is to recognize that discount rates are variable, an insight at the foundation of leading frameworks in modern finance.[5] By benchmarking against the $\mathbb{Q}$ representation of models, any excessive volatility we document must arise from sources other than discount rate variation. In short, we analyze the affine-$\mathbb{Q}$ class as a null model for our analysis because it explicitly accounts for what has become the de facto explanation for excess volatility, time-varying discount rates.

I.A. A One-Factor Example

Our main empirical finding is that in every asset class that we analyze, long-maturity prices overreact to short-maturity price fluctuations relative to the predictions of an affine-$\mathbb{Q}$ model. A simple example illustrates the nature of this overreaction.

Consider a term structure of claims to the one-factor cash flow process x_t. For concreteness, think of x_t as the realized variance of the aggregate stock market return during period t, and consider valuing a derivative contract whose payoff is determined by x_t.

---

4. More specifically, the $\mathbb{Q}$ measure incorporates variation in risk premia, which is the primary driver of total discount rate variation. Throughout we refer to discount rates and risk premia interchangeably.

5. See, for example, Campbell and Shiller (1987, 1988a,b, 1991), Fama and Bliss (1987), Campbell (1987, 1991, 1995), Cochrane (1992, 2008, 2011), and Cochrane and Piazzesi (2009).

Under the pricing measure $\mathbb{Q}$, cash flows evolve according to[6]

$$x_t = \rho^{\mathbb{Q}} x_{t-1} + \epsilon_t^{\mathbb{Q}}.$$

We abstract from constants and risk-free rate adjustments in this example in the interest of simplicity. The price of a n-maturity forward claim on these cash flows is

$$(1) \qquad\qquad f_{t,n} = E_t^{\mathbb{Q}}[x_{t+n}].$$

The term structure of forward prices at maturities $1, \ldots, N$ is therefore given by

$$(2) \qquad f_{t,1} = \rho^{\mathbb{Q}} x_t, \quad f_{t,2} = (\rho^{\mathbb{Q}})^2 x_t, \quad \ldots, \quad f_{t,N} = (\rho^{\mathbb{Q}})^N x_t.$$

The key cross-equation restrictions in this model require that the term structure of prices obeys a strict one-factor structure, and that the only admissible shape for the price curve is one in which the factor loadings follow a geometric progression in $\rho^{\mathbb{Q}}$ (the parameter governing cash flow dynamics under $\mathbb{Q}$). This restriction is equivalently represented with prices of cumulative claims, defined as $p_{t,n} = E_t^{\mathbb{Q}}[x_{t+1} + \ldots + x_{t+n}]$, in which case the term structure takes the form:

$$p_{t,n} = (\rho^{\mathbb{Q}} + (\rho^{\mathbb{Q}})^2 + \ldots + (\rho^{\mathbb{Q}})^N) x_t.$$

Tests of the model's restrictions hinge on an estimate of $\rho^{\mathbb{Q}}$. Fortunately, $\rho^{\mathbb{Q}}$ is easily estimated from regressions of prices onto prices. For example, let the first maturity forward price, $f_{t,1}$, stand in for the latent factor x_t. Let b_2 denote the (population) slope coefficient in a regression of the price at maturity two, $f_{t,2}$, on $f_{t,1}$. According to equation (2), b_2 exactly recovers $\rho^{\mathbb{Q}}$. This regression is intuitive. The relative valuation of the first two claims reveals the cash flow persistence that investors perceive. If investors price assets as though x_t is very persistent, a rise in the short price $f_{t,1}$ will coincide with a rise in $f_{t,2}$ of nearly the same magnitude,

---

6. Autoregressive models for variance are standard in the time series and derivatives pricing literature. See for example Andersen et al. (2003) and our discussion of variance swaps in Section III. The shock, $\epsilon_t^{\mathbb{Q}}$, is orthogonal to x_{t-1} and mean 0, but is otherwise general. For example, $\epsilon_t^{\mathbb{Q}}$ may possess a conditional distribution that ensures x_t is nonnegative, as in standard stochastic volatility models.

which indicates that $\rho^{\mathbb{Q}}$ is near one under the investors' subjective pricing measure.

If we project prices for remaining maturities $3, \ldots, N$ onto the short-maturity price $f_{t,1}$, we recover a sequence of regression coefficients denoted $b_3, \ldots, b_N$ that are unrestricted in the sense that they are not forced to be jointly determined by $\rho^{\mathbb{Q}}$ as would be implied by equation (2). At the same time, these regressions can be recast in their "restricted" form, where the restriction in equation (2) relates, for example, b_N to b_2 by:

$$(3) \qquad\qquad b_N = (b_2)^{N-1}.$$

We convert this restriction into a test of excess volatility by constructing a variance ratio statistic for each maturity N:

$$VR_N = \frac{Var(b_N f_{t,1})}{Var((b_2)^{N-1} f_{t,1})}.$$

The numerator, $Var(b_N f_{t,1})$, is the explained variance in the unrestricted regression of long-end prices ($f_{t,N}$) onto the short end ($f_{t,1}$). The denominator, $Var((b_2)^{N-1} f_{t,1})$, is the explained variance of the same regression under restriction (3).[7] Under the null model, the restricted and unrestricted variances are the same and $VR_N = 1$. If the ratio statistic significantly exceeds 1, the price at maturity N varies more than is justified by the behavior of the short end of the term structure. The same variance ratio test can be applied to cumulative claims as well.

This one-factor example is intentionally simplified to illustrate our approach for testing excess volatility along the term structure. In Section II, we develop an estimation and inference approach for VR_N in general K-factor affine specifications that is implementable with OLS regressions. We also develop an approach that uses the Kalman filter and maximum likelihood to build variance ratio tests that are robust to measurement error in term structure prices (for example, due to illiquidity).

I.B. A Representative Term Structure

Figure I illustrates the behavior of variance ratios in one of our data sets—the term structure of variance swaps—which are

---

7. Note that VR_N is simply the squared ratio of the unrestricted regression coefficient to the restricted coefficient.
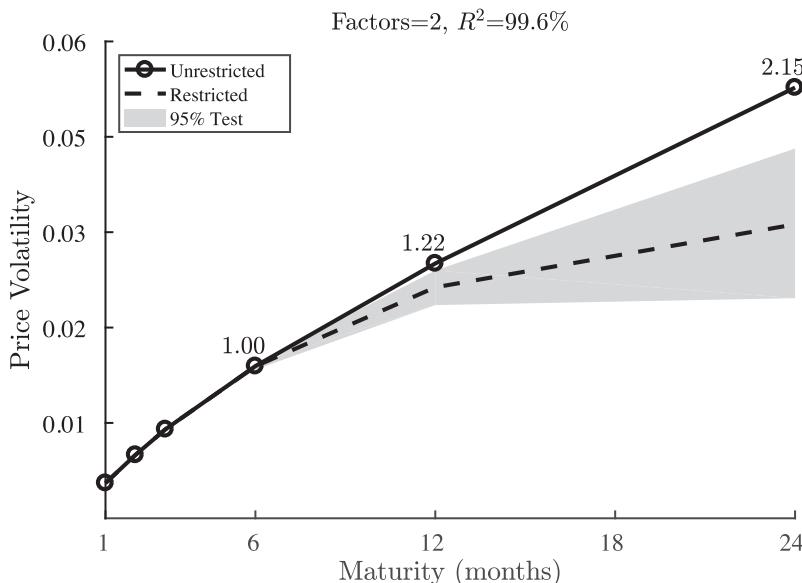
Variance Swap Tests

The figure plots the standard deviation of prices under the unrestricted factor model (solid line) and under the restricted model (dashed line). The circles in the unrestricted line represent the maturities we observe in the data. Numbers adjacent to circles are the variance ratios at each maturity. The shaded area represents the 2.5th to 97.5th percentiles of the model-implied variance in bootstrap simulations.

claims to the cumulative variance of the S&P 500 index over the life of the contract.[8] An unrestricted linear two-factor model provides an excellent description of the term structure, delivering an R^2 of 99.6% for the panel of prices. The solid black line plots the explained swap price volatility from an unrestricted regression of each long-maturity claim on the first two short-maturity claims. The dashed line plots the explained variation from the regression that imposes the model restrictions. The variance ratio statistic for each maturity is printed above the unrestricted volatility estimates and the shaded region represents the pointwise 95% bootstrap confidence interval for price volatility in the restricted model.

8. These data are described in detail in Section III.

At 24 months, the variance ratio statistic reaches 2.15, meaning that the variability in long-maturity prices is more than twice as large as that allowed by the affine-model restriction and is highly statistically significant. The high variance ratio can be thought of in the following way. The concave shape of price volatility at the short end of the curve suggests that cash flows mean revert fairly quickly under $\mathbb{Q}$. But this appears inconsistent with indications of much higher persistence implied from the long end. As a result, unrestricted price volatility increases with maturity at a much faster rate than the price volatility predicted by the model. The high variance ratio indicates that prices at the long end of the curve react to the short end much more strongly in the data than affine-model dynamics allow.[9]

The excess volatility of long-maturity claims is not explained by movements in discount rates. Discount rate variation that is describable within the affine class is subsumed by our model. Nor do high variance ratios merely reflect a poor fit from the factor model. The R^2 from the unrestricted factor specification is nearly 100% in all of our term structures, meaning that an unconstrained linear model does an excellent job describing the data. Instead, the high variance ratio is a violation of the cross-equation restrictions of the affine model. That is, the data are exceedingly well described by a linear factor model, but with factor loadings that differ from those implied by model restrictions.

Behavior of the variance swap term structure is representative of our broader empirical findings. All of the asset classes we study exhibit excess volatility of long-maturity prices similar to that in Figure I.

I.C. Potential Explanations

Tests of excess volatility are fundamentally tests of market efficiency and are therefore subject to the joint-hypothesis problem described by Fama (1970, 1991):

> Market efficiency per se is not testable. It must be tested jointly with some model of equilibrium, an asset-pricing model. . . . As a result, when we find anomalous evidence on the behavior of returns, the

9. We interpret high variance ratios as excess volatility at long maturities, rather than a dearth of volatility at short maturities, because short-maturity prices appear appropriately anchored to (with nearly identical volatility as) the underlying physical cash flow. For example, short-dated variance swaps and inflation swaps closely track realized variance and CPI inflation, respectively.

way it should be split between market inefficiency or a bad model of market equilibrium is ambiguous.

In the last part of the article, we investigate how the sources of excess volatility should be "split between market inefficiency," that is, mispricing along the term structure, "or a bad model of market equilibrium" in the form of model misspecification. While it is impossible to draw unambiguous conclusions or to exhaust the list of possible explanations, analyzing leading candidates helps refine our basic facts. In Section IV, we examine five potential explanations for our findings: omitted factors, nonlinear dynamics, long-memory dynamics, measurement error, and temporary mispricing of long-maturity claims.

First, if the true data-generating process is a K-factor affine model but we use fewer than K factors in our analysis, the variance ratio statistic is likely to diverge significantly from 1. In robustness checks, we show that gradually increasing the number of factors (thereby pushing the factor model R^2 even closer to 100%) still produces variance ratios significantly in excess of 1.

Second, we explore a wide range of long-memory models in the stationary ARFIMA family. These models can exhibit persistence that decays much more slowly than the autoregressive structure assumed in affine-$\mathbb{Q}$ specifications. The vast majority of ARFIMA specifications appear well approximated by simple affine models and do not lead to high variance ratios. However, as the long-memory parameter reaches the boundary of the nonstationary range, we show that it is possible to generate variance ratios as high as 3 at the 24-month maturity. But when we allow for an extra factor, the variance ratios again shrink to 1, which is inconsistent with the behavior we find in the data.

Third, we explore a large class of nonlinear dynamic specifications known as smooth-transitioning autoregressive (STAR) models. In most parameterizations, STAR models are very closely approximated by a low-dimension affine model and therefore do not produce variance ratios above 1. For the most extreme nonlinear specifications it is possible to generate variance ratios that statistically reject the affine restrictions, but even in these cases the variance ratios are substantially smaller than those found in the data.

Fourth, we evaluate the role of measurement error in our empirical tests. We conduct a variety of robustness tests establishing

that measurement error is a quantitatively unviable explanation of our findings.

Finally, we explore the possibility of mispricing as a potential driver of excess volatility in two ways. First, we construct a trading strategy to quantify the economic magnitude of the deviation from the affine-$\mathbb{Q}$ specification. It trades long-maturity claims when they are misvalued relative to the affine model and hedges with an offsetting short-maturity position in the exact proportion dictated by the estimated model. If violations of the affine model are small or infrequent, then the trading strategy will perform poorly in terms of risk-adjusted returns. But if the hypothesized mispricing exists, then the strategy may appear profitable even after adjusting for risk.[10]

In the variance swap market, we find that the trading strategy yields an annualized out-of-sample Sharpe ratio of 1.3 on average. We show that this performance is not explained by exposure to standard risk factors and discuss limits to arbitrage in the swap market that can lead these mispricings to persist (Shleifer and Vishny 1997). This is suggestive but not conclusive evidence of mispricing, as high average returns may represent compensation for some risk that we have not considered. In this case, the strategy's performance quantifies the economic importance of risk factors missed by affine-$\mathbb{Q}$ models.

Second, we explore theoretical underpinnings of excess volatility. To do so, we present a model of investor extrapolation in the natural expectations framework (Fuster, Laibson, and Mendel 2010). This framework posits that investors price assets by averaging two different expectations, one formed according to the true cash flow-generating process and another based on misspecified, extrapolative beliefs. We then derive the model's term structure implications. We show that long-maturity excess volatility is an inherent prediction of the natural expectations model and show that model averaging is the key mechanism for qualitatively matching our empirical facts. Finally, we calibrate the model to variance swap data and show that it provides a close quantitative match to the data. The ability of natural expectations to fit term

---

10. The existence of profitable trading opportunities is not a necessary condition for mispricing, in the sense that price is not equal to value. It is possible that mispricings exist yet there is too much noise for arbitrage. Despite noise in the data, we provide evidence of high compensation for trading on affine-model violations that supports our excess-volatility interpretation of the facts.

structure patterns is remarkable because the idea was originally conceived to target time series patterns for macro aggregates, not term structure prices.

