Excess Volatility: Beyond Discount Rates*

Stefano Giglio Bryan Kelly

University of Chicago and NBER

January 6, 2016

Abstract

We document a form of excess volatility that is irreconcilable 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. Our findings are pervasive. 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.

*We are grateful to Robert Barro, Jonathan Berk, Oleg Bondarenko, John Campbell, John Cochrane, Josh Coval, Drew Creal, Ian Dew-Becker, Hitesh Doshi, Gene Fama, Xavier Gabaix, Lloyd Han, Valentin Haddad, Lars Hansen, Lawrence Jin, Ralph Koijen, Ahn Le, Martin Lettau, Hanno Lustig, Matteo Maggiori, Tim McQuade, Stavros Panageas, Monika Piazzesi, Seth Pruitt, Martin Schneider, Andrei Shleifer, Jeremy Stein, Pietro Veronesi, and Cynthia Wu for helpful comments, seminar participants at ASU, Berkeley, Case Western, Chicago, Harvard, Houston, Stanford, and UT Dallas, and conference participants at CITE and IFSID.

1 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 expectations, 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 much (or too little) covariation between long and short maturity prices, relative to the covariation allowable within a model, 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 paper, we document a form of excess volatility in prices along the term structure that is irreconcilable 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 both statistically and economically. Perhaps most interestingly, 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 default risk, commodities, and inflation. Only for the term structure of Treasuries do we find that violations of model restrictions are economically small, consistent with the findings of a large literature on interest rate models.[3]

We define as "standard" any model in which cash flows are driven by a vector autoregression under the risk-neutral pricing measure, a class of models that we refer to as "affine-$\mathbb{Q}$."[4] This class encompasses many leading asset pricing paradigms, from structural equilibrium models with long run risks (Bansal and Yaron, 2004) or variable rare disasters (Wachter,

[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).

[2] For example, a linear three-factor model explains the panel of Treasury interest rates for maturities of one up to thirty years with an R^2 in excess of 99%.

[3] See, for example, Joslin, Singleton and Zhu (2011).

[4] Common notation refers to the real-world statistical measure as "$\mathbb{P}$" and the risk-adjusted pricing measure as "$\mathbb{Q}$." Section 2 discusses the origins and interpretation of the $\mathbb{Q}$ measure in detail.

2013),[5] to reduced-form models ubiquitous in fixed income and derivatives pricing (Duffie, Pan and Singleton, 2000). The affine-$\mathbb{Q}$ class has become pervasive precisely due to its convenience in delivering closed-form solutions in diverse valuation settings.[6]

We focus on the risk-neutral, or "$\mathbb{Q}$," representation of structural and reduced-form models based on a feature that is crucial for thinking about excess price volatility. By its definition, the $\mathbb{Q}$ measure incorporates all potential variation in discount rates.[7] Therefore, any inference regarding price volatility based on estimates of the $\mathbb{Q}$ measure explicitly accounts for discount rate behavior. This is in contrast to the notion of excess volatility famously documented by Shiller (1979, 1981) and others, 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 that lies at the foundation of leading frameworks in modern finance.[8] By using the $\mathbb{Q}$ representation of models in our analysis, any excessive volatility that we document must be coming from sources other than the types of discount rate variation that can be represented within an affine-$\mathbb{Q}$ model. In short, we choose the affine-$\mathbb{Q}$ specification as the null model for our analysis based on its great flexibility for nesting many leading economic frameworks and because it explicitly accounts for what has become the de facto explanation for excess volatility, time-varying discount rates.

In addition to estimating the magnitude and pervasiveness of excess volatility, we also characterize the specific nature of the affine-$\mathbb{Q}$ violation and show that all asset classes deviate from the model in the same way. We also show that the data is inconsistent with several non-linear models that have been studied in the literature. Finally, we find evidence that trading against long maturity excess volatility is profitable, even after adjusting for exposure to standard risk factors.

[5] In Appendix C we discuss affine structural models in more detail. The long run risks and disaster models are affine under the assumption of unit intertemporal elasticity of substitution (IES). Yet even with non-unit IES, researchers typically analyze these models in their loglinearized form so that the models predictions are studied under an affine approximation. Appendix C also discusses affine models with learning.

[6] Furthermore, the affine-$\mathbb{Q}$ class nests a wide range of dynamics by allowing the data to choose the appropriate number of driving factors. Its flexibility accurately approximates non-linear models as well, a property that we illustrate empirically in Section 4.

[7] 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.

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

1.1 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. Under the pricing measure $\mathbb{Q}$, cash flows evolve according to

$$x_t = \rho x_{t-1} + \Gamma \epsilon_t.$$

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 $f_{t,n} = E_t^{\mathbb{Q}}[x_{t+n}]$. The term structure of forward prices at maturities $1, ..., N$ is therefore given by

$$f_{t,1} = \rho x_t, \quad f_{t,2} = \rho^2 x_t, \quad ..., \quad f_{t,N} = \rho^N x_t. \tag{1}$$

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 ρ (the parameter governing cash flow dynamics under $\mathbb{Q}$). This restriction is equivalently represented with prices of *cumulative* claims, defined as $p_{t,n} = E_t^{\mathbb{Q}}[x_{t+1} + ... + x_{t+n}]$, in which case the term structure takes the form:

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

Tests of the model's restrictions hinge on an estimate of ρ. Fortunately, ρ is easily estimated from regressions of prices onto prices. For example, let the first maturity claim 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, $f_{t,2}$, onto $f_{t,1}$. According to Equation (1), b_2 exactly recovers ρ. This regression is intuitive. The relative valuation of the first two claims perfectly 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, which indicates that ρ is near one under the investors' subjective pricing measure.

If we project prices for remaining maturities $3, ..., N$ onto the short price, we recover a sequence of regression coefficients denoted $b_3, ..., b_N$ that are "unrestricted" in the sense that they are not forced to be jointly determined by ρ according to (1). At the same time, these regressions can be recast in their "restricted" form, where the restriction in (1) relates, for

example, b_N to b_2 by:

$$b_N = (b_2)^{N-1}. \tag{2}$$

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 (2). Under the null model, the restricted and unrestricted variances are the same and $VR_N = 1$. If the ratio statistic significantly exceeds one, the price at maturity N varies more than is justified by the behavior of the short end of the term structure. Note that 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 2, we develop an estimation and inference approach for VR_N in fully general K-factor affine specifications.

1.2 A Representative Term Structure

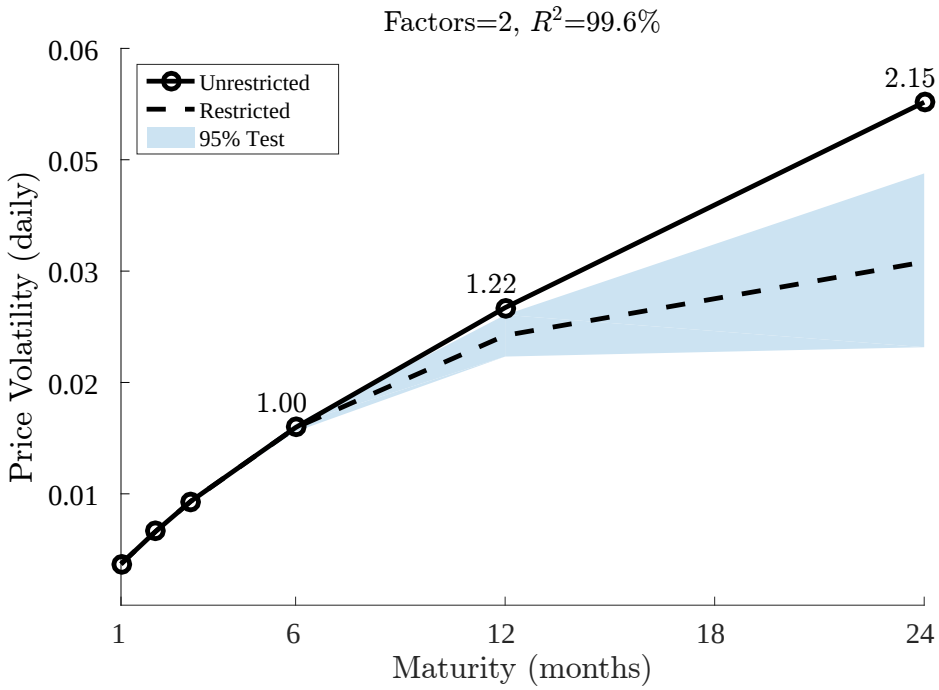
Figure 1 illustrates the behavior of variance ratios in one of our datasets—the term structure of variance swaps—which are claims to the cumulative variance of the S&P 500 index over the life of the contract.[9] An unrestricted linear two-factor model provides an excellent description of the term structure, delivering an R^2 of 99.6% for the panel of prices.[10] 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 blue shaded region represents the point-wise 95% bootstrap confidence interval for price variance in the restricted model.