I.D. Literature Review

Perhaps the most important predecessor of our article is the seminal contribution of Stein (1989), who compares the volatility of short- and long-maturity S&P 100 index options. He finds excess volatility of one-year option prices and interprets it as evidence of investor overreaction. Our article builds on Stein's original insight with a few key differences. First, he analyzes comovement of long- and short-maturity prices relative to cash flow persistence estimated from the $\mathbb{P}$ measure. In other words, the reference model of Stein (1989) does not account for discount rate variation, nor do the interest rate volatility tests of Shiller (1979) or the equity volatility tests of Shiller (1981) and LeRoy and Porter (1981). Our excess volatility test explicitly accounts for discount rate variation by estimating cash flow dynamics under the $\mathbb{Q}$ measure. In addition, Stein (1989) uses a one-factor model for volatility, whereas our approach allows for an arbitrary number of factors and extends to a wide range of asset classes. Our findings are also related to Gurkaynak, Sack, and Swanson (2005), who show that the responsiveness of long-run Treasury bond yields to macroeconomic announcements is excessive relative to established new Keynesian DSGE models. More recently, Hanson and Stein (2015) study overreaction at the long end of the Treasury yield curve focusing on FOMC announcement days. An interesting aspect of our work is that long-maturity excess volatility is even more pronounced in other asset classes.[11]

Our evidence further encourages efforts to reconcile models of investors' expectation formation with financial market

---

11. The Treasury yield curve is the subject of a large literature that works extensively with affine-$\mathbb{Q}$ specifications. For a review and recent contributions see, for example, Dai and Singleton (2002), Duffee (2002), Ang and Piazzesi (2003), Le, Singleton, and Dai (2010), Piazzesi (2010), and Joslin, Singleton, and Zhu (2011). Our focus is on volatility of prices at different maturities. A distinct literature studies risk premia along various term structures. Backus, Boyarchenko, and Chernov (2015) study a few of the term structures that we analyze. Van Binsbergen, Brandt, and Koijen (2012) and van Binsbergen et al. (2013) analyze risk premia of dividend strips. Giglio, Maggiori, and Stroebel (2015, 2016) study the term structure of risk premia in housing markets. Dividend strip and housing data do not have maturity structures rich enough for our analysis.

fluctuations.[12] Our trading strategy analysis in Section IV.E suggests there are large costs borne by investors who overreact due to extrapolative expectations or other belief distortions. Our analysis emphasizes that asset term structures, whose prices depend on how investors form expectations over multiple horizons, offer a fruitful setting for future behavioral research.

II. TERM STRUCTURE MODEL

In this section we develop and test the internal consistency restrictions implied by affine term structure models.

Our focus is on the joint price behavior of claims to an underlying cash flow process x_t that have different maturities. For most of our analysis, we focus on linear claims to the x_t process. We discuss the extension to linear exponential claims in Online Appendix B.

II.A. Claims with Linear Payoff Structures

At time t, a linear n-maturity forward claim promises a one-time stochastic cash flow of x_{t+n} to be paid in period $t+n$. Under the weak assumption that a model admits no arbitrage opportunities, there exists a pricing measure $\mathbb{Q}$ under which prices of such claims are expectations of future cash flows discounted at the risk-free interest rate. We assume such a measure exists, thus the n-maturity forward price is representable as

$$(4) \qquad f_{t,n} = E_t^{\mathbb{Q}} \left[x_{t+n} \frac{S_t}{S_{t+n}} \right],$$

where S_t is the value of a risk-free account that pays the short-term rate. In our empirical analysis, risk-free rate variation is negligible compared to risky asset price variation in almost all asset classes.[13] To reduce notation in the remainder of this section,

12. Recent examples of research into expectation formation related to our analysis include Cecchetti, Lam, and Mark (2000), Hansen (2014), Gennaioli, Shleifer, and Vishny (2015), Bordalo, Gennaioli, and Shleifer (2015), Barberis et al. (2015a,b), Glaeser and Nathanson (2015), and Hirshleifer, Li, and Yu (2015), among others.

13. The obvious exception is the Treasury bond market, in which case we account for risk-free rate variation explicitly using the standard model.

we assume that S_t is constant and equal to 1, and we provide a detailed discussion of risk-free rate considerations and associated robustness checks in Online Appendix C.

Prices of one-off forward claims aggregate into linear cumulative claims that promise a sequence of cash flows through maturity. The time-t price of an n-maturity cumulative claim is a sum of forward prices,

$$p_{t,n} = E_t^{\mathbb{Q}}[x_{t+1} + \ldots + x_{t+n}] = f_{t,1} + \ldots + f_{t,n}.$$

Under no arbitrage, the pricing measure possesses a martingale property that binds prices together across time and maturity,

$$f_{t,n} = E_t^{\mathbb{Q}}[f_{t+1,n-1}] \quad \text{and} \quad p_{t,n} = E_t^{\mathbb{Q}}[p_{t+1,n-1}] + f_{t,1},$$

which follows from the law of iterated expectations.

II.B. Affine-$\mathbb{Q}$ Model Setup

Our baseline model is defined by the following assumptions.

ASSUMPTION 1. The cash flow process, x_t, is a linear function of K latent factors, H_t,

$$(5) \qquad x_t = \delta_0 + \delta_1' H_t,$$

where δ_0 is a scalar and δ_1 is a $K \times 1$ vector.

ASSUMPTION 2. Under the $\mathbb{Q}$ measure, H_t evolves as

$$(6) \qquad H_t = c^{\mathbb{Q}} + \rho^{\mathbb{Q}} H_{t-1} + \Gamma \epsilon_t^{\mathbb{Q}},$$

where $\epsilon_t^{\mathbb{Q}}$ is a vector of uncorrelated mean-zero shocks that is orthogonal to the history of the system through $t - 1$, and $\Gamma \Gamma'$ is a positive-definite covariance matrix.

ASSUMPTION 3. The matrix $\rho^{\mathbb{Q}}$ is diagonal, $c^{\mathbb{Q}}$ is 0, and δ_1 is a vector of ones.

Assumptions 1 and 2 ensure that the price of all cash flow claims will be linear functions of H_t, since prices are determined as expectations of x_t under $\mathbb{Q}$. Assumption 2 emphasizes our article's

focus on the $\mathbb{Q}$ measure. In particular, our baseline model requires linear factor dynamics under $\mathbb{Q}$.[14]

Because the H_t factors are latent, the system specification in Assumptions 1 and 2 is identified only up to a linear transformation of the factors. Assumption 3 describes the parameter normalization needed to achieve econometric identification. This normalization imposes no economic restrictions but ensures that the model we bring to the data has exactly as many parameters as there are observables. For a detailed discussion of our normalization choices, see Joslin, Singleton, and Zhu (2011) and Hamilton and Wu (2012).

The price of a linear cumulative claim with maturity n is given by

$$(7) \qquad p_{t,n} = n\delta_0 + \mathbf{1}'(\rho^{\mathbb{Q}} + \ldots + (\rho^{\mathbb{Q}})^n)H_t + \nu_{t,n}.$$

This equation follows from Assumptions 1–3. In addition, we follow convention in the term structure literature and include a noise term ($\nu_{t,n}$) to account for potential measurement error in prices under the physical measure. Equation (7) constitutes a collection of cross-equation restrictions implied by the affine model. Prices at all maturities must obey a strict factor structure so that physical comovement among prices is driven by H_t. Furthermore, the loadings at each maturity are tightly restricted—they must follow a geometric progression in $\rho^{\mathbb{Q}}$—reflecting the fact that prices along the term structure arise from investors' iterated expectations under the $\mathbb{Q}$ measure. Our empirical tests investigate the extent to which observed term structures adhere to the model restrictions.

II.C. Tests of the Affine-$\mathbb{Q}$ Model

We propose two approaches to testing the internal consistency of asset term structures in the affine-$\mathbb{Q}$ setting.

---

14. We report a simple example illustrating the link between $\mathbb{P}$ and $\mathbb{Q}$ measures in Online Appendix D. An attractive feature of our testing framework is that we do not require a linearity assumption for the $\mathbb{P}$ model. In some asset classes like variance swaps, models often include additional assumptions such as conditional heteroskedasticity to ensure the x_t process remains positive. For pricing of linear claims, heteroskedasticity does not affect the pricing formula in equation (1) because the error term remains conditionally mean 0. We abstract from heteroskedasticity in our main analysis, and find that our conclusions are unchanged if we account for heteroskedasticity in residuals via GLS or GARCH regression.

1. Regression-Based Tests. Our first set of excess volatility tests require only OLS regressions of prices at long maturities on prices at the short end of the term structure. Regression-based tests have the virtue of simplicity, do not require assumptions about detailed $\mathbb{P}$ dynamics, and do not require distributional assumptions for model shocks. These tests do, however, require the following additional assumption that prices are well behaved under $\mathbb{P}$ and short-dated claims are free of measurement error.

ASSUMPTION 4REG. Under the $\mathbb{P}$ measure, the term structure of prices satisfies standard regularity conditions for OLS and Wald test consistency.[15] In addition, for a K-factor model, $v_{t,n} = 0$ for $n = 1, \ldots, K$.

To construct the regression-based excess volatility test, let $P_{t,1:K}$ be the $K \times 1$ vector of prices $p_{t,1}$ through $p_{t,K}$. Denote the loading of the jth maturity price on the latent factor as $b_j = (\rho^{\mathbb{Q}} + \ldots + (\rho^{\mathbb{Q}})^j)\mathbf{1}$. According to Assumption 4REG, we can represent the K latent factors in terms of prices for the first K maturities:

$$(8) \qquad H_t = B_{1:K}^{-1}(P_{t,1:K} - \delta_0[1, \ldots, K]'),$$

where $B_{1:K} = [b_1, \ldots, b_K]'$ is the $K \times K$ matrix of stacked factor loadings for the short-end prices.

Because observed short-maturity prices $P_{t,1:K}$ stand in for latent factors, we can recover an estimate of $\rho^{\mathbb{Q}}$ via OLS regression of the $K + 1$ maturity price on the first K prices:

$$(9) \qquad p_{t,K+1} = \alpha_{K+1} + \beta'_{K+1}P_{t,1:K} + v_{t,K+1}.$$

Under the affine-$\mathbb{Q}$ specification, the population slope coefficients satisfy

$$(10) \quad \beta'_{K+1} = b'_{K+1}(B_{1:K})^{-1} = \mathbf{1}'\left(\rho^{\mathbb{Q}} + \ldots + \rho^{\mathbb{Q}^{K+1}}\right)$$
$$\times \left(\left[\rho^{\mathbb{Q}}\mathbf{1}, \ldots, \left(\rho^{\mathbb{Q}} + \ldots + \rho^{\mathbb{Q}^K}\right)\mathbf{1}\right]'\right)^{-1}.$$

15. See, for example, Hayashi (2000), Proposition 2.3. In particular, the term structure must be stationary, have a nondegenerate covariance matrix, and have residuals that satisfy a central limit theorem. These conditions, together with the continuous-mapping theorem, ensure consistency of our regression-based variance ratio tests, which are based on the same asymptotic principles as a Wald test.

This is a system of K polynomial equations in the K unknown parameters of $\rho^{\mathbb{Q}}$. We obtain the estimate $\hat{\rho}^{\mathbb{Q}}$ by numerically inverting system (10) given the OLS slope estimate $\hat{\beta}_{K+1}$, as in Hamilton and Wu (2012). We do not impose a priori that the $\mathbb{Q}$ dynamics are stationary, though they are estimated to be stationary in all of our asset classes.

This estimate of $\hat{\rho}^{\mathbb{Q}}$ forms the basis of our excess volatility test. The test also requires, for any maturity $j > K + 1$, a regression of $p_{t,j}$ onto the short-maturity prices

$$(11) \qquad p_{t,j} = \alpha_j + \beta_j' P_{t,1:K} + v_{t,j}.$$

We construct the variance ratio statistic at maturity j by comparing the explained price variation from restricted and unrestricted versions of regression (11).

The restricted version of the regression respects constraints on the factor loadings implied by the affine model. In particular, we denote the restricted regression slope estimate as $\hat{\beta}_j^R$, and calculate this by plugging $\hat{\rho}^{\mathbb{Q}}$ into the right side of equation (10) (extended to maturities above $K + 1$). By evaluating equation (10) at model parameters estimated from the short end, we impose that the regression model for the j-maturity contract is exactly consistent with behavior of the first $K + 1$ prices. The explained price variation for maturity j in the restricted model is then given by

$$(12) \qquad V_j^R = \hat{\beta}_j^{R'} \hat{V}(P_{t,1:K}) \hat{\beta}_j^R,$$

where $\hat{V}(P_{t,1:K})$ is the sample covariance estimate for short-end prices under $\mathbb{P}$.

Next, the unrestricted version of regression (11) ignores constraints that the affine model places on the factor loadings. Instead, it estimates the factor loading as the unrestricted OLS slope estimate, denoted $\hat{\beta}_j^U$. The explained price variation for maturity j in the unrestricted model is then

$$(13) \qquad V_j^U = \hat{\beta}_j^{U'} \hat{V}(P_{t,1:K}) \hat{\beta}_j^U.$$

Note that the $\hat{V}(P_{t,1:K})$ matrix enters the same way in both V_j^R and V_j^U, so the test amounts to a comparison of the restricted and unrestricted factor loadings. Also note that measurement error variance at long maturities does not directly enter into the

model-explained variance expressions, which is why Assumption 4REG rules out measurement error only for short-maturity prices and not at the long end.

Finally, the variance ratio statistic for maturity j is given by

$$(14) \qquad VR_j = \frac{V_j^U}{V_j^R}.$$

VR_j calculates the fraction of price variation at maturity j that is consistent with variation at other maturities according to the model's cross-equation restrictions. Under the null of a K-factor affine model, $VR_j = 1$. Any deviation from unity (above and beyond that due to sampling variation) indicates a violation of the model's restrictions. Variance ratios that are significantly greater than unity indicate that long-maturity prices overreact to movements at the short end, relative to the affine model.

There are many potential ways to formulate tests of the affine model's restrictions, and many of these are asymptotically equivalent. Our specific test construction has the attractive interpretation as a measure of excess volatility relative to a benchmark model. Our test choice is inspired by, and designed to remain comparable with, the rich history of excess-volatility tests studied by Shiller (1981), Stein (1989), Campbell and Shiller (1988a), Campbell (1991), Cochrane (1992), and others.

Under the null of an affine no-arbitrage model, the restricted and unrestricted loading vectors $\hat{\beta}_j^U$ and $\hat{\beta}_j^R$ should be equal element by element. When there is more than one factor in the model, it raises the question of how to best evaluate joint restrictions that apply to multiple loadings. An attractive feature of the variance ratio test is that it offers a sensible aggregation of all of the loading comparisons. The total explained variance in the restricted and unrestricted models are

$$\sum_{k=1}^{K}\sum_{l=1}^{K} \tilde{b}_{j,k}\tilde{b}_{j,l}\hat{\sigma}_{k,l} \quad \text{and} \quad \sum_{k=1}^{K}\sum_{l=1}^{K} \hat{b}_{j,k}\hat{b}_{j,l}\hat{\sigma}_{k,l},$$

where $\hat{\sigma}_{k,l}$ is the (k, l) element of $\hat{V}(P_{t,1:K})$. Rather than comparing loadings element-wise, the variance ratio sums loadings into a scalar to compare alternative models. The weights assigned to elements in the sum are based on the (co)variances of the short-maturity prices. The prices that most strongly covary are also the most informative about the dynamics of the model, and their factor loadings receive the largest weights in our test.

Equations (12) and (13) illustrate why the regression-based test does not require specification of physical factor dynamics. The test statistic requires only two inputs, coefficients in a contemporaneous regression (11) and the unconditional covariance matrix of the factors (represented via the short-end prices). Both of these elements can be estimated without consideration of physical price dynamics other than requiring that the factor covariance matrix is finite.

In Online Appendix E we describe a bootstrap procedure for conducting inference in our regression-based variance ratio test. Our bootstrap provides a *p*-value calculation to answer the question, "How likely are we to observe a variance ratio as large as the one we observe in the data, under the null of an affine model, given the sampling error of model parameter estimates?" Online Appendix E also reports simulations demonstrating the finite-sample performance of our estimation and testing approach.

2. Maximum Likelihood Tests. The regression-based variance ratio test has a number of benefits, but has the shortcoming of requiring that short-end prices are observed without error. If this assumption is violated, the estimate of $\rho^{\mathbb{Q}}$ suffers attenuation bias, which in turn biases the variance ratio statistic.