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

[9] These data are described in detail in Section 3.

[10] This panel R^2 is computed as the fraction of the total variance explained by the first two principal components.

Figure 1: Variance Swap Tests



Note. 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. The numbers next to each circle are the Variance Ratios at each maturity. The shaded area encloses the 97.5^{th} and 2.5^{th} percentiles of the model-implied variance in bootstrap simulations. The left axis reports the volatility of prices.

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. Note that both curves represent "explained" price volatilities from regressions of long prices onto short prices. That is, both the unrestricted and restricted model describe the portion of long maturity price fluctuations that is captured by behavior at the short end. The high variance ratio therefore indicates that the prices at the long end of the curve react to the short end much more strongly in the data than affine model dynamics allow.

The excess volatility of long maturity claims cannot be explained by movements in discount rates. Any discount rate variation that is describable within the affine class is subsumed by our model. Our price-on-price regressions estimate the dynamics of the latent factors under the pricing measure, and we allow the data to determine the appropriate number of factors driving the term structure. This gives our approach the flexibility to consistently es-

timate any specification in the affine-$\mathbb{Q}$ class, regardless of whether the factors are driven by discount rate variation or physical cash flows. Nor can high variance ratios be explained by 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 the loadings implied by model restrictions.

Behavior of the variance swap term structure is representative of our broader empirical findings. In all asset classes that we study we document excess volatility of long maturity prices similar to that in Figure 1. Only in the term structure of Treasuries is excess volatility economically small (though statistically significant), indicating that affine models are especially well suited to the interest rate market, a finding that we return to below.

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

In the last part of the paper, we investigate how the sources of excess volatility should be "split between market inefficiency," i.e. 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 4, we examine four potential explanations for our findings: omitted factors, non-linear dynamics, long memory dynamics, and temporary mispricing of long maturity claims.[11]

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 one. However, omitted factors are unlikely to explain our findings because an unrestricted factor model explains more than 99% of the variation in each term structure we study.

[11] We also consider there the role of measurement error in our empirical tests, and show that it is a quantitatively unviable explanation of our findings.

We show via simulation that it is essentially infeasible for an omitted factor to generate a variance ratio far above one while at the same time producing an unrestricted R^2 over 99%. Additionally, we conduct robustness checks that gradually increase the number of factors used in our tests. This pushes the factor model R^2 even closer to 100% yet still produces variance ratios significantly in excess of one.

Second, we explore a large class of non-linear 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 one. For the most extreme non-linear 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.

Third, 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 non-stationary range, we show that it is possible to generate variance ratios as high as three at the 24 month maturity. But when we allow for one or two extra factors, the variance ratios again shrink to one, which is inconsistent with the behavior we find in the data.

Finally, we construct a trading strategy to explore the possibility of mispricing as a potential driver of excess volatility, and to quantify the economic magnitude of the deviation from the affine-$\mathbb{Q}$ specification. The strategy posits that the estimated affine model reflects the true value of claims, so that any excessive fluctuations of long maturity prices are temporary and potentially exploitable mispricings. The trade is implemented by buying (selling) long maturity claims when they are undervalued (overvalued) relative to the affine model, and hedges this position by selling (buying) short maturity claims in the exact proportion dictated by the estimated model. If there is no mispricing in the true data generating process, then we expect the trading strategy to 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.

In the variance swap market, we find that the trading strategy yields an annualized out-of-sample Sharpe ratio of 1.2 on average, and is not explained by exposure to standard risk factors. The strategy's performance is not driven by any single subsample, and its largest losses are unassociated with the two recessions in our sample (the Great Recession and the 2001 recession). This is not conclusive evidence of mispricing—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.

1.4 Literature Review

An important predecessor of our paper—especially given its focus on the term structure of volatility—is Stein (1989), who compares the volatility of short and long maturity S&P 100 options. He finds excess volatility of one-year option prices and interprets it as evidence of investor overreaction. Our paper 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, while our approach allows for an arbitrary number of factors and extends to a wide range of asset classes.[12]

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. As in Shiller (1979, 1981) and Stein (1989), this reference model does not account for rational discount rate variation. More recently, Hanson and Stein (2015) study overreaction at the long end of the Treasury yield curve focusing on FOMC announcement days. An interesting distinction from our work is that long maturity Treasury rates exhibit by far the *least* excess volatility among the asset classes we study.

The Treasury yield curve has been subject of a large literature. Early contributions by Shiller (1979) and Singleton (1980) demonstrated excess volatility of long-term bonds relative to the expectations hypothesis model, while later literature has worked extensively with affine-$\mathbb{Q}$ specifications that explicitly account for time variation in discount rates. For a review and recent contributions, see for example Ang and Piazzesi (2003), Dai and Singleton (2002), Duffee (2002), Le, Singleton and Dai (2010), Piazzesi (2010), and Joslin, Singleton and Zhu (2011). That literature has typically found that affine no-arbitrage restrictions hold quite well in the interest rate market. We confirm this fact by showing that model

[12] Pontiff (1997) documents excess volatility of closed-end mutual funds that also rules out discount rate variation as an explanation.

violations in the Treasury market are economically small compared to violations in derivatives markets.[13]

Our evidence lends support to recent efforts to understand the key role of expectations formation in financial markets (for example, Hansen, 2014; Greenwood and Shleifer, 2014; Barberis et al., 2015*a*,*b*; Bordalo, Gennaioli and Shleifer, 2015; Gennaioli, Shleifer and Ma, 2015). Our trading strategy analysis in Section 4.5 suggests there may be high costs borne by investors who overreact due to extrapolative expectations or other belief distortions. While observing prices and their comovement allows us to detect overreaction at the long end of the curve, it does not allow us to determine the underlying mechanism driving this overreaction. Our findings highlight a potentially fruitful setting for future research into how agents form expectations over multiple horizons.

2 Asset Term Structures in Linear Models

In this section we develop our general approach to testing the internal consistency of asset term structures in the affine-$\mathbb{Q}$ setting.

2.1 Claims by Maturity

Our focus is on the joint price behavior of claims to the same underlying cash flow process but with different maturities. Let x_t denote a scalar cash flow. For most of our analysis, we focus on linear claims to the x_t process. We study the extension to exponential-linear claims in Section 2.5.1.

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 that no-arbitrage is satisfied, thus the n-maturity forward price is representable as

$$f_{t,n} = E_t^{\mathbb{Q}}\left[x_{t+n}\frac{S_t}{S_{t+n}}\right] \tag{3}$$

[13] Our focus is on volatility of prices at different maturities. A distinct and growing 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*a*,*b*) 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.

where S_t is the value of a risk-free account that pays the instantaneous short-term rate. In our empirical analysis, risk free rate variation is negligible compared to risky asset price variation in almost all asset classes.[14] So, to reduce notation in the remainder of this section, we assume that S_t is constant and equal to one. We return to a detailed analysis of risk-free rates and associated robustness checks in Appendix D.

The pricing of forward claims is easily recast in terms of 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,

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

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,

$$f_{t,n} = E_t^{\mathbb{Q}}\left[x_{t+n}\right] = E_t^{\mathbb{Q}}\left[E_{t+1}^{\mathbb{Q}}[x_{t+n}]\right] = E_t^{\mathbb{Q}}[f_{t+1,n-1}].$$

2.2 Cash Flow Dynamics Under $\mathbb{Q}$

Affine models assume that the cash flow process obeys a linear factor structure, and that these factors evolve as a vector autoregression (VAR) under $\mathbb{Q}$. In particular, let H_t be a vector of K factors with $\mathbb{Q}$-dynamics given by

$$H_t = \rho H_{t-1} + \Gamma \epsilon_t. \tag{4}$$

Under $\mathbb{Q}$, the $K \times K$ parameter matrices ρ and Γ govern transition probabilities and ϵ_t is mean zero and orthogonal to H_{t-1}. Cash flows are determined by the factors according to

$$x_t = \delta_0 + \delta_1' H_t$$

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

Since the factors, H_t, are latent in our setting, model identification requires a normalization of model parameters. We impose the normalization that the matrix ρ is diagonal

[14] The obvious exception is the Treasury bond market, in which case we account for risk-free rate variation in the standard way.

(so that its diagonal elements, which are also its eigenvectors, directly reveal factors' decay rates under $\mathbb{Q}$), and that δ_1 is a vector of ones. These identification assumptions impose no economic restrictions, but ensure 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).[15]

Finally, for notational ease in this section, we set $\delta_0 = 0$. This is not without a loss of generality, as δ_0 determines the overall price level of claims in the term structure. But, for our purpose of understanding the volatility of prices, δ_0 is constant and eventually drops from our analysis. We can now rewrite the cash flow as

$$x_t = \mathbf{1}' H_t. \tag{5}$$

We refer to the class of models satisfying Equations (4) and (5) as "affine-$\mathbb{Q}$." Models with this structure are ubiquitous in the asset pricing literature due to their convenience for describing prices of linear (and exponentially-linear) cash flow claims.

2.3 Term Structure of Prices

Given (4) and (5), the price of a linear forward claim with maturity n is

$$f_{t,n} = \mathbf{1}' \rho^n H_t. \tag{6}$$

Equation (6) contains a set of cross-equation restrictions implied by the affine-$\mathbb{Q}$ model. Prices at all maturities must obey a strict factor structure so that any and all comovement among prices must be due to H_t. Furthermore, the loadings at each maturity must abide by a specific structure—they must follow a geometric progression in ρ. It is also easy to see that this specification satisfies the law of iterated expectations:

$$E_t^{\mathbb{Q}}[f_{t+1,n-1}] = E_t^{\mathbb{Q}}[\mathbf{1}' \rho^{n-1} H_{t+1}] = \mathbf{1}' \rho^{n-1} E_t^{\mathbb{Q}}[H_{t+1}] = \mathbf{1}' \rho^{n-1} \rho H_t = \mathbf{1}' \rho^n H_t = f_{t,n}.$$

Our empirical test investigates the extent to which observed term structures adhere to the model restrictions.

[15] In the exponential-affine setting, we require the additional assumption that ϵ_t is Gaussian. In this setting, the covariance matrix of the errors, Γ, is the same under $\mathbb{P}$ and $\mathbb{Q}$.

2.4 A Convenient Recursive Representation

Equation (6) implies that the latent factors H_t are exactly recoverable from any set of K prices (either forwards or cumulative claims) at different maturities. In turn, this also implies that the price at any maturity j can be represented as an exact linear function of a set of K different prices at any maturity other than j.

In particular, denote the $K \times 1$ vector of time t prices for forwards with maturities 1 to K as $F_{t,1:K} = (f_{t,K}, ..., f_{t,1})'$, and likewise for $F_{t,2:K+1}$, $F_{t,3:K+2}$, and so forth. Define $b = (b_1, ..., b_K)'$ to be the coefficient in a projection of $f_{t,K+1}$ onto $F_{t,1:K}$. In this model, the projection is exact so there is no residual,

$$f_{t,K+1} = b' F_{t,1:K}. \tag{7}$$

This equation simply states that in a linear model with K factors, the $(K+1)$-period forward can be expressed as an exact linear combination of maturities 1 to K. Because the vector $F_{t,1:K}$ plays a special role the rest of the paper, we refer to it simply as the "short end" of the term structure; i.e., the set of short-term claims that exactly span the full term structure.

This equation only links maturities 1 through $K+1$. Next, we derive a recursive relation that links the entire price curve to the short end in a convenient way. In particular, any two blocks of K consecutive forward prices with maturity shifted by one period (for example, $F_{t,1:K}$ and $F_{t,2:K+1}$) are linked by the equation:

$$F_{t,j+1:K+j} = BF_{t,j:K+j-1}, \quad B = \begin{bmatrix} b_K & b_{K-1} & ... & b_2 & b_1 \\ 1 & 0 & ... & 0 & 0 \\ 0 & 1 & ... & 0 & 0 \\ \vdots & & ... & & \\ 0 & 0 & ... & 1 & 0 \end{bmatrix}. \tag{8}$$

By the definition of b in (7), the relationship in (8) holds for $j = 1$. It follows from the law of iterated expectations that (8) holds for $j = 2$ because

$$E_t^{\mathbb{Q}}[F_{t+1,2:K+1}] = BE_t^{\mathbb{Q}}[F_{t+1,1:K}] \Leftrightarrow F_{t,3:K+2} = BF_{t,2:K+1}.$$

A recursive argument therefore establishes (8). It pins down the price of any forward on the term structure with the prices at the K immediate neighboring maturities via the matrix B.

Iteratively substituting (8) into itself implies

$$F_{t,j+1:K+j} = BF_{t,j:K+j-1} = B^2 F_{t,j-1:K+j-2} = ... = B^j F_{t,1:K}. \tag{9}$$

The geometric recursion in (9) further shows that prices at any maturity are pinned down by *any* K prices, even those at distant maturities. In particular, the equation links any price to the "short-end" vector $F_{t,1:K}$, where the coefficients are entirely determined by the powers of B.

Equation (9) is merely a restatement of the cross-equation restrictions summarized by Equation (6). However, the restrictions in (6) face the practical difficulty that they relate the restrictions to unobserved factors. What makes (9) powerful is that the restrictions are stated only in terms of observable prices. Specifically, the affine model structure requires not only that prices are perfectly correlated with the rest of the maturity curve, but limits the admissible shapes of the curve to those with *geometrically decaying loadings* (B^j) in regressions involving prices at different maturities.

While forwards are more convenient in mathematical derivations, it is more convenient to work with cumulative prices in empirical analyses (we discuss this further below). The representation of restriction (9) in terms of prices of cumulative claims is

$$P_{t,j+1:K+j} = (I + B + ... + B^j)R^{-1}P_{t,1:K}. \tag{10}$$

where $P_{t,j:m} = (p_{t,m}, p_{t,m-1}, ..., p_{t,j})'$. The B matrix is from Equation (8) and R is the $K \times K$ upper-triangular matrix of ones, which facilitates the algebraic adjustment from forwards to sums of forwards.[16] Representations (9) and (10) are exactly equivalent and we work interchangeably with the two.

2.5 Testing for Excess Comovement

We present now our general test for overreaction. For our analysis, we use prices for the first K maturities on the short end of the term structure to represent the K latent factors. In the preceding discussion we considered population projections, but to formalize the test we work with sample regressions. First, for each maturity $j = K + 1, ..., N$, we regress $f_{t,j}$ on

[16] Because $p_{t,n} = p_{t,n-1} + f_{t,n}$ for all n, we can write $P_{t,2:K+1} = P_{t,1:K} + F_{t,2:K+1}$. $P_{t,1:K} = RF_{t,1:K}$ because cumulative prices are sums of forwards, and $F_{t,2:K+1} = BF_{t,1:K}$ by iterated expectations. Substituting, we reach $P_{t,2:K+1} = RF_{t,1:K} + BF_{t,1:K} = (R + B)F_{t,1:K}$, and one more substitution arrives at $P_{t,2:K+1} = (R + B)R^{-1}P_{t,1:K} = (I + B)R^{-1}P_{t,1:K}$.

to $F_{t,1:K}$,

$$f_{t,j} = \hat{a}_j + \hat{b}_j' F_{t,1:K} + u_{t,j}. \tag{11}$$

We allow for a small measurement error term $u_{t,j}$ to avoid stochastic singularity following the term structure literature.