If we relax Assumption 4REG to allow measurement error in short-maturity prices, the factor space is no longer observable. Estimation of the model, then, requires latent factor techniques. The system's structure lends itself naturally to maximum likelihood estimation via Kalman filtering, which is the estimation approach we pursue (we refer to it in shorthand as KF-MLE).

In addition to Assumptions 1–3, to use KF-MLE we must also specify the $\mathbb{P}$-dynamics of factors and make a distributional assumption for the errors.

ASSUMPTION 4MLE. Physical factor dynamics follow

$$(15) \qquad H_t = c + \rho H_{t-1} + \Gamma \epsilon_t,$$

and prices obey

$$(16) \qquad P_{t,1:N} = \delta_0 [1 \ \ldots \ N]' + [I_K \ \beta_{K+1} \ \ldots \ \beta_N]' H_t + \nu_t,$$

where $\epsilon_t \sim N(0, I)$ and is i.i.d. The vector of measurement error across maturities is likewise i.i.d. and obeys $\nu_t \sim N(0, \Sigma)$.

Under this assumption the model constitutes a linear Gaussian state space system and is therefore efficiently estimated with KF-MLE. In the state space setting we can construct long-maturity variance ratio statistics that are exactly analogous to the regression-based variance ratios described earlier.

In both the restricted and unrestricted models, the physical state transition equation is the same and is given by equation (15). The term structure of prices constitutes the system's observation equations. Due to the presence of measurement error at all maturities, we can no longer use the price representation of equation (11), and instead represent the price vector as equation (16). There is an observation equation for every price in the term structure. The first block of the factor-loading matrix is fixed at I_K. This identifies the system by anchoring the factors to the short-end prices, and is the noisy price version of the factor representation used in equation (8). Like Assumption 3, this ensures econometric identification but imposes no further economic restrictions on the model.

The specification of factor loadings $(\beta_{K+1}, \ldots, \beta_N)$ depends on whether affine restrictions are imposed on the system. In the unrestricted model, we estimate separate unconstrained factor loadings for each long-maturity price. On the other hand, factor loadings in the restricted model are jointly determined by the same K underlying parameters in $\rho^{\mathbb{Q}}$ (with the loadings' specific functional form described in equation (10)).

There are two further details of our state space specification. First, our tests focus on restrictions among factor loadings and leave the intercepts of the model free in both versions.[16] Second, because the variance of cumulative prices increases with maturity, we specify the measurement error covariance matrix Σ such that its diagonal elements are in fixed proportion to the unconditional variance of prices. We allow measurement error to be correlated across maturities according to a single correlation parameter.[17]

---

16. In particular, we estimate a separate intercept d_n in each observation equation, rather than restricting $d_n = \delta_0 n$. This choice keeps our tests in the state space setting conceptually identical to the regression-based tests.

17. More specifically, the measurement error covariance specification is reduced to two parameters, σ and ζ, where $\Sigma = DRD$, with $D = \text{diag}(\sigma\sqrt{\hat{V}(p_{t,1})}, \ldots, \sigma\sqrt{\hat{V}(p_{t,N})})$ and $R = (1-\zeta)I + \zeta\mathbf{11}'$. In the interest of notational simplicity, we admit a slight abuse of notation as Σ is in fact a function of data.

We estimate both the restricted and unrestricted model by maximizing the conditional likelihood derived from the Kalman filter. We refer to the unrestricted loading estimates as $\hat{\beta}_j^U$ ($j = K + 1, \ldots, N$), keeping the KF-MLE notation in line with that of the regression-based analysis. Note that specification (16) normalizes the latent factors so that estimated loadings are interpreted as coefficients on the first K prices, after removing measurement error.

Optimizing the likelihood of the restricted model produces an estimate of the deeper persistence parameter, $\hat{\rho}^{\mathbb{Q}}$. We transform this into a joint estimate of the restricted factor loadings $\hat{\beta}_j^R$ by plugging $\hat{\rho}^{\mathbb{Q}}$ into equation (10) for each j. With KF-MLE loading estimates for both models in hand, the variance ratio is constructed identically to that in the regression-based test following equations (12)–(14). In other words, the Kalman filter allows us to estimate the regression equation (11) while allowing for measurement error throughout the curve. As a result, the loading estimates and resulting variance ratio statistics are unbiased by the presence of measurement error. The KF-MLE test also has the advantage that the model is estimated from all maturities simultaneously, rather than from separate regressions for each maturity.

Because variance ratios are continuous nonlinear functions of the parameters estimated in the model, their asymptotic standard errors are easily obtained via the delta method. In addition, because the restricted model is nested in the unrestricted specification, we can also conduct a more powerful likelihood ratio test of the restricted model versus the unrestricted model. Unlike the variance ratio statistics which test for excess volatility maturity by maturity, the likelihood ratio statistic allows us to jointly test the affine-model restrictions using all maturities at once.

III. Empirical Findings

This section presents our main empirical findings. We study term structures of variance swaps, equity options, currency options, credit default swaps, commodity futures, inflation swaps, and Treasury bonds. In each asset class, we describe details of the term structure data and model specification, then report excess volatility test results.

Wherever possible, we take the number of factors K from specification analysis in previous literature. For example, it is standard practice to use three factors when describing the term structure of Treasury yields and two factors for the variance swap market. We then check that a K-principal-component model explains at least 99% of the variation in the panel of prices at all available maturities. In asset classes where the literature provides no guidance on K, we set K to the number of principal components necessary to explain at least 99% of the term structure price variation.[18]

We first report detailed analyses of one particular example, variance swaps, then show that the same results hold in other asset classes. Interested readers can find further details on data and model specification for each asset class in Online Appendix F.

III.A. S&P 500 Variance Swaps

The variance swap market allows investors to trade direct claims on the riskiness of equities. A long variance swap position receives cash flows at maturity proportional to the sample variance of the S&P 500 over the life of the contract. Let RV_t denote the sum of squared daily log index returns during calendar month t. The payoff of an n-maturity variance swap is $\sum_{j=1}^{n} RV_{t+j}$. Ignoring risk-free rate variation (as is typical in this literature), the price of a variance swap corresponds to the $\mathbb{Q}$-expectation of the payoff:

$$p_{t,n} = E_t^{\mathbb{Q}} \left[\sum_{j=1}^{n} RV_{t+j} \right].$$

This structure maps directly into the simple affine framework of Section II with $x_t = RV_t$. We model RV_t using $K = 2$ latent factors, as in Egloff, Leippold, and Wu (2010), Ait-Sahalia, Karaman, and Mancini (2015), and Dew-Becker et al. (2015).

Variance swaps are traded in a liquid over-the-counter market with a total outstanding notional of around \$4 billion in "vega" at the end of 2013, meaning that a movement of one point in volatility would result in \$4 billion changing hands between counterparties. We obtained a sample of daily variance swap transactions

---

18. We discuss the robustness of our findings to inclusion of additional factors in Section IV.A.

TABLE I

TERM STRUCTURE LIQUIDITY

Asset class	Liquidity measure	Maturity				
		0–12 mo.	13–24 mo.	25–36 mo.	37–60 mo.	>60 mo.
Variance swaps	Vega (mil.)	10.0	7.2	4.2	0.5	0.1
		0–6 mo.	7–12 mo.	13–18 mo.	19–24 mo.	25–36 mo.
Apple options	Volume	390.2	51.3	56.3	35.7	29.5
Citigroup options	Volume	128.5	81.7	106.1	57.2	29.1
Euro options	Volume	20.8	9.2	2.6	1.5	1.1
Yen options	Volume	7.4	4.2	4.1	3.1	1.7
STOXX 50 options	Volume	1,183.3	451.9	277.9	128.6	55.9
DAX options	Volume	321.0	98.3	68.2	40.9	14.1
Gold futures	Volume (thous.)	152.8	2.8	0.8	0.4	0.8
Crude oil futures	Volume (thous.)	504.3	48.1	17.2	8.4	9.1
		0–5 yr.	6–10 yr.	11–20 yr.	25 yr.	30 yr.
U.S. inflation swaps	Bid-ask spread	1.7%	0.8%	0.8%	0.7%	0.8%
EU inflation swaps	Bid-ask spread	3.2%	1.9%	2.1%	2.0%	1.9%
			0–3 yr.	4–6 yr.	7–11 yr.	>11 yr.
Treasuries	Volume ($bil.)		165.2	124.5	116.8	29.0
			0–2 yr.	3–4 yr.	5–6 yr.	7–12 yr.
Brazil CDS	Volume (% of tot.)		5.4	4.1	82.0	8.4
Russia CDS	Volume (% of tot.)		22.6	7.0	58.8	11.5
GE CDS	Volume (% of tot.)		39.5	23.7	21.1	15.8
BofA CDS	Volume (% of tot.)		13.2	18.3	53.9	14.6

Notes. Volume and vega statistics are daily averages. Volumes are reported in number of contracts for options and futures markets and in dollar volume for Treasuries. CDS averages are the fraction of trades occurring in each maturity bin. Percentage bid-ask spreads are defined as $100 \frac{\text{ask-bid}}{\frac{1}{2}(\text{ask+bid})}$.

collected by DTCC between March 2013 and June 2014, and the first row of Table I summarizes trading volume from this sample. Volumes are reported as the average daily vega (in millions) transacted between counterparties; the table shows that swaps are frequently traded at all maturities up to 24 months, and there is even some volume at longer maturities. Bid-ask spreads for maturities up to 24 months are relatively low at 1–2% of the claim price.[19] Our estimation uses daily price data for cumulative claims

19. In addition, the liquidity of the swap market is supported by option market liquidity. Variance swaps are anchored to the prices of S&P 500 index options by a no-arbitrage relationship because options can be used to synthetically replicate the swap.

at maturities 1, 2, 3, 6, 12, and 24 months during the period 1996–2013.

Figure I presents variance ratios from the regression-based test. The unrestricted price variance at 24 months more than doubles the variance allowed under the affine-pricing model's restriction. Comovement among prices at the short end of the curve suggests that cash flows mean revert relatively quickly under $\mathbb{Q}$. But this is not borne out on the long end—model-restricted volatilities increase with maturity at a much slower rate than the unrestricted volatility. This long maturity overreaction is relative to the short end, and relative to the estimated affine model. Table II, Panel A reports the variance ratio test using the KF-MLE method. These results are economically and statistically the same as the regression-based test results.

Plotting price variability in terms of standard deviation is convenient for visualizing the degree of cash flow persistence under the pricing measure. For example, in a one-factor model with $\rho^{\mathbb{Q}} > 0$, the model-based standard deviation of an n-maturity claim is $\left(\sum_{j=1}^{n} (\rho^{\mathbb{Q}})^j \right) \sqrt{Var(p_{t,1})}$. If $\rho^{\mathbb{Q}} = 1$, so that cash flows are integrated under the pricing measure, then the standard deviation is a linear function of maturity. On the other hand, if $\rho^{\mathbb{Q}} < 1$, then the standard deviation of price is a concave function of maturity. For variance swaps (indeed, for all other term structures we study), the unrestricted estimate of price volatility is concave in maturity, suggesting stationarity of cash flows under the pricing measure.

Three points warrant emphasis regarding these results. First, the excess volatility of long-maturity claims cannot be explained by discount-rate variation that is describable within the affine class, as this is subsumed by the $\mathbb{Q}$ model. Second, the data are exceedingly well described by a linear factor model (as evident from the unrestricted R^2), but with factor loadings that sharply differ from those implied by model restrictions. Figure II separately plots regression loadings of prices on each factor for both the restricted and the unrestricted model.[20] The figure shows that long-maturity prices overreact because they load too heavily on both factors, relative to the loadings predicted by the null model. Third, the close similarity between likelihood-based and regression-based variance ratios suggests that the excess

---

20. These plots look essentially identical in the KF-MLE approach.

TABLE II

VARIANCE RATIO TESTS

Asset	Model		Estimation method					
	K	R^2	Regression			KF-MLE		
Panel A: Equity variance								
			6 mo.	12 mo.	24 mo.	6 mo.	12 mo.	24 mo.
Variance swaps	2	99.7	1.00	1.22**	2.15**	1.03	1.31**	2.49**
			12 mo.	18 mo.	24 mo.	12 mo.	18 mo.	24 mo.
Apple IV	2	99.3	1.21**	1.56**	2.01**	1.30**	1.80**	2.42**
Citigroup IV	2	99.7	1.82**	3.17**	4.68**	1.33**	0.99	0.61
STOXX 50 IV	2	99.4	1.22**	1.68**	2.27**	1.16**	1.50**	1.97**
DAX IV	2	99.4	1.22**	1.68**	2.31**	1.17**	1.56**	2.08**
Panel B: Currency variance								
			12 mo.	18 mo.	24 mo.	12 mo.	18 mo.	24 mo.
Euro IV	2	99.8	1.22**	1.65**	2.14**	1.13*	1.38**	1.65**
Yen IV	2	98.5	1.67	2.85*	4.57*	1.15	1.18	1.22
Panel C: Interest rates								
			20 yr.	25 yr.	30 yr.	20 yr.	25 yr.	30 yr.
Treasuries	3	99.9	1.20**	1.39**	1.64**	1.43	1.92	2.04
Panel D: Inflation								
			20 yr.	25 yr.	30 yr.	20 yr.	25 yr.	30 yr.
U.S. infl. swaps	4	99.4	3.37**	5.54**	7.47**	2.10**	2.85**	3.91**
EU infl. swaps	4	99.1	1.74**	2.45**	2.89**	2.65**	5.48**	8.51**
Panel E: Commodities								
			6 mo.	12 mo.	24 mo.	6 mo.	12 mo.	24 mo.
Crude oil fut.	2	99.6	1.01**	1.19**	1.63**	0.99	1.01	1.14
Gold fut.	2	99.5	1.04*	1.19**	1.53**	1.13*	2.46*	9.33
Panel F: Credit								
			5 yr.	7 yr.	10 yr.	5 yr.	7 yr.	10 yr.
Brazil CDS	2	99.8	1.19**	1.64**	3.08**	1.28**	1.71**	2.60**
Russia CDS	2	99.8	1.14**	1.46**	2.18**	1.19**	1.63**	2.71**
GE CDS	2	99.5	1.12**	1.13**	1.45**	1.33**	1.76**	3.50**
BoA CDS	2	99.7	1.06**	1.14**	1.38**	1.03	1.00	1.02

Notes. The table reports long-maturity variance ratio test results. Regression-based estimates are reported on the left side and likelihood-based estimates are on the right. Significance for the one-sided test that the variance ratio is greater than one at the 1% level is denoted by ** and at the 5% levels by *.
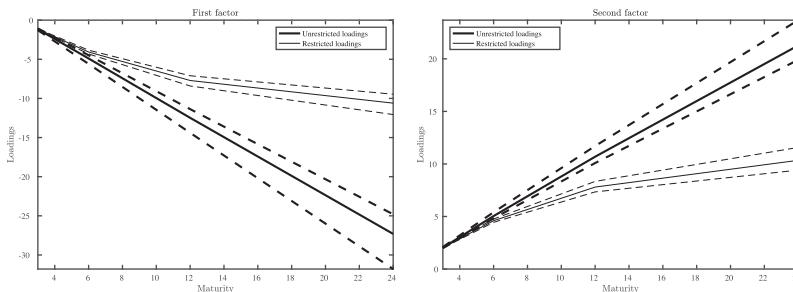
FIGURE II

Variance Swaps: Individual Factor Loadings

The figure plots loadings of prices at each maturity on the two factors. Thick lines indicate the unrestricted model and thin lines the restricted model. Dashed lines are 95% confidence bands.
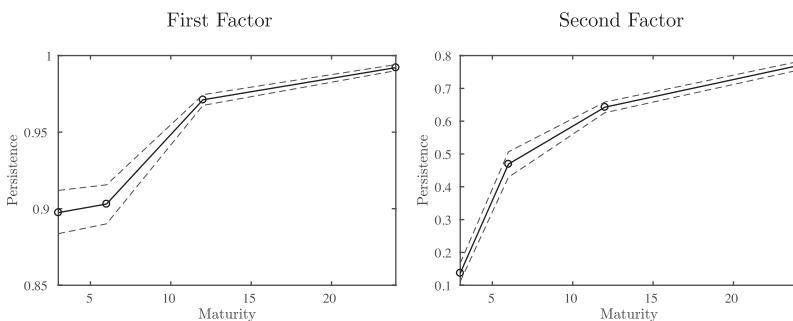

FIGURE III

Estimates of $\rho^{\mathbb{Q}}$ by Maturity

The figures plot estimates of persistence parameters in the two-factor variance swap model from different points in the term structure. The left panel shows loadings on the first factor, the right panel loadings on the second factor. Dashed lines are 95% confidence bands.

volatility is a robust phenomenon and is not explained by measurement error.