We construct a test of overreaction in the form of a variance ratio statistic at each maturity j. The coefficient $\hat{b}_j$ in Equation (11) is the *unrestricted* OLS regression estimate. The numerator of the variance ratio statistic for any maturity $j > K+1$ is the explained variance in the unrestricted regression and equals $\hat{b}_j' \hat{\Sigma}_{1:K} \hat{b}_j$, where $\hat{\Sigma}_{1:K}$ is the sample covariance matrix of $F_{t,1:K}$.

The denominator of the variance ratio is the explained variance in the *restricted* version of (11), where the constraint is the cross-equation restriction in Equation (9). The estimate $\hat{B}$ for the recursions in Equation (8) and (9) is obtained by using the estimated $\hat{b}_{K+1}$ as the first row of B and leaving all other rows unchanged. The *constrained* loading of $f_{t,j}$ on $F_{t,1:K}$, denoted $\tilde{b}_j$, is the first row of the matrix $\hat{B}$ raised to the power $j-K$:

$$\tilde{b}_j' = e_1(\hat{B}^{j-K}), \quad e_1 = (1, 0, ..., 0). \tag{12}$$

The explained variance in the constrained regression is therefore $\tilde{b}_j' \hat{\Sigma}_{1:K} \tilde{b}_j$, and the test statistic is

$$VR_j = \frac{\hat{b}_j' \hat{\Sigma}_{1:K} \hat{b}_j}{\tilde{b}_j' \hat{\Sigma}_{1:K} \tilde{b}_j}. \tag{13}$$

As we consider the time series variation of some long maturity price $f_{t,j}$, we wonder the extent to which this variation is consistent with variation at other maturities, from the point of view of an affine K-factor model. The VR_j statistic calculates the unconditional covariation of the long and short end prices and reports the fraction of this variation consistent with the model's cross-equation restrictions. Under the null of an affine K-factor 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 model predictions.

The variance ratio in (13) is based on forward prices, but the test is equivalently formulated from prices of cumulative claims. The test structure is identical, only the unrestricted and restricted regression coefficients ($\hat{b}$ and $\tilde{b}$) need modification. In analogy with (11), let $\hat{d}_j$ be the OLS slope estimate from an unconstrained regression of $p_{t,j}$ on $P_{t,1:K}$. The constrained

regression coefficient, denoted $\tilde{d}_j$, comes from the cross-equation restriction in Equation (10):

$$\tilde{d}_j' = e_1(I + \hat{B} + ... + \hat{B}^{j-K})R^{-1} \tag{14}$$

and the variance ratio test statistic is

$$VR_j = \frac{\hat{d}_j'\hat{\Sigma}_{1:K}\hat{d}_j}{\tilde{d}_j'\hat{\Sigma}_{1:K}\tilde{d}_j}. \tag{15}$$

Our empirical work uses the cumulative form in (15) for the following reason. If the eigenvalues of the risk-neutral cash flow persistence matrix ρ are below one in absolute value, then the system is stationary. In this case, forward prices at long maturities converge to a constant (and the denominator of VR_j converges to 0) because cash flows mean revert under the pricing measure. So, in the stationary case, infinite maturity assets have undefined variance ratios under the null. While this is not a pressing practical concern (most claims have maturities up to a few years), it is avoided by testing with cumulative claims data rather than forwards.

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 (1988*a*), Campbell (1991), Cochrane (1992), and many others.

Under the null of an affine no-arbitrage model, the restricted and unrestricted loading vectors $\hat{b}_j$ should equal $\tilde{b}_j$ element-by-element. When there is more than one factor in the model, it raises the question of how to best evaluate the 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

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

Rather than comparing loadings element-wise, the variance ratio sums loadings into a scalar in order 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 (co)vary are also the most informative about the dynamics of the model, and their factor

loadings receive the largest weights in our test.

In Appendix B.1 we describe a bootstrap procedure for conducting inference in small samples, which we use to construct confidence intervals in our main analysis. These calculations answer the question, "How likely are we to observe a given variance ratio given the sampling error of model parameter estimates?" In Appendix B.2, we report simulations demonstrating the finite sample performance of our estimating and testing approach. In particular, we show that model parameters and their standard errors are accurately estimated from short maturity prices alone, even when one of the factors has very low variance and is extremely persistent. We then show that our approach of estimating on the short end and testing on the long end produces accurate inference about the validity of the affine model.

2.5.1 Exponential-affine Models

The linear claim structure of Equation 3 is well suited for modeling variance claims,[17] which comprise several of the term structures we study. Claims in other asset classes, such as interest rates or credit default swaps (CDS), are more naturally modeled as exponential-affine claims. In that case, it is the log of x_t that is linear in factors H_t.

The model restrictions and testing procedures we derived above also apply in the exponential-affine setting under two additional assumptions regarding the distribution of factor innovations, $\Gamma\epsilon_t$, in Equation (4). First, ϵ_t follows a Gaussian distribution under $\mathbb{Q}$. Second, $\Gamma\epsilon_t$ is homoskedastic or, alternatively, it is heteroskedastic but its conditional volatility is uncorrelated with the factors (as in unspanned volatility models).

In exponential-affine models, the price of a cumulative claim is:

$$p_{t,n} = E_t^{\mathbb{Q}} \left[\exp \left(x_{t+1} + ... + x_{t+n} \right) \right].$$

Interest rate claims are the leading example in this class, where r_t is the instantaneous interest rate and $x_t = -r_t$. Prices are then related to factors according to[18]

$$\log p_{t,n} = \mathbf{1}' \left[\rho + \rho^2 + ... + \rho^n \right] H_t + \text{constant}. \tag{16}$$

For some claims it is preferable to model individual forwards with an affine-exponential form:

$$\log f_{t,n} = \log E_t^{\mathbb{Q}} \left[\exp \left(x_{t+n} \right) \right] = \mathbf{1}' \rho^n H_t + \text{constant}. \tag{17}$$

[17] See Egloff, Leippold and Wu (2010), Ait-Sahalia, Karaman and Mancini (2014), and Dew-Becker et al. (2015).

[18] A minor adaptation for the case of bonds is that powers of ρ range from 0 to $n-1$ rather than from 1 to n, though this is inconsequential for our variance ratio test.

The pricing formulas of Equations (16) and (17) differ from the simple affine form in (6) only by a constant due to assumptions on the distribution of factor innovations. Thus, (16) and (17) recover all the necessary structure to perform estimation and testing as described above, subject to the modification that we analyze log prices rather than price levels.

In the remainder of the paper, we focus on the homoskedastic case for three reasons. First, many of the asset classes we analyze (such as variance and inflation swaps) are typically modeled as claims to the level of x_t, in which case heteroskedasticity does not affect pricing. Conditional variance enters only in exponential models through the Jensen inequality term.

Second, conditional heteroskedasticity affects the loadings on the factors in exponential-affine models only to the extent that the factors themselves span the volatility of the errors. The term structure literature finds evidence of a large unspanned volatility component in interest rates (see, for example, Collin-Dufresne and Goldstein, 2002). So-called unspanned volatility models fix the loadings of bond prices on volatility factors to be zero. In this case, the factor loadings for log prices follow the same recursion as in standard homoskedastic models (Creal and Wu, 2015).[19]

In the remaining case where factor shock volatility is in fact spanned by prices, the magnitude of the effect on factor loadings is shown in the bond market to be small relative to the part of the loading coming from the claim's direct exposure to factors. Nonetheless, spanned volatility models can potentially affect our variance ratio test and Appendix D.3 performs robustness tests that directly account for heteroskedasticity. Our main conclusion from this check is that heteroskedasticity of factor innovations is not a central driver of our results.

3 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.

3.1 Implementation

For each asset class, we map the term structure to either the linear or the exponentially-affine specification (discussed below, case by case). To minimize the potential confounding

[19]For further discussion of the unspanned volatility case, see Collin-Dufresne and Goldstein (2002), Dai and Singleton (2003), Joslin (2006), Bikbov and Chernov (2009), and Creal and Wu (2015).

effects of illiquidity in these term structures, we focus on the most liquid contracts available for each asset class.

A key input to our tests is an estimate for the number of factors, K. For each term structure, we set K to the number of principal components necessary to explain at least 99% of the variation in the panel of prices at all available maturities.[20] We make an exception for Treasury yields by directly assuming three factors based on standard practice in the interest rate literature, though two principal components explain 99.9% of the variation in the panel. In later sections, we establish the robustness of our results to different choices of K.

From here, we use the first K short maturity prices to represent the factor space. We regress the price of the $K+1$ maturity claim on the first K maturities to estimate the baseline loadings (vector b in Equation (7)). These serve as the basis for restricted regression coefficients for maturities $K+2$ through N.[21]

3.2 Term Structure Tests by Asset Class

For each asset class, we describe the data and discuss any contract-specific or institutional features that need to be considered in the empirical analysis. Appendix E provides further in-depth descriptions of our data.

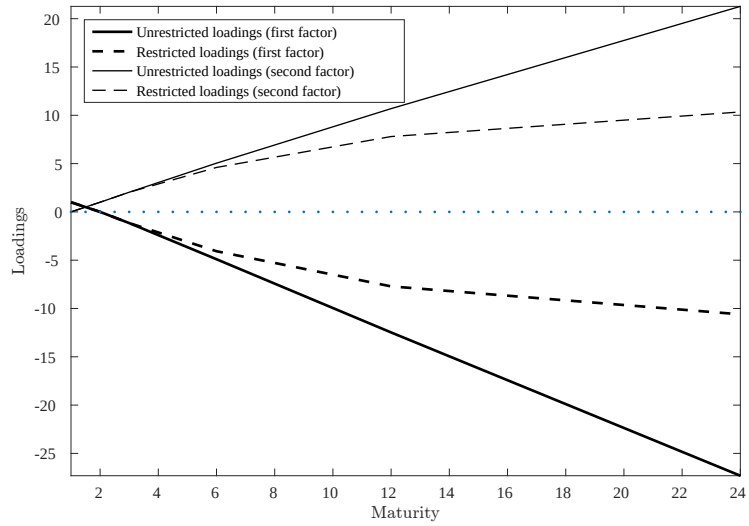
3.2.1 S&P 500 Variance Swaps

The first market we study is for variance swaps on the S&P 500 index. The variance swap market has the fascinating feature that it 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),

[20] We conduct this principal components step using variance-standardized prices, so that all points in the term structure are on equal footing in determining the number of factors.

[21] In Section 5, we discuss why treating the first K prices as an exact representation of the latent factor space is a powerful approach for detecting violations of the model's internal consistency conditions. It differs from a common practice in the term structure literature of estimating factors as principal components using price data at *all* available maturities, which is motivated by arguments of efficiency and overcoming potential measurement error in prices. Section 5.2 explains why model comparisons that rely on data from the full term structure can have limited power to detect the patterns of no-arbitrage violations we uncover in this paper. Appendix D.4 explains why our results are inconsistent with effects of measurement error.

Figure 2: Variance Swap Tests: Loadings



Note. The figure plots the loadings of prices of each maturity on the two factors (1-month and 2-month price). Solid lines indicate loadings in the unrestricted model, dashed lines indicate loadings in the restricted model.

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 2 with $x_t = RV_t$. We model RV_t as a linear function of latent factors, and explore robustness to using an exponential-affine specification in Appendix D.2.

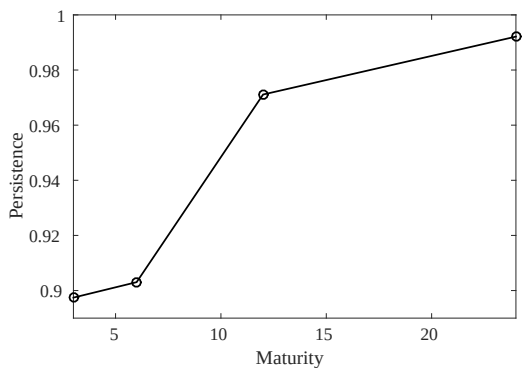
Variance swaps are traded in a liquid over-the-counter market with a total outstanding notional of around \$4 billion in "vega" at the end of 2013, meaning that a movement of one point in volatility would result in \$4 billion changing hands. Bid-ask spreads for maturities up to 24 months are relatively low, at around 1-2%. 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.[22]

We use daily price data for cumulative claims at all available maturities (1, 2, 3, 6, 12,

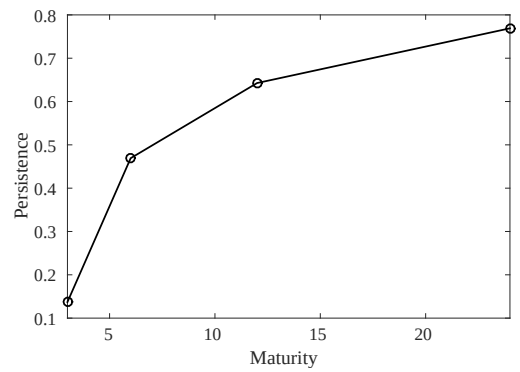
[22] Dew-Becker et al. (2015) show that the term structure of variance swap prices indeed closely matches the term structure of options-based synthetic swaps (more commonly known as the VIX).

Figure 3: $\mathbb{Q}$-persistence Estimated Along the Term Structure

(a) First factor



(b) Second factor



Note. Estimated persistence parameters in the two factor variance swap model. Persistences are re-estimated using data from various points in the term structure. Each value is plotted against the longest maturity claim used in its estimation.

and 24 months) during the period 1995 to 2013. Our baseline test uses $K = 2$, as two components explain 99.6% of the variance in the panel, consistent with existing literature.[23]

Our main findings for variance swaps are reported in Figure 1 in the introduction. The horizontal axis shows maturity of claims in months and the vertical axis shows the time series standard deviation of daily swap prices. The solid black line plots the explained swap price volatility from an unrestricted regression of each long maturity claim on the first K short maturity claims—this is the square root of the numerator in the variance ratio test. Points corresponding to observed maturities are marked with a circle. The dashed line plots the explained variation from the restricted regression that imposes the affine model's consistency conditions based on coefficient estimates in a regression of price $K+1$ on prices for the first K maturities—this the square root of the test's denominator. The variance ratio statistic for each maturity is printed above the unrestricted volatility estimates. The test statistic is only available for maturity $K+2$ and higher because the first $K+1$ maturities are used to estimate model parameters. Finally, the blue shaded region represents the point-wise 95% bootstrap confidence interval of predicted price volatility.

Plotting price variability in terms of standard deviation is convenient for visualizing the degree of cash flow persistence under the pricing measure. For a cumulative claim, the coefficient in a regression of long prices onto short prices is a geometric series in the persistence parameter, ρ. For example, in a one-factor model, the model-based standard

[23] E.g., Egloff, Leippold and Wu (2010),Ait-Sahalia, Karaman and Mancini (2014), and Dew-Becker et al. (2015).

deviation of an n-maturity claim is $\left(\sum_{j=1}^{n} \rho^j\right) \sqrt{Var(p_{t,1})}$. If cash flows are integrated under the pricing measure (ρ=1), then the standard deviation is a linear function of maturity. On the other hand, if the $\mathbb{Q}$-persistence of cash flows is in (0,1), 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, indicating stationarity of cash flows under the pricing measure. This is a first suggestion that variability on the long end is inconsistent with integrated or explosive model dynamics under $\mathbb{Q}$.