Figure III provides a different visualization of how the data deviate from the affine model. In the regression test, we estimate factor persistences by regressing the third-shortest-maturity claim on the first two. Figure III shows estimated factor persistences when we use data from different points along the maturity curve. First, we estimate $\rho^{\mathbb{Q}}$ from a regression of maturity 3 on maturities 1 and 2, then from a regression of 6 on 2 and

3, then 12 on 3 and 6, and finally 24 on 12 and 6. Under the null of the affine model, both sets of factor loading estimates should be flat, as the implied factor persistence should be internally consistent along the curve. Instead, the figure shows that estimated persistence increases with maturity (for both factors). In other words, it is as though investors treat factors as more persistent when valuing longer maturity claims.

III.B. *Equity Implied Variance*

We next turn to option markets. Like variance swaps, options allow investors to trade term structures of equity volatility. But the options market is much richer in that claims are liquidly traded for hundreds of underlyings beyond the S&P 500 index. A relative drawback of the option market is that while variance swaps fall neatly into the affine framework of Assumptions 1 and 2, options do not.

Fortunately, well-known results in the option-pricing literature establish that variance swaps and, closely related, volatility swaps, can be accurately approximated using options. Britten-Jones and Neuberger (2000) and Jiang and Tian (2005) show how a portfolio of options with different strike prices replicates a variance swap, while Carr and Lee (2009) show that at-the-money Black-Scholes implied volatility approximates the price of a volatility swap. Synthetic swaps constructed from options are frequently encountered in practice. The most prominent example is the VIX index maintained by the Chicago Board Options Exchange, whose squared value replicates the price of a variance swap on the S&P 500 index.

Following the seminal work of Stein (1989) on excess volatility in the options market, and motivated by the synthetic swap results referenced above, we treat implied variances as proxies for the price of a claim to realized variance. That is, we conduct our tests using at-the-money Black-Scholes implied variances as the term structure of prices. We study option term structures for two individual stocks, Apple and Citigroup, which are the two most actively traded single-name term structures in Option-Metrics by contract volume. These data are from the IVY DB US file and are available from 1996 to 2014. We also study options for the two most liquid international stock indexes, STOXX 50 and DAX, which are from the IVY DB Europe file covering 2002–2013. Options for all underlyings are liquidly traded up to at least

24 months to maturity, as shown in Table I. We set $K = 2$ following the variance swap literature, and find that two factors explain more than 99% of the panel variation in implied variances for each of the option term structures we examine.

Table II, Panel A reports excess volatility tests and shows that equity option term structures possess the same excess volatility pattern as S&P 500 variance swaps. Variance ratios at the longest maturities range between 2.01 and 4.68, depending on the underlying and the estimation method, and are significantly different than one at the 1% significance level or better. The exception is Citigroup, for which the KF-MLE variance ratio drops below one and the likelihood ratio test of joint restrictions fails to reject the affine model.

III.C. Currency Implied Variance

We next test the term structure of claims to exchange rate volatility. To do so, we analyze the currency option market and use the same model specification that we used for equity options. Currency options are traded on the Chicago Mercantile Exchange (CME), and the two most liquid term structures are for options on the euro and yen versus the U.S. dollar. Our tests use options on currency ETFs from OptionMetrics, whose data are more complete and avoid recording errors that occasionally surface in the CME data. Not coincidentally, the euro and yen (via the FXE and FXY tickers of the Guggenheim CurrencyShares ETF family) are also the most liquid currency ETF options. Table I shows that there is daily volume for these contracts at maturities up to at least 24 months.[21] Our sample runs from 2007 to 2014.

The currency patterns in Table II, Panel B are qualitatively similar to those of equity variance claims. Regression-based variance ratios at 24 months are 2.14 and 4.57 for euro and yen, respectively, and are significant at the 1% level and 5% level, respectively. The KF-MLE variance ratio is 1.65 for the euro, again highly significant. It drops to 1.22 for the yen and is insignificant. However, the yen likelihood ratio test rejects the joint affine restrictions for all maturities with a *p*-value below 1%.

---

21. We have also verified that CME and OptionMetrics implied variances share an extremely close correspondence on the subset of days for which reliable CME data are available.

III.D. *Interest Rates*

U.S. government bond prices are among the most well-studied data in all of economics. U.S. bond data comes from Gurkaynak, Sack, and Wright (2006). Our tests use daily zero-coupon nominal bond yields with maturities of 1 through 30 years for the period 1985–2014. The U.S. Treasury market is also among the most liquid markets in the world. The Securities Industry and Financial Markets Association (SIFMA) provides average daily dollar volumes in coarse maturity bins for 2002–2014, which we report in Table I.

We estimate a standard homoskedastic exponential-affine model for yields. We choose $K = 3$ factors based on broad consensus in the interest rate literature. We discuss this specification in detail in Online Appendix B and perform a robustness analysis with respect to heteroskedasticity.

The variance ratio tests in Table II, Panel C show excess volatility at long maturities in the Treasury curve. The variance ratio reaches 1.64 at 30 years in the regression test and 2.04 with KF-MLE. While the KF-MLE variance ratio is not statistically significant, the magnitude of excess volatility is in line with the regression test and our findings for other asset classes. Furthermore, the likelihood ratio test rejects the joint affine restrictions using all maturities with a *p*-value below 1%.

III.E. *Inflation Swaps*

Inflation swaps are claims whose payoffs are proportional to realized CPI inflation over the life of the contract. We obtain U.S. and EU inflation swap price data from Bloomberg. This includes a full term structure of maturities between 1 and 30 years observed at the daily frequency and available for the period 2004–2014. Our inflation swap data do not include volume. We do observe bid-ask spreads, however, and report average spreads in Table I to provide a sense of liquidity. Spreads are approximately 1% of the U.S. inflation swap price, 2% for the EU data, and are somewhat larger at short maturities. Our U.S. inflation swap data are also studied in Fleckenstein, Longstaff, and Lustig (2013). The Federal Reserve report of Fleming and Sporn (2013) notes that "The U.S. inflation swap market is reasonably liquid and transparent. That is, transaction prices for this market are quite close to widely available end-of-day quoted prices, and realized bid-ask spreads are modest."

The term structure model for inflation swaps falls within the exponential-affine specification of Section II, as we describe in Online Appendix F. We set $K = 4$, which is the number of factors required to explain at least 99% of the variation in the panel of swap prices. Table II, Panel D shows that 20-year regression-based variance ratios are 3.37 in U.S. data and rise to 7.47 for 30 years. In EU data, the 20-year regression-based variance ratio is 1.74 and the 30-year is 2.89. KF-MLE corroborates the excess volatility assessed by the regression method.

III.F. Commodity Futures

We next analyze the term structure of commodity futures. We study the most liquid energy commodity, crude oil, and the most liquid metal, gold, based on volume data from CME Group. Contracts for both commodities are heavily traded at both the short end (1 month) and long end (24 months) of the term structure, as shown in Table I.

Commodity futures reflect $\mathbb{Q}$-expectations of the future price of the underlying, which is in turn linked to the current price of the underlying and to the $\mathbb{Q}$-expectation of the convenience yield. One of the advantages of modeling only the $\mathbb{Q}$ measure is that we do not have to explicitly model or estimate the physical process for the convenience yield and can instead work solely with futures prices. Online Appendix F describes how we map futures prices into the exponential-affine setup. We apply our tests with $K = 2$ factors.

Table II, Panel E shows that the regression-based 24-month variance ratio reaches 1.63 for oil and 1.53 for gold, both significant at the 1% level. The KF-MLE analysis corroborates this pattern, but with weaker statistical significance.

III.G. Credit Default Swaps

Credit default swaps (CDS) are the primary security used to trade and hedge default risk of sovereigns and corporations. As of December 2014, the outstanding notional value of single-name CDS was \$10.8 trillion. Our CDS data are from MarkIt, which reports maturities from 1 to 30 years.

We analyze CDS for the two most liquid sovereign names (Brazil and Russia) and two most liquid corporate names (Bank of America and General Electric) based on average daily notional dollar volume reported by the DTCC and aggregated over all

maturities. Using more detailed confidential DTCC data, Siriwardane (2015) summarizes the distribution of daily contract volume by maturity for the term structures we study from 2010–2014, which we report in Table I. These show that while much of the volume is concentrated near the five-year contract, there is substantial activity in maturities below three years and above seven years. We study maturities of 1, 2, 3, 5, 7, and 10 years over the 2007–2014 sample.

In Online Appendix F we describe how we map CDS prices into the framework of Section II. The link to the affine setup is based on an exponential-affine specification for defaultable bonds from Duffie and Singleton (1999), noting that the CDS spread can be expressed as an approximate linear function of the yield of a defaultable bond. Our CDS model sets $K = 2$ following a literature using two-factor models to describe term structures of credit spreads.[22]

Regression-based 10-year variance ratios for sovereign CDS of Brazil and Russia are 3.08 and 2.18, respectively. They are 1.45 and 1.38 for General Electric and Bank of America, respectively. In all four cases the regression-based statistics are significant at the 1% level. The KF-MLE results are similar: they remain large and significant for Russia, General Electric, and Brazil (but not for Bank of America).

The general conclusion from Table II is that excess volatility of long-maturity claims is a pervasive phenomenon. The simple regression-based tests indicate that excess volatility is economically large and highly significant in all asset classes. The likelihood-based tests, which are robust to measurement error, appear somewhat noisier but convey the same overall picture as the regression analysis.

IV. POTENTIAL SOURCES OF VIOLATION

In this section we explore potential explanations for excess price volatility relative to the affine-$\mathbb{Q}$ model. We classify possibilities into two categories. The first category is model misspecification, such as the rejection of $\mathbb{Q}$ restrictions being due to dynamics that are not affine under $\mathbb{Q}$. The section shows that a broad range of nonaffine no-arbitrage models cannot explain the excess

22. See Madan and Unal (2000) and Christensen and Lopez (2008).
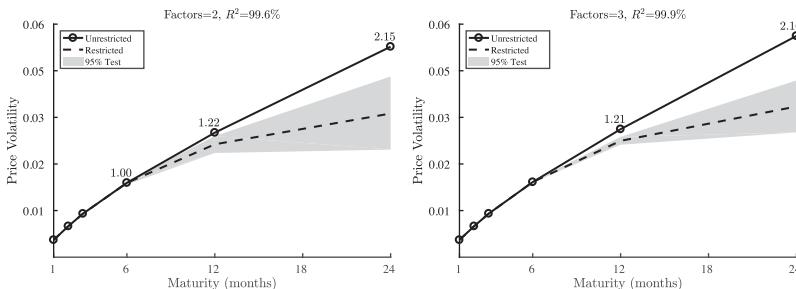
FIGURE IV

Variance Swaps: Varying the Number of Factors

See Figure I.

volatility patterns, mainly because affine models approximate nonlinear models remarkably well.

The second category of explanations we explore is mispricing arising from expectation errors. We analyze term structure pricing predictions in a leading model of extrapolative expectations. The model produces long-maturity excess volatility closely consistent with observed data patterns and offers insight into the key modeling features that generate these patterns. We also show that trading against excess volatility appears profitable above and beyond the risk endured.

Our intention in this section is not to exhaustively explore alternative explanations. Nor can we categorically rule out some forms of misspecification. Instead, our aim is to provide the reader with intuition for how various affine-model violations impact the behavior of variance ratios.

IV.A. Missing Factors

Even if the true model were an affine-factor model, prices might appear excessively volatile if the estimated model has too few factors relative to the truth.

Figure IV shows variance swap tests when we increase the number of factors in the null model from two to three. The two-factor case is the main result reported in Figure I, which has an R^2 of 99.6% and a long-end variance ratio of 2.15. With three factors, the R^2 exceeds 99.9%, and the regression-based test continues to produce large economic and statistical rejections of the affine model with essentially identical variance ratios ($VR_{24} = 2.16$).

Similarly, with three factors in the KF-MLE specification, we find $VR_{24} = 2.04$ with a p-value below .001. We see this type of behavior throughout the asset classes we study, and provide further details in Table II of Online Appendix G.[23]

IV.B. Long Memory

Excessive volatility of long-lived claims intuitively raises the possibility that our findings are due to long-memory cash flow dynamics that are poorly captured by the more rapid, geometric mean reversion inherent in affine models.

Our data suggest that cash flows are stationary under $\mathbb{Q}$ in all asset classes we study; this is for example evident from the concave shape of price volatility versus maturity. However, it is possible that cash flows are stationary under $\mathbb{Q}$ yet they mean revert more slowly than an autoregression would suggest. Granger and Joyeux (1980) propose the broad class of fractionally integrated, or ARFIMA, models to capture precisely this type of long-memory behavior. An ARFIMA process is indexed by a parameter d that determines its degree of long-range dependence. When d is in the interval (0,0.5), it is positively fractionally integrated yet stationary (the special case of $d = 0$ corresponds to a standard ARMA process).

We investigate the effect of estimating an affine (short-memory) model when the data are in fact fractionally integrated. No-arbitrage term structure prices become intractable to derive analytically in the ARFIMA setting, but are easily evaluated via simulation. We simulate term structure prices assuming an ARFIMA(1,d,0) model using a grid of values for $d \in (0, 0.5)$ and values of the AR coefficient of 0.25, 0.50, or 0.75.[24] Figure V demonstrates the range of long-memory behavior that is embedded in our simulated term structure. The extremely slow decay for the case $d = 0.49$ illustrates how an ARFIMA process is difficult to

---

23. There is always a factor model that delivers variance ratios equal to 1—it is a model with the number of factors equal to the number of observed maturities. This extreme specification is a reminder that the modeler's objective is to maximize the variety of phenomena explained while minimizing the number of inputs and parameters necessary to do so. Adding factors eats up valuable cross-equation restrictions that give the model its economic and statistical content.