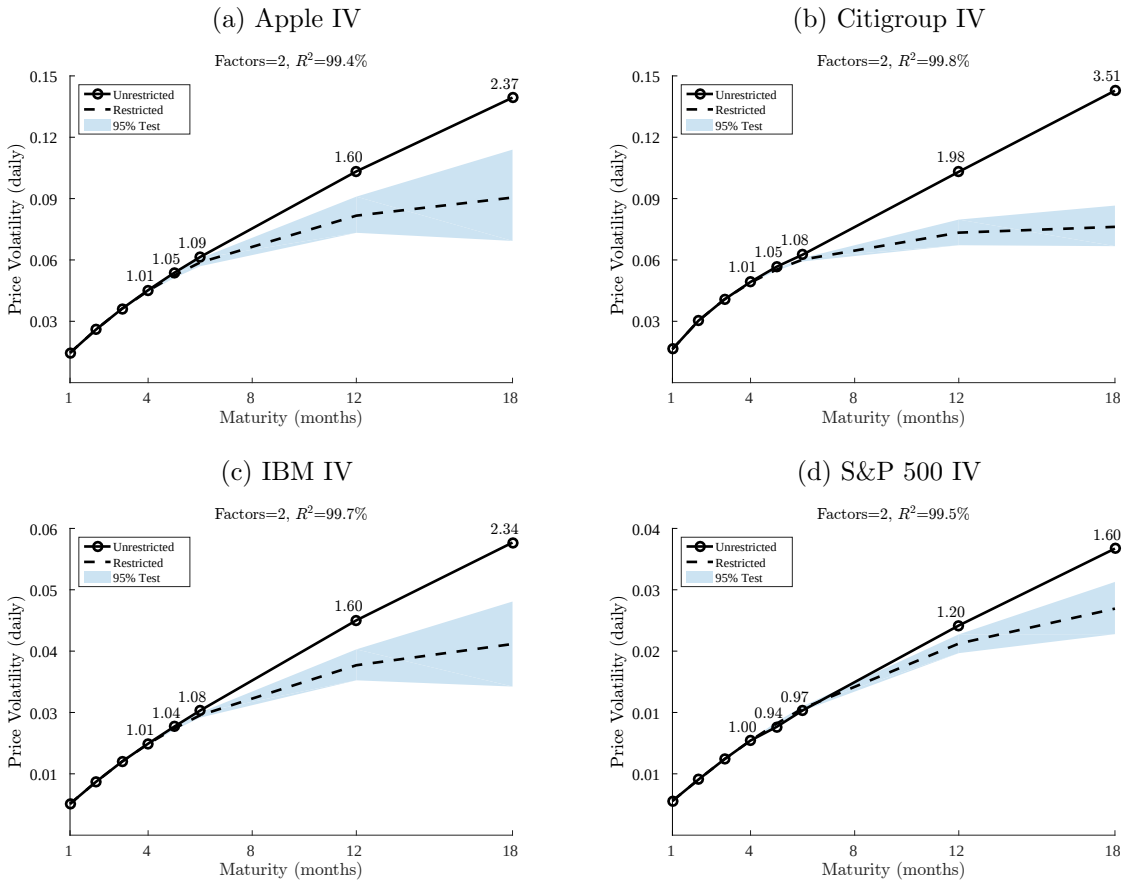
As described in the introduction, the unrestricted price variance at 24 months more than doubles the variance allowed under the affine pricing model's restriction. Comovement among prices at the short end of the curve suggests that cash flows mean revert relatively quickly under $\mathbb{Q}$. But this is not borne out on the long end—model-restricted volatilities increase with maturity at a much slower rate than the unrestricted volatility. Recall that these are "explained" price volatilities from regressing onto short-end prices. The high variance ratio therefore indicates that the prices at the long end of the curve react to the short end much more strongly than the affine model dynamics allow. There is *overreaction* at the long end relative to the short end, and relative to the estimated affine model.

Figure 2 plots estimated loadings of prices at each maturity on the model's two factors for both the restricted and the unrestricted model. The figure shows that long maturity prices overreact because they load too heavily on *each* factor, relative to the loadings predicted by the null model.

Two points warrant special emphasis regarding these results. First, the excess volatility of long maturity claims cannot be explained by movements in discount rates, as any discount rate variation that is describable within the affine class is subsumed by the $\mathbb{Q}$ model. Second, the data are exceedingly well described by a linear factor model (evident from an unrestricted R^2 near 100%), but with factor loadings that sharply differ from those implied by model restrictions.

In our main test, we estimate the two factor persistences from the short end of the curve, regressing the third shortest maturity claim on the two shortest maturities. Figure 3 provides another visualization of how the data deviate from the affine model. We estimate the two factor persistences from each point along the maturity curve. First, we estimate them from a regression of maturity 3 prices on prices for 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. The figure shows the persistence parameter estimates at different points on the curve. Under the null of the affine model both lines should be flat, as the implied factor persistence should be internally consistent along

Figure 4: Implied Volatilities of Equities (I)



Note. See Figure 1.

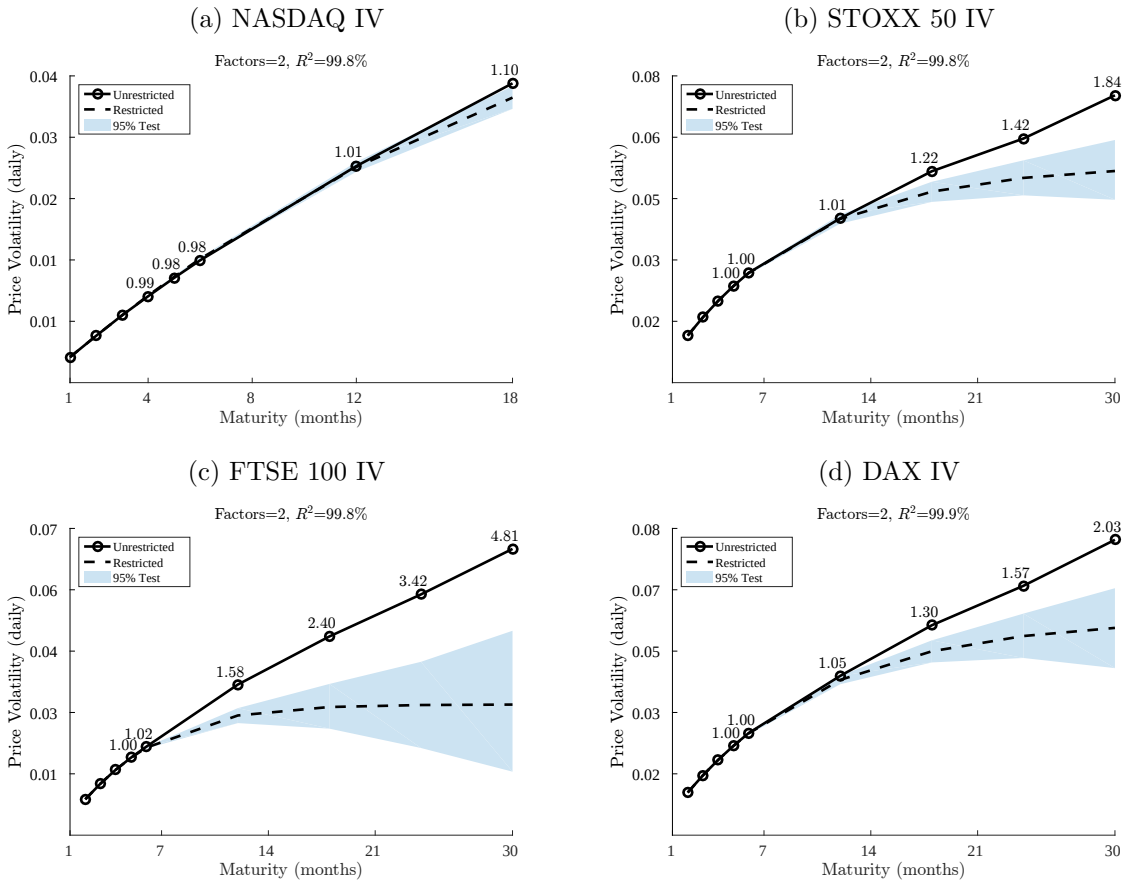
the curve. Instead, the figure shows that estimated persistence increases with maturity (for both factors). In other words, the data behave as though they are generated by a linear two factor model, but that investors implicitly think of factors as more persistent when they value longer maturity claims.