24. For these simulations as well as the simulations of Table VI, we generate data assuming that the $\mathbb{P}$ distribution is the same as the $\mathbb{Q}$ distribution, thus imposing that risk premia are 0.
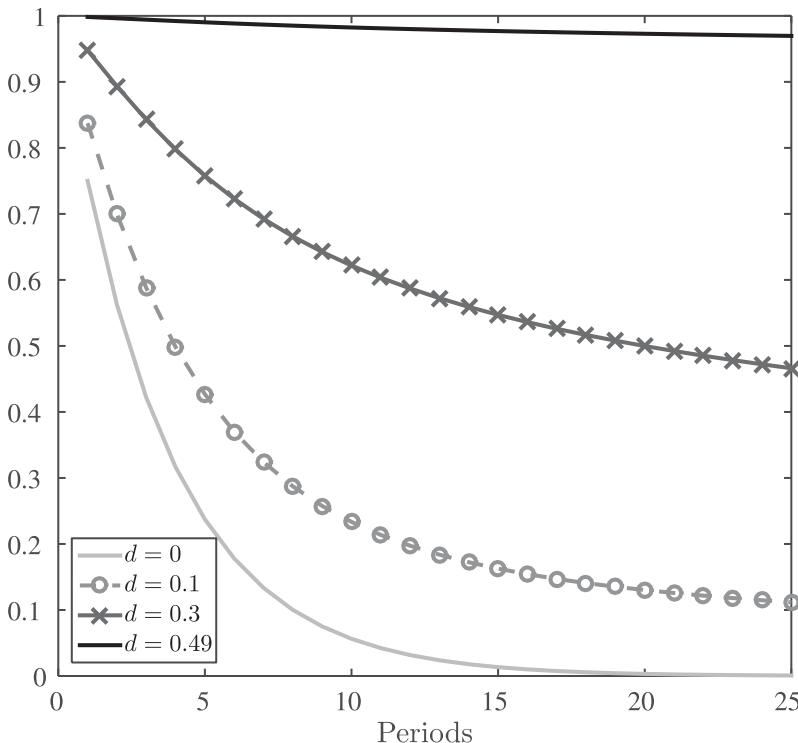
FIGURE V

Long-Memory Mean Reversion

ARFIMA(1,d,0) reversion from a one standard deviation shock to the process's mean value of 0 over 25 periods, assuming an AR(1) coefficient of 0.75 and d values of 0, 0.10, 0.30, and 0.49.

distinguish from an integrated process as d approaches the upper limit of the stationary range.

We calculate prices at maturities up to 24 periods and use a time series sample size of 1,000 periods. Then we estimate and construct variance ratio tests using the misspecified, short-memory affine model with either one, two, or three factors. Results reported in Table III show that it is uncommon to find a model that produces an R^2 greater than 99% along with a variance ratio above two. When this does occur, it is because the long-memory behavior is close to nonstationary. In these cases, inclusion of an "extra" factor brings variance ratios close to one. Evidently,

TABLE III

EFFECTS OF LONG MEMORY

		AR(1) = 0.25			AR(1) = 0.50			AR(1) = 0.75		
d	K	R^2	VR_{12}	VR_{24}	R^2	VR_{12}	VR_{24}	R^2	VR_{12}	VR_{24}
0.10	1	96.8	2.0	2.9	99.1	1.3	1.7	99.9	1.0	1.1
0.10	2	100.0	1.0	1.2	100.0	1.0	1.0	100.0	1.1	1.2
0.10	3	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0
0.20	1	97.1	2.4	4.1	98.9	1.5	2.2	99.9	0.9	0.9
0.20	2	100.0	1.0	1.2	100.0	1.0	1.0	100.0	1.1	1.3
0.20	3	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0
0.30	1	97.7	2.5	4.8	99.1	1.5	2.4	99.9	0.7	0.6
0.30	2	100.0	1.0	1.1	100.0	1.2	1.4	100.0	1.0	1.3
0.30	3	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.1
0.40	1	98.3	2.4	5.0	99.4	1.3	2.2	99.9	0.5	0.3
0.40	2	100.0	1.0	1.1	100.0	1.5	2.8	100.0	1.0	1.2
0.40	3	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.1
0.49	1	98.7	2.3	4.9	99.6	1.1	1.8	99.9	0.4	0.1
0.49	2	100.0	1.0	1.0	100.0	1.4	2.7	100.0	1.0	1.2
0.49	3	100.0	1.0	1.1	100.0	1.0	1.0	100.0	1.0	1.2

Notes. Variance ratios and R^2 computed in simulations of an ARFIMA(1,d,0) model. d corresponds to the order of integration; K is the number of factors used in the variance ratio test. VR_{12} and VR_{24} are the variance ratios at 12- and 24-month maturities. AR(1) is the autoregressive coefficient in the ARFIMA model.

despite its incorrect specification, the affine model with two factors is an accurate enough approximation of the ARFIMA process that the misspecification can go undetected.

IV.C. Nonlinearities

A third potential explanation of our findings is that cash flows evolve nonlinearly. We explore the effects of estimating and testing restrictions of a misspecified affine model when the true cash flow process has nonlinear dynamics. To do so, we study a class of processes known as STAR models.[25] As emphasized by Granger and Teräsvirta (1993), STAR models encompass a broad variety of nonlinear dynamics that have proven successful in modeling economic time series. Although far from exhaustive, they allow us to gain some insight into the role that nonlinearities play in our empirical results.

25. See Teräsvirta (1994) for a detailed econometric treatment of STAR models.
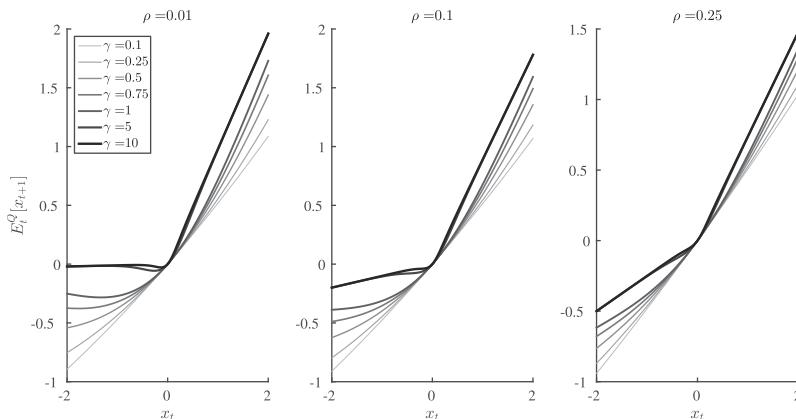
FIGURE VI

Nonlinear Cash Flow Dynamics

The figure shows how the conditional mean of a logistic STAR process depends on the current value of the process x_t. The lines and panels correspond to different parameterization of the STAR process that vary γ and ρ parameters.

We assume that cash flows evolve according to the one-factor nonlinear process

$$(17) \qquad x_t = \rho x_{t-1}(1 - (1 + e^{-\gamma(x_{t-1}-c)})^{-1})$$
$$+ (1 - \rho)x_{t-1}(1 + e^{-\gamma(x_{t-1}-c)})^{-1} + \epsilon_t.$$

Equation (17) is the most commonly used variant in the STAR class and is known as the logistic STAR model. It nests the standard linear autoregression, but allows the process to transition between high and low serial correlation depending on the state of the process.[26] The degree of nonlinearity is governed by two parameters, ρ and γ.

Figure VI plots the model-implied relationship between x_t and $E_t^{\mathbb{Q}}[x_{t+1}]$, illustrating the extent of nonlinearity accommodated by STAR models. When ρ is close to either 0 or 1, the model exhibits extreme state-dependence in cash flows, transitioning between dynamics that are very persistent in some periods and nearly

26. By incorporating time variation in autocorrelation, the STAR model's nonlinearities mimic parameter instability that may arise, for example, from investors learning about ρ.

TABLE IV

EFFECTS OF NONLINEARITY

γ	K	$\rho = 0.01$			$\rho = 0.10$			$\rho = 0.25$		
		R^2	VR_{12}	VR_{24}	R^2	VR_{12}	VR_{24}	R^2	VR_{12}	VR_{24}
0.1	1.0	100.0	1.00	1.00	100.0	1.00	1.00	100.0	1.00	1.00
0.1	2.0	100.0	1.00	1.00	100.0	1.00	1.00	100.0	1.00	1.00
0.1	3.0	100.0	1.00	1.00	100.0	1.00	1.00	100.0	1.00	1.00
0.5	1.0	98.6	1.22	1.49	99.9	1.04	1.04	100.0	1.00	1.00
0.5	2.0	100.0	1.04	1.16	100.0	1.01	1.02	100.0	1.00	1.00
0.5	3.0	100.0	1.01	1.09	100.0	1.00	1.00	100.0	1.00	1.00
1.0	1.0	99.8	1.02	1.04	99.7	1.05	1.07	100.0	1.01	1.01
1.0	2.0	100.0	1.01	1.01	100.0	1.01	1.01	100.0	1.00	1.00
1.0	3.0	100.0	1.00	0.98	100.0	0.99	0.99	100.0	1.00	1.00
5.0	1.0	99.9	1.00	1.01	99.9	1.01	1.02	100.0	1.00	1.00
5.0	2.0	100.0	1.00	1.00	100.0	1.00	1.00	100.0	1.00	1.00
5.0	3.0	100.0	1.00	0.99	100.0	1.00	0.99	100.0	1.00	1.00

Notes. Variance ratios and R^2 computed in simulations of a logistic STAR model with parameters γ and ρ. K is the number of factors used in the variance ratio test. VR_{12} is the variance ratio at 12 months maturity, and VR_{24} is the test at 24 months.

i.i.d. in others. For a given value of ρ, higher γ produces higher curvature and can even mimic a kink when γ is very large.

We simulate no-arbitrage prices in the STAR model at maturities up to 24 periods and use a time series sample size of 1,000 periods. Then we estimate and construct variance ratio tests using the misspecified affine model with up to three factors. The results are reported in Table IV. In this large family of nonlinear models (including rather extreme nonlinearities under certain parameterizations), the variance ratio does not rise far above one in any specification. In other words, the affine specification is a very good approximation to the true nonlinear $\mathbb{Q}$-dynamics and the variance ratio does not detect significant violations of cross-equation restrictions.

In Online Appendix H we explore more complex nonaffine specifications, including heteroskedastic STAR models, mixture STAR/long-memory models, and multifractal models. The behavior of variance ratios in these simulated settings is similar to those in Tables III and IV.

IV.D. Measurement Error

A fourth form of model misspecification that can lead to high variance ratio estimates is measurement error in short-maturity

prices. The KF-MLE analysis of Section III suggests that our findings persist after accounting for noise in prices. Here we expand on this evidence in two ways.

First, we calibrate the magnitude of measurement error needed to generate the excess-volatility patterns we see in the data. To do so, we simulate data from an affine model and ask how much error is needed on the short end of the curve to match observed variance ratios. To simulate affine models that are as close as possible to variance swaps and Treasuries, we estimate the model from the short end of each curve and construct the new data set using the fitted prices from the model. These artificial prices satisfy affine-model restrictions at all maturities by construction. Next, we add i.i.d. measurement error to this artificial data set, reestimate the model, and calculate variance ratios. We choose the size of the measurement error to match the variance ratios from our regression-based tests at the longest available maturity. For variance swaps, we find that measurement error must have a standard deviation of more than two volatility points at the short end of the curve to match long-maturity variance ratios. This is five times larger than the average bid-ask spread of short-dated variance swaps. For Treasuries, we need measurement error to have a standard deviation of at least 10 basis points, or more than 10 times the average bid-ask spread of short-maturity bonds. Thus, in both markets, we require unrealistically large measurement error to produce variance ratios as high as those we document.

We conduct a related calibration in which, rather than adding simulated i.i.d. measurement error to fitted affine prices, we add actual estimated measurement errors from the unrestricted KF-MLE estimation. Regression-based variance ratios for these generated prices show that estimated KF-MLE measurement error is likewise unable to produce variance ratios as high as we observe in the actual data (results are reported in Table IV of Online Appendix G). For example, counterfactual variance ratios reach at most 1.18 for variance swaps, and are smaller in other asset classes.[27]

State space methods (KF-MLE) are one way to account for this error and achieve unbiased estimates. Another way to overcome errors in variables is by finding suitable instruments for latent factors and using instrumental variables (IV) regression. We use

---

27. We are grateful to our referee for suggesting this.
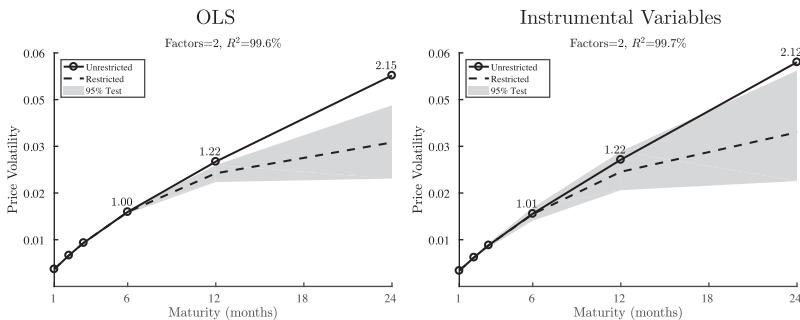
FIGURE VII

Instrumental Variables Adjustment for Measurement Error

This figure compares OLS regression-based variance ratios (left panel) to those based on IV regression (right panel) for the term structure of variance swaps.

IV to construct a modified regression-based test that is robust to measurement error and evaluate how close the resulting variance ratio statistic is to the OLS method.[28]

We conduct this test for the variance swap term structure, and instrument the two latent factors using S&P 500 index return realized variance and the VIX. Both variables are closely related to variance swap prices, but are valid instruments because they have no direct relationship with measurement error in the variance swap prices. Realized variance depends only on the return of the S&P 500, and VIX is calculated from exchange-traded options while variance swaps are OTC.

Figure VII compares variance ratios from OLS and IV regression approaches. Test statistics based on the IV adjustment are very similar to those in the baseline estimation, further indicating that measurement error does not explain the excess-volatility patterns we find.

1. Cautionary Note on R^2. Throughout our analysis we find term structure panel R^2's above 99% using a small number of factors in our regression-based analysis. A high regression R^2 does not rule out misspecification due to omitted factors or measurement error that is unaccounted for. That is, the intuition that a regression R^2 of 99% is almost the same as 100% is potentially flawed.

28. See Online Appendix E for IV test construction details.

To illustrate, Online Appendix K describes a two-factor affine-pricing example. One factor is highly volatile and has little persistence. The other has very low volatility but is highly persistent. In addition, the model includes a small amount of measurement error in prices. Measurement error volatility is less than 1% of the total price volatility.[29]

If we incorrectly specify this model to have a single factor, we essentially identify the high-volatility factor and this is enough to produce a panel R^2 of 99% in a regression on the first maturity alone. If we correctly specify this model to have two factors, we find that the regression R^2 exceeds 99.5%. In both cases, however, we find long-maturity variance ratios that significantly exceed one in regression-based tests. In the first case, this occurs primarily because a factor has been omitted, and this omission would have been hard to detect due to the high R^2. This highlights the value of robustness tests in Figure IV and Table II in which we consider specifications that allow for additional factors.

In the second case, we see that comparatively small measurement error in an otherwise correctly specified model can bias the regression test toward a mistaken rejection of the affine null. This case highlights the importance of our alternative testing schemes. The KF-MLE likelihood function explicitly accounts for measurement error on the short end of the curve. In doing so it produces an unbiased variance ratio statistic and therefore does not reject the affine model. When instruments for the latent factors are available, IV estimation likewise does not reject the (correct) affine null. We refer readers to Online Appendix K for additional detail about this example, including a comparison of variance ratio statistics from regression, KF-MLE, and IV tests.