3.2.2 Equity Implied Variance

A well known result in option pricing establishes that variance swaps can be synthesized from a portfolio of put and call options with different strike prices.[24] Synthetic variances swaps are frequently encountered in practice. A 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.

[24]See Britten-Jones and Neuberger (2000) and Jiang and Tian (2005).

Figure 5: Implied Volatilities of Equities (II)



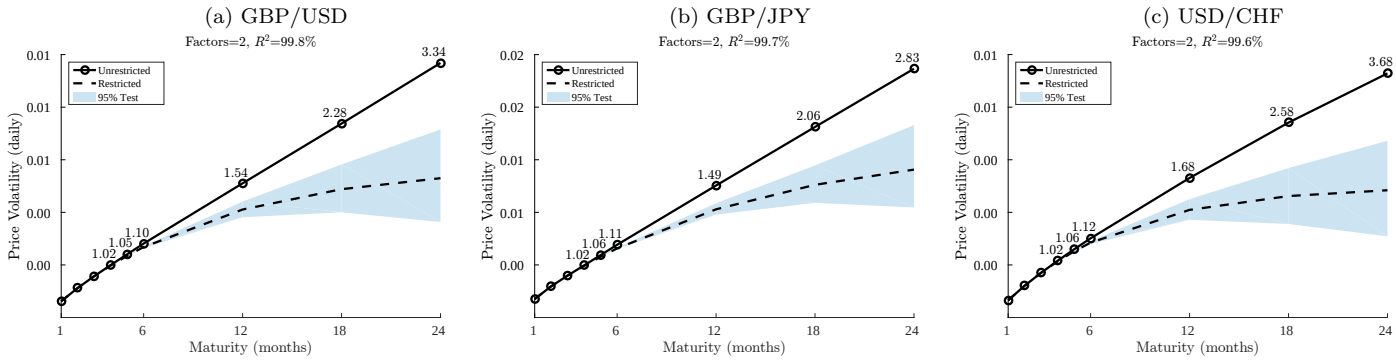
Note. See Figure 1.

For many option underlyings, however, a reliable VIX construction is unavailable due to the lack of deep out-of-the-money options. As an alternative, we study term structures of at-the-money (ATM) option implied volatilities.[25] Motivated by Carr and Lee (2009), who show that ATM implied volatilities approximate prices of claims to realized volatility ($\sqrt{RV}$), we treat implied variances as proxies for the price of a claim to realized variance. This is the same approach taken in Stein (1989)'s seminal work on excess volatility in the options market.

Figures 4 and 5 show variance ratio tests for term structures of equity options. We report results for three individual stocks (Apple, Citigroup, and IBM), two domestic stock indices (S&P 500 and NASDAQ), and three international stock indices (STOXX 50, FTSE 100, and

[25] Appendix E.1 discusses synthetic variance swap (VIX) term structures in more detail. We also conduct a robustness analysis comparing VIX and ATM implied volatility. In data sets with a sufficient number of long dated OTM options to construct the VIX term structure, we show that VIX-based variance ratio tests behave the same as our main results that use ATM implied volatility.

Figure 6: Currency Implied Variance



Note. See Figure 1.

DAX).

The results corroborate those observed for variance swaps. Variance ratios at the longest maturities (from 18 to 30 months) range between 1.6 and 4.8 and are significantly different from one at the 95% level. The only exception is NASDAQ, for which the variance ratio is 1.1 at the long end and is insignificant.

3.2.3 Currency Implied Variances

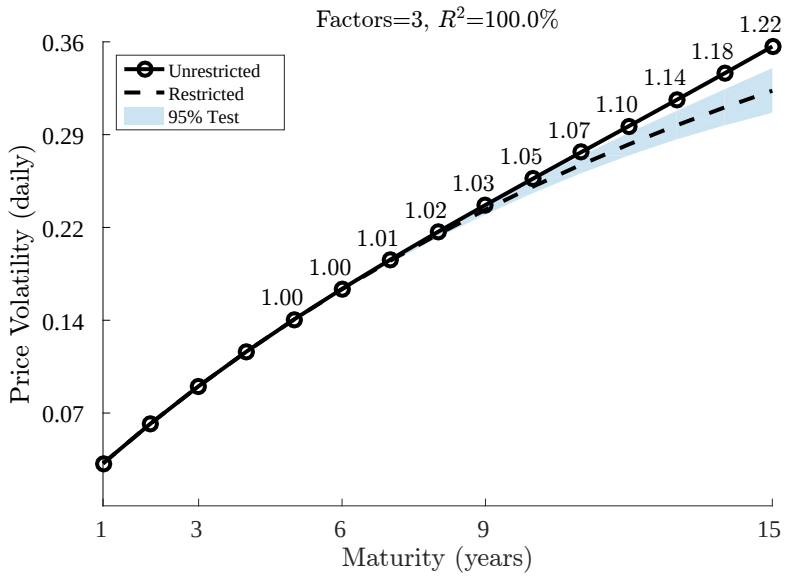
We next study the term structure of currency options. As in the case of equity options, we treat implied variances at different maturities as proxies for variance swaps, and apply the same variance ratio test for linear claims that we used to study variance swaps. Our data are for three of the highest volume currency pairs (GBP-USD, GBP-JPY, and USD-CHF) from JP Morgan, covering the period 1998-2014, with maturities up to 24 months. Variance ratio tests based on currency options are plotted in Figure 6, and share the same patterns found in term structures of other volatility claims.

3.2.4 Interest Rates

US government bond prices are among the most well studied data in all of economics. Our US bond data comes from Gurkaynak, Sack and Wright (2006). The data consist of zero-coupon nominal bonds with maturities of 1 to 15 years for the period 1971 to 2014, and is available at the daily frequency (we do not use higher maturities because their sample starts later). The term structure is bootstrapped from coupon bonds and uses strict interpolation (no extrapolation), so that a maturity is present only if enough coupon bonds are available for interpolation at that maturity.

The pricing model we use for interest rates is the standard homoskedastic exponential-

Figure 7: Treasuries



Note. See Figure 1.

affine model. We discuss this specification in detail in Appendix D and show empirically that heteroskedasticity plays a minor role in the variance ratios we estimate. It is well known from the interest rate literature that three factors provide a good fit of the model, and our estimates confirm that three factors explain more than 99.9% of the common variation among log yields.