IV.E. Excess Volatility and Mispricing

A fifth possibility for explaining variance ratios greater than one is that some claims are subject to temporary mispricing. This is another way of stating the joint hypothesis problem that arises in any asset-pricing model test: is a rejection indicating that the null model is incorrect, or that the model is right on average but asset prices sometimes deviate from "true" value?

Two questions arise as we consider the possibility that prices occasionally reflect mispricing. First, can we find evidence that

---

29. We are grateful to our referee for suggesting this example.

favors this view over the alternative of an incorrect econometric model with no mispricing? Second, what type of investor behavior might lead to mispricing? We address these questions in turn.

1. Trading Strategy Evidence. An approach that begins to address the joint hypothesis problem is to understand whether model deviations appear profitable, above and beyond equilibrium compensation for bearing risk. If there exists a strategy that exploits deviations from the null model to earn large trading profits while taking on little risk, it may be evidence of mispricing as a driver of excess volatility.

Under the null of a *K*-factor affine model, we can check at any point whether a long-maturity claim is over- or underpriced relative to the model by comparing traded prices against model-fitted prices. Our evidence of long-maturity overreaction suggests that large increases in short-maturity prices tend to drive long-maturity prices above their model-predicted values. Similarly, large drops in the short end tend to push long-end prices below their predicted value. These amount to temporary mispricings of long claims relative to the model.

The strategy presumes that the estimated affine model is correct on average, so that observed price deviations from the model are temporary and expected to correct. Under this presumption, an investor who detects that traded prices at some maturity have deviated from those predicted by the model can exploit the deviation and can hedge the underlying factor risk using claims at other maturities.

To make the strategy concrete, consider taking a position at time t in a cumulative claim with maturity $N + n > K$ and holding this position for n periods.[30] At $t+n$, the maturity of the position has shortened to N, and is expected to have a correct price (based on the model) of

$$(18) \qquad p_{t+n,N} = \alpha_N + \beta_N' P_{t+n,1:K},$$

where α_N and β_N are model-implied coefficients as in equations (9) and (10). Over the n-period investment period, the claim has paid out cash flows of $x_{t+1}, \ldots, x_{t+n}$.

---

30. In this section we focus on cumulative claims. Online Appendix J presents an alternative trading strategy based on forwards, suggested by our referee.

TABLE V

REPLICATION STRATEGY FOR TRADING

	Strategy $\mathcal{A}$		Strategy $\mathcal{B}$	
Date	Ongoing value	Cash flows	Ongoing value	Cash flows
t	$p_{t,N+n}$	0	$\beta_N' P_{t,1+n:K+n} + \left(1 - \beta_N'\mathbf{1}\right)P_{t,n}$	0
$t+1$	$p_{t+1,N+n-1}$	x_{t+1}	$\beta_N' P_{t,1+n-1:K+n-1} + \left(1 - \beta_N'\mathbf{1}\right)P_{t,n-1}$	x_{t+1}
$\vdots$				
$t+n$	$p_{t+n,N}$	x_{t+n}	$\beta_N' P_{t,1:K} + 0$	x_{t+n}

Notes. Portfolio $\mathcal{A}$ buys the $N+n$-maturity claim at a price of $p_{t,N+n}$. Portfolio $\mathcal{B}$ replicates $\mathcal{A}$ under the affine null model, investing the present value of α_N in the n-maturity risk-free bond (we simplify with a risk-free rate of 0), buying all claims with maturities of $n + 1, \ldots, n + K$ with the number of shares in each claim given by the vector β_N, and buying $\left(1 - \beta_N'\mathbf{1}\right)$ shares of an n-maturity claim.

Construction of the strategy works backward from $t + n$ (when the trade is unwound) to initiation of the trade at time t. In particular, we seek a trade that is expected to have zero liquidation value at $t + n$, but that generates a positive cash flow at initiation. Equation (18) suggests comparing the prices of two portfolios at time t. Portfolio $\mathcal{A}$ simply buys the $(N + n)$-maturity claim at a price of $p_{t,N+n}$. After holding $\mathcal{A}$ for n periods, it has yielded cash flows of $x_{t+1}, \ldots, x_{t+n}$ and has ongoing value of $p_{t+n,N}$.

Portfolio $\mathcal{B}$ is designed to replicate the right-hand side of equation (18). First, it invests the present value of α_N in the n-maturity risk-free bond (for simplicity let us assume that the risk-free rate is 0). Next, it buys all claims with maturities of $n + 1, \ldots, n + K$, corresponding to the price vector $P_{t,n + 1:n+K}$. The exact number of shares purchased in each claim is given by the vector β_N. Third, it buys $\left(1 - \beta_N'\mathbf{1}\right)$ shares of an n maturity claim with price $p_{t,n}$.

After n periods, the risk-free bond has matured with a value of α_N and the position $\beta_N' P_{t,n+1:n+K}$ has ongoing value of $\beta_N' P_{t+n,1:K}$. The n-maturity claim has expired with no remaining value, but has ensured that the intermediate cash flows generated over the life of the trade are exactly $x_{t+1}, \ldots, x_{t+n}$. In short, portfolio $\mathcal{B}$ exactly replicates the expected future value of portfolio $\mathcal{A}$ and exactly matches all intermediate cash flows generated by $\mathcal{A}$, as described in Table V.

Because portfolio $\mathcal{B}$ is an exact hedge to portfolio $\mathcal{A}$ according to the model, any difference in the time t initiation prices of $\mathcal{A}$ and $\mathcal{B}$ represents a mispricing. If the price of $\mathcal{B}$ exceeds that of $\mathcal{A}$, the strategy establishes a long position in $\mathcal{A}$ and a short position in $\mathcal{B}$, and vice versa. This strategy generates a strictly positive

cash flow at time t, exactly offsets all intermediate cash flows, and has zero liquidation value in expectation.[31] Note that even when the investor's presumed affine model is correct on average (so that the investor can accurately detect temporary deviations from the model) this is not a pure arbitrage. It is rather a "good deal on average," as the investor faces uncertainty about when the deviation will correct and whether it will widen before shrinking.

We implement the trading strategy in the variance swap market. We compute the return to this strategy taking into account realistic margin constraints.[32] We also execute the strategy on a purely out-of-sample basis. That is, when deciding on a trade at time t, estimated model parameters and position choices only use data that an investor would have access to in real time. We reestimate the model each day using the most recent 250 trading days. We only trade in periods when the initiation profit Π is sufficiently large, which avoids trading on small mispricings that are indistinguishable from estimation noise. We consider trading thresholds based on the historical distribution of Π. Therefore, at each date t, the initial profit is being compared only with backward-looking information and the trading choice preserves the out-of-sample character of the trade.

The "Variance swaps" column in Table VI reports the annualized Sharpe ratios of a trading strategy using month-end prices, for a one-month holding period ($n = 1$), with various choices for the maturity of the long-end claim being traded ($N + n = 15, 18, 21,$ or 24 months), and with various thresholds for trade initiation

---

31. In practice, the liquidation equation (18) does not hold exactly. To minimize the liquidation risk, α_N and β_N are based on unrestricted regressions of N-maturity prices on prices for maturities 1 through K. This minimizes the squared liquidation error.

32. We assume that each trade must be fully collateralized on both the long position and short position. That is, if the strategy is allocated C dollars of capital to invest, the absolute value of costs for the buy and sell positions must not exceed C. We denote q as the number of units we trade, which we solve for given the capital requirement. Z_S is the per unit cost of the short position, and Z_L the per unit cost of the long position. We write $Z_L = Z_S - \Pi$, where $\Pi > 0$ is the immediate per unit profit realized from the trade (no-arbitrage is equivalent to $\Pi = 0$). Therefore, the number of units traded, q, must satisfy $q \leqslant \frac{C}{2Z_S - \Pi}$. This caps the number of units that can be traded depending on capital and margin. Larger positions can be taken when more capital is available and when haircuts are smaller. These constraints also have the attractive feature that the size of the trade is increasing in the size of the initial profit, Π, relative to a unit position in one leg of the trade, Z_S. We normalize trading capital C to one each period.

TABLE VI

TRADING-STRATEGY SHARPE RATIOS

Mispricing threshold	Longest maturity traded	Variance swaps	Simulations		
			Missing factor	Long memory	Nonlinear
50	15	0.73	−0.01	−0.01	0.00
50	18	1.17	−0.01	0.03	0.00
50	21	0.94	−0.01	0.00	0.01
50	24	0.56	−0.01	−0.02	0.01
75	15	1.43	0.00	0.00	0.01
75	18	1.68	0.00	0.00	0.01
75	21	1.37	0.00	−0.01	0.02
75	24	0.50	0.00	0.02	0.02
90	15	1.56	0.00	0.05	0.03
90	18	1.96	0.00	−0.02	0.03
90	21	1.91	0.00	−0.05	0.04
90	24	1.61	0.00	−0.05	0.05
Average		1.28	0.00	−0.01	0.02

Notes. The table reports annualized Sharpe ratios for trading strategies that exploit mispricing relative to the affine-$\mathbb{Q}$ model. All strategies are implemented using information available to the investor at the time of the trade, and use a one-month holding period ($n = 1$) for each trade. The first column reports at what level of mispricing (relative to the historical distribution) a trade is executed. The second column reports which maturity ($N + n$) the trading occurs on. The third column reports the trading strategy applied on actual variance swap data, while the remaining columns implement the trading strategy on different simulated data sets. Simulations are based on affine-$\mathbb{Q}$ models and therefore the investor operating the trading strategy is using a misspecified model.

(equal to the 50th, 75th, or 90th historical percentile for Π).[33] We obtain consistently high Sharpe ratios in all cases, often above 1.5, and we find higher Sharpe ratios in cases where Π is required to exceed a higher threshold (cases in which the model identifies a large mispricing).

As highlighted earlier in this section, variance ratios above one may arise due to model misspecification in the sense that observed claims are never mispriced but the true model is not affine. Trading based on a misspecified model, when in fact no mispricings exist, should not produce trading profits. To confirm this intuition, we also report results for our trading strategy applied in simulated no-arbitrage models. We compare against three models in which long-maturity variance ratios are greater than one

33. The threshold maps approximately into the fraction of days traded, with the 50th percentile trade triggered about half of the time and 90th percentile trade initiated roughly 1 day in 10.

because the estimated affine model is misspecified, but in which the simulated claims are always correctly priced. These include

i. the two-factor affine model with $\rho_1 = 0.9$ and $\rho_2 = 0.5$, but estimated assuming a one-factor structure;

ii. the long-memory ARFIMA model with $d = 0.3$ and AR(1) coefficient 0.25;

iii. the nonlinear logistic STAR model with parameters $\rho = 0.01$ and $\gamma = 0.5$.

In each of these cases, we simulate a sample of 10,000 term structure observations and run the same trading strategy that we use for the variance swap data. As expected, Sharpe ratios in these cases are uniformly close to 0.

While the Sharpe ratios in the variance swap trade are on average quite high, this is not evidence per se that long-maturity claims are subject to mispricing. It is possible, for example, that a trading strategy based on a misspecified model would yield high average returns by inadvertently loading heavily on risk factors that are not well captured by the affine model.

To test whether this is the case, we compute the alpha of the trading strategy relative to various asset-pricing factors. We focus on the 18-month maturity with a mispricing threshold of 50% and one-month holding period. We scale the trading strategy to have a yearly standard deviation of 20%, comparable with the market. The average annualized return of this strategy is 23% and its Sharpe ratio is 1.26. The alpha relative to the Fama and French (1993) three-factor model is 21% per annum and is highly statistically significant, meaning almost none of the strategy's performance is captured by exposure to the Fama-French factors. We obtain nearly identical results (alpha of 22%) when we add two more factors representing shocks to the level and slope of the variance swap curve.[34] The Sharpe ratios associated with this trading strategy thus do not seem explained by exposure to standard risk factors.

Figure VIII further details the performance of the trading strategy. The upper left panel shows when the strategy calls for a

34. We construct variance swap term structure factors by first calculating monthly returns to variance swaps at all maturities, then extracting the first two principal components from this return panel. We construct alphas with respect to a factor model that includes the Fama-French factors plus the two variance swap factors. See Dew-Becker et al. (2015) for additional details.
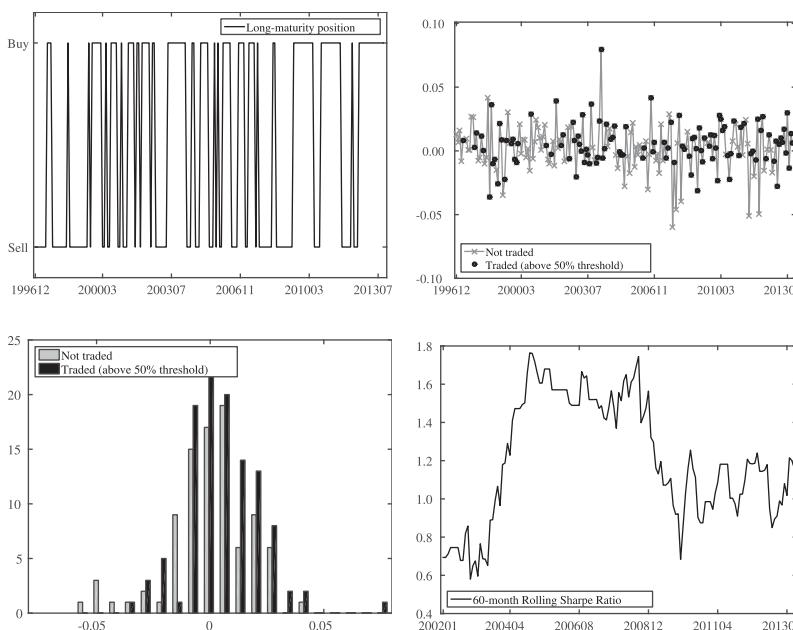
Variance Swap Trading-Strategy Performance

Behavior of one-month holding period returns when the trading strategy focuses on long-end claims with 18 months to maturity and uses a backward-looking mispricing threshold of 50% to determine whether a trade is initiated. The strategy is scaled to have an annual standard deviation of 20%. Clockwise from the upper left, we report the direction of trade in the long-maturity claim, time series of monthly realized returns, rolling 60-month Sharpe ratio, and histogram of realized returns.

buy or a sell position in the long maturity swap. The strategy frequently changes the direction of the trade. In the average month, the long-maturity claim is 26% likely to be traded in the opposite direction from the previous month. This frequent sign switching is the reason the strategy's returns are essentially uncorrelated with standard risk factors.

The upper right panel shows the time series of returns to the strategy. It only trades when the signal is sufficiently strong (when the deviation from the model price is greater than the median historical mispricing). Returns during traded months are shown by black circles, and returns in nontraded (weak signal) months are shown in gray crosses. The histogram for returns in traded

and nontraded months is shown in the lower left panel. Traded returns are positively skewed. While some of the largest losses occur during risky episodes, including a loss of 3.6% in August 1998 amid the Russian default and LTCM crisis and a loss of 3.1% in January 2009, the overall Sharpe ratio during the financial crisis is 0.49. The lower right panel shows subsample annualized Sharpe ratios for the strategy calculated over a 60-month rolling window. No one subsample appears to drive the strategy's overall performance, and the rolling Sharpe ratio never falls below 0.5.

Trading strategy results for variance swaps indicate that an investor who treats the affine model as the true value process and trades against deviations of actual prices from model predictions earns high average returns, and these are not easily explained as compensation for bearing risk. This supports the view that overreaction of long-maturity claims reflects temporary mispricing. Yet it is by no means conclusive evidence of mispricing. It is always possible that high average returns represent compensation for some risk that we have not accounted for in our model. In this case, our trading strategy can be viewed as quantifying the economic importance of risk factors and risk premia that are missed by affine-$\mathbb{Q}$ models.

If the attractive performance of the excess volatility trading strategy is due to mispricing, it is important to understand barriers that prevent arbitrageurs from exploiting and eliminating the anomaly (Shleifer and Vishny 1997). The most natural limits to arbitrage to consider are transactions costs, which can be substantial in an OTC derivatives market such as that for variance swaps. Industry sources suggest that variance swap transaction costs are typically 1% to 2% of the value of a position, consistent with the findings of Avellaneda and Cont (2011). We analyze the strategy's performance assuming trading costs of this magnitude for all legs of the trade (long and short, at initiation and liquidation). We assume that an investor takes these costs into consideration and only initiates a trade when the mispricing is sufficiently large after costs.

Table VII, Panel A reports Sharpe ratios and Panel B reports the fraction of periods in which a trade is triggered for each version of the strategy. Trading costs erode a substantial portion of the strategy's profits. A proportional cost of 2% entirely eliminates the benefit of the one-month holding period strategy, indicating that prices do not converge enough over one month to cover the cost of trading. Convergence improves with longer holding periods of

TABLE VII

TRADING STRATEGY WITH TRANSACTION COSTS

Mispricing percentile	Longest maturity traded	0% TC			1% TC			2% TC		
		1 mo.	3 mo.	6 mo.	1 mo.	3 mo.	6 mo.	1 mo.	3 mo.	6 mo.
Panel A: Sharpe ratio										
50	15	0.73	0.80	0.69	−0.75	0.30	0.56	−2.01	−0.23	0.16
50	18	1.17	1.26	0.98	−0.11	0.77	0.84	−1.38	0.40	0.52
50	21	0.94	1.12	1.10	−0.27	0.69	0.80	−1.49	0.31	0.34
50	24	0.56	0.69	0.49	−0.88	0.22	0.16	−1.95	−0.24	−0.13
75	15	1.43	0.84	1.17	−0.09	0.49	0.86	−1.51	0.35	0.35
75	18	1.68	1.34	1.52	0.50	0.99	1.13	−0.87	1.11	0.73
75	21	1.37	1.46	1.43	0.14	0.97	1.02	−0.91	0.59	0.72
75	24	0.50	0.72	0.63	−0.70	0.23	0.51	−1.47	−0.22	0.18
90	15	1.56	1.82	1.25	−0.08	1.07	1.02	−1.96	0.55	1.33
90	18	1.96	2.26	1.70	1.05	2.28	1.59	−0.69	1.69	1.19
90	21	1.91	2.45	1.54	0.75	2.18	1.20	−0.22	1.49	1.04
90	24	1.61	0.58	0.93	0.17	0.23	0.60	−1.46	0.50	0.54
Average		1.28	1.28	1.12	−0.02	0.87	0.86	−1.33	0.53	0.58
Panel B: Trading frequency										
50	15	0.54	0.50	0.51	0.47	0.44	0.39	0.38	0.34	0.32
50	18	0.55	0.50	0.49	0.47	0.45	0.43	0.41	0.39	0.35
50	21	0.54	0.51	0.49	0.49	0.43	0.45	0.41	0.37	0.38
50	24	0.57	0.52	0.54	0.50	0.44	0.45	0.40	0.38	0.35
75	15	0.33	0.31	0.29	0.30	0.25	0.24	0.24	0.21	0.18
75	18	0.33	0.28	0.30	0.28	0.26	0.24	0.23	0.23	0.22
75	21	0.33	0.28	0.31	0.30	0.25	0.28	0.25	0.22	0.22
75	24	0.33	0.32	0.34	0.27	0.29	0.27	0.21	0.21	0.21
90	15	0.16	0.12	0.14	0.14	0.10	0.11	0.13	0.09	0.06
90	18	0.16	0.14	0.14	0.14	0.10	0.12	0.13	0.09	0.09
90	21	0.15	0.13	0.14	0.14	0.09	0.13	0.13	0.08	0.11
90	24	0.16	0.17	0.14	0.15	0.13	0.11	0.11	0.12	0.10

Notes. Panel A reports annualized Sharpe ratios for variance swap trading strategies that exploit mispricing relative to the affine-$\mathbb{Q}$ model assuming all positions pay a transactions costs (TC) of 0%, 1%, or 2% of the value of the position. We consider holding periods of one month, three months, and six months. Panel B reports the fraction of periods in which mispricings are sufficiently large to trigger a trade.

three or six months, in which cases the Sharpe ratio remains above 0.50 on average after costs. This represents more than a 50% decline from the Sharpe ratio ignoring trading costs and requires that arbitrageurs stomach convergence risk over longer intervals. The table also suggests that, in response to trading costs, an arbitrageur can boost Sharpe ratios by only trading on very large

mispricings (such as those above the 90th percentile). Requiring such a high threshold, however, reduces the number of tradable periods to roughly 1 in 10. This is costly to arbitrageurs whose undeployed capital idly awaits trading opportunities.

In summary, Table VII suggests that excess volatility of long-maturity claims may be perpetuated by limits to arbitrage in the form of transaction costs, infrequent profit opportunities, and long holding periods.

2. A Model of Extrapolation. A number of recent models explore the usefulness of extrapolative expectations in matching asset-pricing phenomena such as excess price volatility in equity and credit markets.[35] These models do not examine how expectation formation varies with the horizon of the expectation, and in particular have not explored the implications that extrapolation may have for excess volatility of long- versus short-maturity claims. Yet given that the affine model's inconsistency stems from long-maturity factor loadings appearing too high—so that the long end of the price curve appears to overreact—extrapolation is a natural candidate for a behavioral bias that might produce systematic mispricing along the term structure. Furthermore, asset markets that have typically been modeled using extrapolation, such as stocks, mortgages, and corporate bonds, are long-duration assets. Excess volatility in these markets is likely to be a phenomenon related to the long-maturity excess volatility that we document in many other markets.

In this section we explore features of extrapolative models that are useful for matching the empirical facts documented in Section III.[36] We focus our analysis on the "natural expectations" framework of Fuster, Laibson, and Mendel (2010), henceforth FLM. Natural expectations are able to generate term structure excess volatility that is both qualitatively and quantitatively consistent with our findings. The ability of this model to fit term structure patterns is remarkable because the natural expectations idea was not conceived with term structure pricing in mind.

---

35. See, for example, Barberis and Shleifer (2003), Greenwood and Shleifer (2014), Barberis et al. (2015a,b), Bordalo, Gennaioli, and Shleifer (2015), and Gennaioli, Shleifer, and Ma (2015).

36. We also provide in Online Appendix I a more general characterization of model misspecification.

Our article therefore provides a test of this model along a previously unexplored dimension.

We first derive new term structure implications from the natural expectations framework, and then calibrate the model to match the salient features of the variance swap term structure. The framework posits that investors price assets using a model that differs from the true data-generating process. Investors construct expectations under both the true model ("rational" expectations) and a more parsimonious but misspecified model ("intuitive" expectations). They then average the two expectations to arrive at their final, "natural," expectation.

We derive term structure prices from the same specification studied in FLM. The true process for the cash flow x_t is an AR(2):

$$(R) \qquad\qquad x_{t+1} = \alpha x_t + \beta x_{t-1} + \eta_{t+1},$$

where α and β are such that x is a persistent but stationary process. We label this equation (R) because it describes the model used to build rational expectations.

The investor's so-called intuitive model is:

$$\Delta x_{t+1} = \phi \Delta x_t + \epsilon_{t+1}.$$

This simplifies the truth by treating the persistence in x as a random walk with an AR(1) adjustment term (i.e., it has one fewer parameter).[37] We can also represent the intuitive model, labeled (I), in levels as an AR(2):

$$(I) \qquad\qquad x_{t+1} = (1 + \phi)x_t - \phi x_{t-1} + \epsilon_{t+1}.$$

The intuitive model of x thus has a unit root, while the true model is stationary. As a result, the intuitive model embeds extrapolative/overreactive beliefs. This is the first fundamental feature of the natural expectations framework: investors treat cash flows as more persistent than they truly are.

---

37. FLM impose an additional restriction on ϕ linked to a specific mechanism through which the investors learn about ϕ from the data. The restriction links ϕ to α and β and therefore removes one further degree of freedom. Since our results do not depend on ϕ, we leave it free in our discussion.

Next, natural expectations are formed as an average of the true and intuitive models with weights given by λ:

$$N_t[x_{t+s}] = \lambda I_t[x_{t+s}] + (1 - \lambda)E_t[x_{t+s}].$$

The notation we use here is the same as in FLM: N_t is the natural expectation, I_t is the expectation under the intuitive process, and E_t is the rational expectation under the true process. $N_t[\cdot]$ forms the basis for valuation—the price of a forward claim is $f_{t,n} = N_t[x_{t+n}]$.

To conveniently analyze claims with different cash flow horizons, we rewrite the AR(2) processes for models (I) and (R) in vector form:

$$y_t = G_R y_{t-1} + \tilde{\eta}_t \qquad \text{and} \qquad y_t = G_I y_{t-1} + \tilde{\epsilon}_t,$$

where

$$y_t \equiv \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix}, \qquad G_R \equiv \begin{bmatrix} \alpha & \beta \\ 1 & 0 \end{bmatrix}, \qquad G_I \equiv \begin{bmatrix} (1+\phi) & -\phi \\ 1 & 0 \end{bmatrix}.$$

From the vector form, it is easy to represent expectations as a function of maturity, n:[38]

$$E_t[x_{t+n}] = [1 \ 0]G_R^n y_t \quad \text{and} \quad I_t[x_{t+n}] = [1 \ 0]G_I^n y_t,$$

so that forward prices are:

$$(19) \qquad f_{t,n} = [1 \ 0]\left(\lambda G_I^n + (1 - \lambda)G_R^n\right)y_t.$$

This equation highlights the second fundamental feature of the natural expectations framework. In affine models, expectations are formed using a single model for the entire term structure. With natural expectations, investors average expectations constructed from models with contradictory persistences. This second feature is the key mechanism that allows natural expectations to replicate the internal inconsistency in short- and long-maturity prices relative to the affine framework that we document in the data. If investors formed expectations using only rational expectations

38. The reader may notice that the transition matrixes of the factors, G_R and G_I, are not diagonal as in the representation we use in Section II. This is without loss of generality. The model can be rotated into the diagonal representation we use, as discussed in Joslin, Singleton and Zhu (2011).

(R) or only using intuitive expectations (I), they would be using an affine model and would satisfy standard consistency along the term structure, because their forecasts adhere to the recursive relations

$$E_t[y_{t+s}] = G_R E_t[y_{t+s-1}] \qquad \text{and} \qquad I_t[y_{t+s}] = G_I I_t[y_{t+s-1}].$$

So, when $\lambda = 0$ or $\lambda = 1$, variance ratios will be exactly one throughout the entire term structure. When instead $0 < \lambda < 1$, the model mixes two inconsistent sets of dynamics. The true cash flow dynamics will dominate the short end of the curve, and the (more persistent) intuitive dynamics will dominate the long end of the curve. Our variance ratio test, which compares the dynamics implied from the short end to those implied by the long end, is designed to identify this type of nonaffinity. In fact, long-maturity variance ratios are always greater than one when $0 < \lambda < 1$ in the FLM model.

3. Model Calibration. We calibrate the model and then ask how well it matches excess volatility observed in the variance swap data. Parameters are chosen to minimize the distance between the factor loadings estimated in regression-based tests, and the factor loadings implied by the natural expectations model.[39] Estimates of the model's four parameters are $\alpha = 0.90$, $\beta = -0.08$, $\phi = 0.12$, and $\lambda = 0.36$. While derived in an entirely separate context, our estimates are close to the parameter values chosen by FLM.[40]

The left panel of Figure IX overlays variance ratios generated from the calibrated natural expectations model (solid line) onto those estimated in the regression-based tests of Figure I. That is, we repeat the variance calculations of Figure I when the true data-generating process is the calibrated natural expectations model.

---

39. As in the affine model, claims prices in the two-factor natural expectations model can be represented via equation (9), where each set of loadings β_j is a function of only natural expectations model parameters α, β, ϕ, λ, and the given maturity, j. Parameters are then estimated via GMM, using OLS regression estimates $\hat{\beta}_j$ throughout the variance swap term structure as moments and using an identity weighting matrix. We estimate the four parameters using eight moment conditions: the regression loadings of maturities 3, 6, 12, 24 onto each of the two short-end prices (maturities 1 and 2).

40. In their analysis of macroeconomic data, FLM use parameters $\alpha = 1.16$, $\beta = -0.24$, $\phi = 0.20$, and $\lambda = 0.50$.
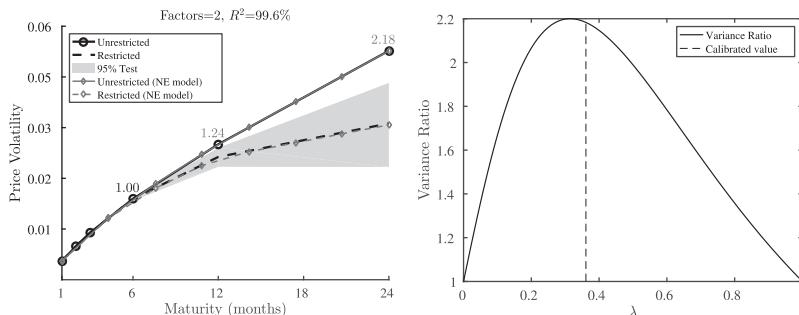
Natural Expectations Calibration

The left panel compares implied variance ratios from the calibrated natural expectations model to the regression-based estimates from Figure I. The red lines ("NE model"; color artwork available at the online version of this article) shows the unrestricted and restricted variances and the variance ratio test obtained from the calibrated natural expectations model, where the investor mixes rational and intuitive expectations with weighting parameter $\lambda = 0.36$. The right panel shows the fitted 24-month variance ratio in the natural expectations model as a function of λ, holding other parameters fixed at $\alpha = 0.90$, $\beta = -0.08$, and $\phi = 0.12$.

The model fits the unrestricted and restricted factor loadings very closely, and therefore implies variance ratios similar to the ones we estimate in the data at all maturities. In particular, the calibrated natural expectations model generates a variance ratio of 2.18 at 24 months, versus a ratio of 2.15 in the data. This plot demonstrates that the natural expectations model can accurately fit the empirical excess volatility patterns.

The key role of expectation mixing for producing high variance ratios is further illustrated in the right panel. We plot the model-predicted 24-month variance ratio as a function of the mixing parameter λ. At $\lambda = 0$ or 1, the natural expectations model is indeed affine and the variance ratio is exactly one. In between, affinity is violated due to model averaging, and extrapolative beliefs drive the variance ratio above one.

V. DISCUSSION AND CONCLUSIONS

We find that prices of long-maturity claims are far more variable than justified by standard models. Our tests of excess volatility exploit the strict overidentification restrictions from term structure asset pricing, in which prices at all maturities

are linked by the law of iterated values and the implied dynamics of the factors driving cash flows. We use the short end of the term structure to learn the implied cash flow dynamics perceived by investors under the pricing measure, $\mathbb{Q}$, and reject the hypothesis that estimated short-end behavior is consistent with prices at long maturities.

Our findings suggest that the puzzle of excess volatility is a pervasive phenomenon, manifesting in a wide variety of markets including those for equity and currency volatility, sovereign and corporate default risk, commodities, interest rates, and inflation. Excess volatility relative to the affine model cannot be explained by time variation in discount rates, as this is accounted for in our estimation of risk-neutral model dynamics.

We show that all asset classes deviate from the model in the same way, with long-maturity claims nearly perfectly correlated with, but overreacting to, fluctuations in short-maturity prices. We also investigate a number of well-studied nonaffine models, none of which appear to capture the behavior of long-maturity claims in the data. We show that trading against long-maturity excess volatility appears profitable after adjusting for exposure to standard risk factors.

We interpret these facts as violations of no-arbitrage restrictions in an affine model. Another potential interpretation of our facts, however, is that apparent affine-model violations reflect the presence of transient risk premia that differentially affect prices along the maturity structure. In this case, the profits that accrue to our trading strategy are can be viewed as evidence of such risk premia.[41]

Last, we study a model of investor extrapolation that is quantitatively consistent with the excess volatility patterns that we document. Models in which expectations are distorted by extrapolation are increasingly prominent in the literature. The exact form that extrapolation takes, however, can vary widely across models. There are differences in the kinds of processes that agents extrapolate. In some cases agents extrapolate fundamentals such as productivity, cash flows, and consumer demand (e.g., Fuster, Hebert, and Laibson 2011; Alti and Tetlock 2014; Greenwood and Hanson 2015; Hirshleifer, Li, and Yu 2015), in other cases they extrapolate risk (e.g., Jin 2015), and in others investors directly extrapolate prices and returns (e.g., Hong and Stein 1999; Barberis

---

41. We thank our referee for suggesting this interpretation of our findings.

et al. 2015b). There are differences in the microfoundations of extrapolation: some are founded on the representativeness heuristic (e.g., Bordalo, Gennaioli, and Shleifer 2015; Gennaioli, Shleifer, and Vishny 2015), others motivated by the availability heuristic and recency bias (e.g., Lansing 2006), and some built on investors using oversimplified models (e.g., Fuster, Laibson, and Mendel 2010).

That extrapolation models do not enforce the internal consistency of rational expectations makes them prime candidates for describing the patterns we document. However, these models are not guaranteed to violate affine-model restrictions. For a model to match the basic long-maturity excess-volatility facts, it must produce a term structure of prices (i) that satisfies a linear factor structure and (ii) whose factor loadings at different maturities diverge from the geometric progression that is the signature of affine models. Many extrapolative models easily satisfy the first requirement—a strict factor structure for term structure prices— by virtue of linear dynamics in the models' driving processes. However, these models also typically imply an affine term structure because their factor loadings follow a geometric progression as in equation (7). Although investors subjectively believe that cash flows are more persistent than they truly are, they nonetheless respect the law of iterated expectations and this implies that loadings are geometric. Greenwood and Hanson (2015) and Bordalo, Gennaioli, and Shleifer (2015) are two examples in which, despite deviating from rational expectations, investor beliefs imply term structures that satisfy the internal consistency conditions of equation (7) and thus do not explain the facts we document.

For a model to give a term structure that is linear in factors, but with loadings that follow a nongeometric progression, the law of iterated expectations must break down. The natural expectations model is one example in which the term structure follows a linear factor model but at the same time violates the law of iterated expectations. It is the act of averaging forecasts from two models with different degrees of persistence that breaks the affine structure and generates an internal inconsistency along the term structure.

It is especially interesting that this inconsistency is difficult to see when studying the behavior of a "single maturity" asset such as the stock market. The effects of natural expectations become clearly evident once term structure implications are drawn. Last, the main message from this analysis is not that the

natural expectations framework is the only model that can match the data but to point out a key ingredient—model averaging—that allows investor expectations and thus model prices to match the excess volatility patterns of the data.

Our exploration into the theoretical origins of excess volatility only scratches the surface of what we believe is a promising future research direction. In particular, our findings call for more investigation into how agents form expectations over multiple horizons and the extent to which investor behavior is consistent with the law of iterated values.

YALE UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH
UNIVERSITY OF CHICAGO AND NATIONAL BUREAU OF ECONOMIC RESEARCH

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online. Data and code replicating the tables and figures in this article can be found in Giglio and Kelly (2017), in the Harvard Dataverse, doi:10.7910/DVN/JA8CFG.

REFERENCES

Ait-Sahalia, Yacine, Mustafa Karaman, and Loriano Mancini, "The Term Structure of Variance Swaps and Risk Premia," Working paper, Princeton University, 2015.

Alti, Aydoğan, and Paul C. Tetlock, "Biased Beliefs, Asset Prices, and Investment: A Structural Approach," *Journal of Finance*, 69 (2014), 325–361.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, "Modeling and Forecasting Realized Volatility," *Econometrica*, 71 (2003), 579–625.

Anderson, Evan W., Lars Peter Hansen, and Thomas J. Sargent, "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection," *Journal of the European Economic Association*, (2003), 68–123.

Ang, Andrew, and Monika Piazzesi, "No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50 (2003), 745–787.

Avellaneda, Marco, and Rama Cont, "Transparency in OTC Equity Derivatives Markets: A Quantitative Study," *Finance Concepts* (2011).

Backus, David, Nina Boyarchenko, and Mikhail Chernov, "Term Structures of Asset Prices and Returns," Working paper, UCLA, 2015.

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

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, "Extrapolation and Bubbles," Unpublished manuscript, Yale University, 2015a.

——, "X-CAPM: An Extrapolative Capital Asset Pricing Model," *Journal of Financial Economics*, 115 (2015b), 1–24.

Barberis, Nicholas, and Andrei Shleifer, "Style Investing," *Journal of Financial Economics*, 68 (2003), 161–199.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Investor Psychology and Credit Cycles," Working paper, Harvard Univeristy, 2015.

Britten-Jones, Mark, and Anthony Neuberger, "Option Prices, Implied Price Processes, and Stochastic Volatility," *Journal of Finance*, 55 (2000), 839–866.

Campbell, John Y., "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18 (1987), 373–399.

———, "A Variance Decomposition for Stock Returns," *Economic Journal*, 101 (1991), 157–179.

———, "Some Lessons from the Yield Curve," *Journal of Economic Perspectives*, 9 (1995), 129–152.

Campbell, John Y., and John H. Cochrane. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107 (1999), 205–251.

Campbell, John Y., and Robert J. Shiller, "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95 (1987), 1062–1088.

———, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1 (1988a), 195–228.

———, "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance*, 43 (1988b), 661–676.

———, "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies*, 58 (1991), 495–514.

Carr, Peter, and Roger Lee, "Volatility Derivatives," *Annual Review of Financial Economics*, 1 (2009), 319–339.

Cecchetti, Stephen G., Pok-sang Lam, and Nelson C. Mark, "Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to Be True?," *American Economic Review*, 90 (2000), 787–805.

Christensen, Jens H. E., and Jose A. Lopez, "Common Risk Factors in the US Treasury and Corporate Bond Markets: An Arbitrage-Free Dynamic Nelson-Siegel Modeling Approach," Manuscript, Federal Reserve Bank of San Francisco, 2008.

Cochrane, John H., "Explaining the Variance of Price-Dividend Ratios," *Review of Financial Studies*, 5 (1992), 243–280.

———, "The Dog That Did Not Bark: A Defense of Return Predictability," *Review of Financial Studies*, 21 (2008), 1533–1575.

———, "Discount Rates," *Journal of Finance*, 66 (2011), 1047–1108, AFA Presidential Address.

Cochrane, John H., and Monika Piazzesi, "Decomposing the Yield Curve," Working paper, University of Chicago, 2009.

Dai, Qiang, and Kenneth J. Singleton, "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics*, 63 (2002), 415–441.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez, "The Price of Variance Risk," Working paper, 2015.

Duffee, Gregory R., "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57 (2002), 405–443.

Duffie, Darrell, Jun Pan, and Kenneth Singleton, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica* (2000), 1343–1376.

Duffie, Darrell, and Kenneth J. Singleton, "Modeling Term Structures of Defaultable Bonds," *Review of Financial Studies*, 12 (1999), 687–720.

Egloff, Daniel, Madrkus Leippold, and Liuren Wu, "The Term Structure of Variance Swap Rates and Optimal Variance Swap Investments," *Journal of Financial and Quantitative Analysis*, 45 (2010), 1279–1310.

Fama, Eugene F., "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25 (1970), 383–417.

———, "Efficient Capital Markets: II," *Journal of Finance*, 46 (1991), 1575–1617.

Fama, Eugene F., and Robert R. Bliss, "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77 (1987), 680–692.

Fama, Eugene F., and Kenneth R. French, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33 (1993), 3–56.

Fleckenstein, Matthias, Francis A. Longstaff, and Hanno Lustig, "Deflation Risk," NBER Technical report, 2013.

Fleming, Michael J., and John Sporn, "Trading Activity and Price Transparency in the Inflation Swap Market," *Economic Policy Review*, 19 (2013).

Fuster, Andreas, Benjamin Hebert, and David Laibson, "Natural Expectations, Macroeconomic Dynamics, and Asset Pricing," NBER Technical report, 2011.

Fuster, Andreas, David Laibson, and Brock Mendel, "Natural Expectations and Macroeconomic Fluctuations," *Journal of Economic Perspectives*, 24 (2010), 67–84.

Gennaioli, Nicola, Andrei Shleifer, and Yueran Ma, "Expectations and Investment," Working paper, Harvard University, 2015.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, "Neglected Risks: The Psychology of Financial Crises," *American Economic Review*, 105 (2015), 310–314.

Giglio, Stefano, and Bryan Kelly, "Replication Data for: 'Excess Volatility: Beyond Discount Rates'," *Harvard Dataverse* (2017), doi:10.7910/DVN/JA8CFG.

Giglio, Stefano, Matteo Maggiori, and Johannes Stroebel, "Very Long-Run Discount Rates," *Quarterly Journal of Economics*, 130 (2015), 1–53.

——, "No-Bubble Condition: Model-free Tests in Housing Markets," *Econometrica*, 84 (2016), 1047–1091.

Glaeser, Edward L., and Charles G. Nathanson, "An Extrapolative Model of House Price Dynamics," NBER Technical report, 2015.

Granger, Clive W. J., and Roselyne Joyeux, "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1 (1980), 15–29.

Granger, Clive W. J., and Timo Teräsvirta, *Modelling Nonlinear Economic Relationships* (Oxford: Oxford University Press, 1993).

Greenwood, Robin, and Samuel G. Hanson, "Waves in Ship Prices and Investment," *Quarterly Journal of Economics*, 55 (2015), 109.

Greenwood, Robin, and Andrei Shleifer, "Expectations of Returns and Expected Returns," *Review of Financial Studies* (2014), hht082.

Gurkaynak, Refet S., Brian Sack, and Eric Swanson, "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models," *American Economic Review*, 95 (2005), 425–436.

Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright, "The U.S. Treasury Yield Curve: 1961 to the Present," Federal Reserve Board Finance and Economics Discussion Series paper 2006-28, 2006.

Hamilton, James D., and Cynthia Wu, "Identification and Estimation of Gaussian Affine-Term-Structure Models," *Journal of Econometrics*, 168 (2012), 315–331.

Hansen, Lars Peter, "Dynamic Valuation Decomposition within Stochastic Economies," *Econometrica*, 80 (2012), 911–967.

——, "Nobel Lecture: Uncertainty Outside and Inside Economic Models," *Journal of Political Economy*, 122 (2014), 945–987.

Hansen, Lars Peter, and Scott F. Richard, "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica* (1987), 587–613.

Hansen, Lars Peter, and Thomas J. Sargent, "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2 (1980), 7–46.

Hansen, Lars P., and Jose A. Scheinkman, "Long-Term Risk: An Operator Approach," *Econometrica*, 77 (2009), 177–234.

Hanson, Samuel G., and Jeremy C Stein, "Monetary Policy and Long-Term Real Rates," *Journal of Financial Economics*, 115 (2015), 429–448.

Hayashi, Fumio, *Econometrics* (Princeton, NJ: Princeton University Press, 2000).

Hirshleifer, David, Jun Li, and Jianfeng Yu, "Asset Pricing in Production Economies with Extrapolative Expectations," *Journal of Monetary Economics*, 76 (2015), 87–106.

Hong, Harrison, and Jeremy C. Stein, "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets," *Journal of Finance*, 54 (1999), 2143–2184.

Jiang, George J., and Yisong S. Tian, "The Model-Free Implied Volatility and its Information Content," *Review of Financial Studies*, 18 (2005), 1305–1342.

Jin, Lawrence J., "A Speculative Asset Pricing Model of Financial Instability," SSRN 2524762, 2015.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu, "A New Perspective on Gaussian Dynamic Term Structure Models," *Review of Financial Studies*, 24 (2011), 926–970.

Lansing, Kevin J., "Lock-in of Extrapolative Expectations in an Asset Pricing Model," *Macroeconomic Dynamics*, 10 (2006), 317–348.

Le, Anh, Kenneth J Singleton, and Qiang Dai, "Discrete-Time AffineQ Term Structure Models with Generalized Market Prices of Risk," *Review of Financial Studies*, 23 (2010), 2184–2227.

LeRoy, Stephen F., and Richard D. Porter, "The Present-Value Relation: Tests Based on Implied Variance Bounds," *Econometrica*, 49 (1981), 555–574.

Madan, Dilip, and Haluk Unal, "A Two-Factor Hazard Rate Model for Pricing Risky Debt and the Term Structure of Credit Spreads," *Journal of Financial and Quantitative Analysis*, 35 (2000), 43–65.

Piazzesi, Monika "Affine Term Structure Models," in *Handbook of Financial Econometrics,* vol. 1 (Amsterdam: Elsevier, 2010), 691–766.

Samuelson, Paul A., "Proof that Properly Anticipated Prices Fluctuate Randomly," *Industrial Management Review*, 6 (1965), 41–49.

Shiller, Robert J., "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," *Journal of Political Economy*, 87 (1979), 1190–1219.

——, "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, 71 (1981), 421–436.

Shleifer, Andrei, and Robert W. Vishny, "The Limits of Arbitrage," *Journal of Finance*, 52 (1997), 35–55.

Siriwardane, Emil N., "Concentrated Capital Losses and the Pricing of Corporate Credit Risk," Harvard Business School Finance Working Paper, 2015.

Stein, Jeremy, "Overreactions in the Options Market," *Journal of Finance*, 44 (1989), 1011–1023.

Teräsvirta, Timo, "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89 (1994), 208–218.

van Binsbergen, Jules H., Michael W. Brandt, and Ralph Koijen, "On the Timing and Pricing of Dividends," *American Economic Review*, 102 (2012), 1596–1618.

van Binsbergen, Jules, Wouter Hueskes, Ralph Koijen, and Evert Vrugt, "Equity yields," *Journal of Financial Economics*, 110 (2013), 503–519.

Wachter, Jessica A, "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?," *Journal of Finance*, 68 (2013), 987–1035.