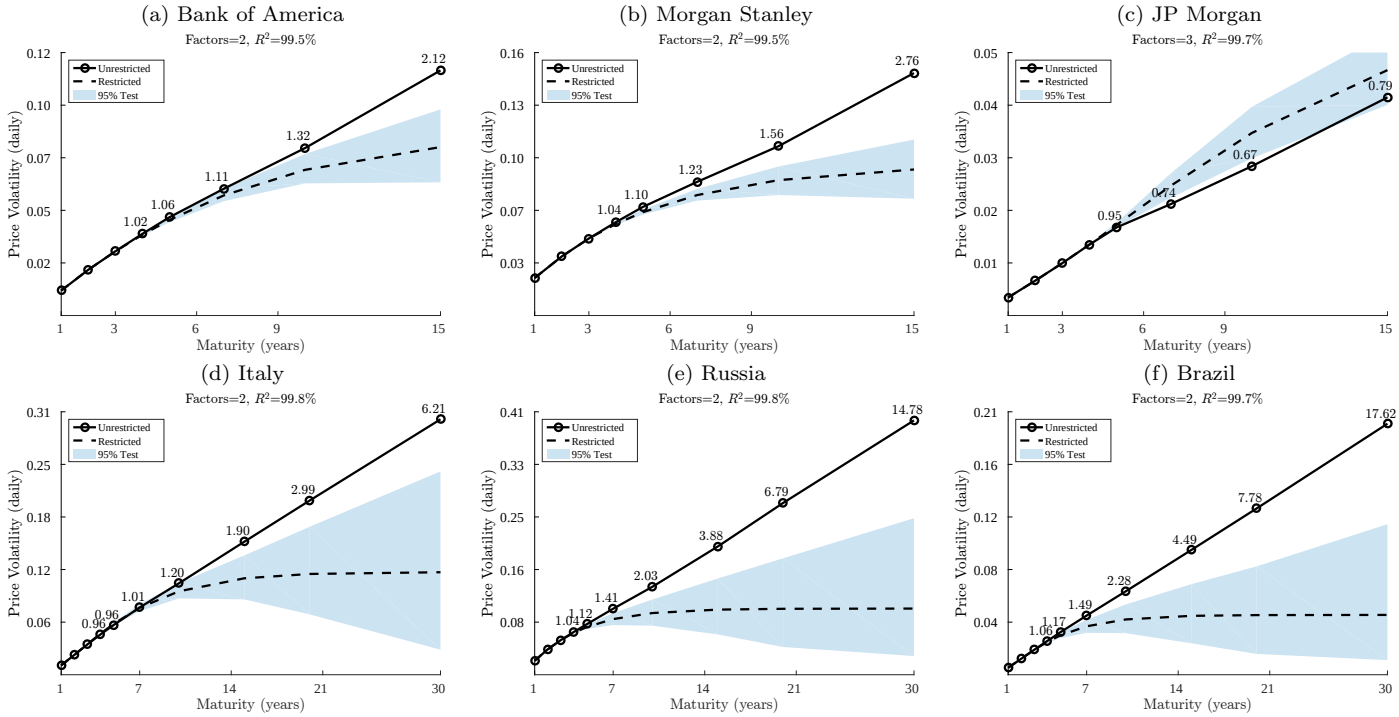
The variance ratios tests in Figure 7 also show that yields deviate only slightly from the affine model restrictions. The maximum variance ratio is 1.2 at 15 years. While this is statistically greater than one and corroborates the excess volatility results of Gurkaynak, Sack and Swanson (2005) and others, it is economically the smallest excess volatility effect that we find among all of the term structures that we study. It is interesting that early affine term structure models were formulated to describe the Treasury market, and this is the market in which affine models indeed appear to perform best.[26]

3.2.5 Credit Default Swaps

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

[26]In Appendix E.2, we estimate ρ at different points of the Treasury curve like we did for variance swaps in Figure 3. The resulting ρ values are nearly flat across maturities for the two largest factors, showing that deviations from the affine model in this market are small and again confirming earlier results from the interest rate literature.

Figure 8: Credit Default Swaps



Note. See Figure 1.

of 1, 3, 5, 7, 10, 15, 20, and 30 years.

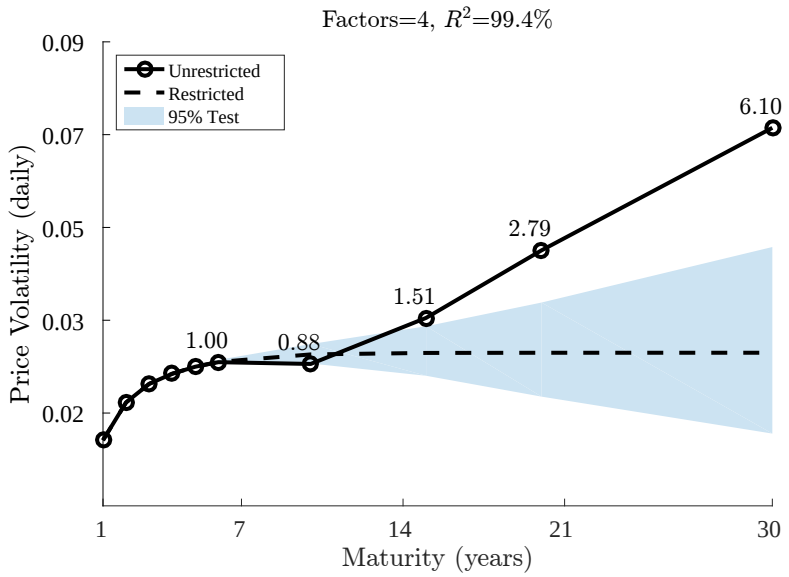
Among the different CDS contracts written on the same reference entity, we choose those with highest liquidity. These are CDS written on senior bonds, with modified-restructuring (MR) clause, and denominated in US dollars.[27] Our CDS analysis focuses on the period from January 2007 onwards. We choose the three most traded sovereigns (Italy, Brazil, Russia) and the three most traded corporates (JP Morgan, Morgan Stanley, Bank of America) according to 2008 volume. In the plots below, we focus on maturities up to 15 years for individual names, and up to 30 years for sovereign CDS. Confidential DTCC data indicate that following 2008 there is positive volume at 15 years for our corporate names and 30 years for our sovereign names.

In Appendix E we describe how to map CDS prices into the framework of Section 2. 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.

Figure 8 reports variance ratio tests for CDS markets. For individual names, variance ratios are as high as 2.8 at 15 years (with the exception of JP Morgan, whose long maturity

[27] For sovereigns, we use contracts with the CR clause, as more data is available than for the MR contracts.

Figure 9: Inflation Swaps



Note. See Figure 1.

variance ratio is insignificantly different from one). For sovereign CDS, data show variance ratios in excess of six at long maturities. Overall, CDS results indicate a qualitatively and quantitatively significant overreaction similar to other asset classes considered in this paper.

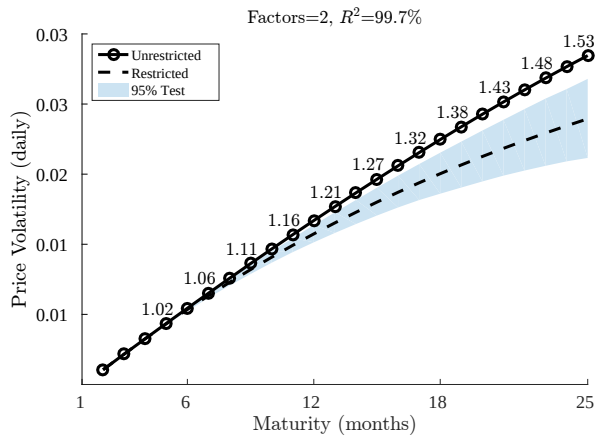
3.2.6 Inflation Swaps

We obtain inflation swaps data from Bloomberg. We observe the full term structure between one and 30 years at the daily frequency over the period 2004 to 2014. As reported in Fleming and Sporn (2013), "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 same data we use is also studied by Fleckenstein, Longstaff and Lustig (2013).

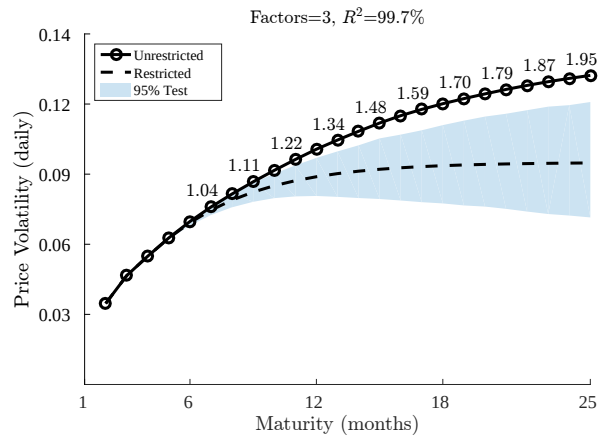
The term structure model for inflation swaps falls neatly within the exponential-affine specification of Section 2 (with additional model details in Appendix E). Figure 9 shows that the variance ratio pattern for inflation swaps is a more extreme version of the pattern in other asset classes, in that price volatility is at first strongly concave in maturity but then rises rapidly between 15 and 30 years to maturity. Variance ratios at the long end reach beyond 6.0, and are thus inconsistent with affine model restrictions.

Figure 10: Commodity Futures

(a) Gold Futures (Comex)



(b) Crude Oil Futures (NYMEX)



Note. See Figure 1.

3.2.7 Commodity Futures

Commodity futures data are from CME Group. We select the two most traded contracts for which we observe at least 1,000 daily observations at both the short end of the term structure (1 month) and at the long end (24 months). These are gold futures and crude oil futures.

Appendix E describes how we map futures prices into the (exponential) affine setup. Note that these contracts 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. Figure 10 shows variance ratios that are significantly higher than one in both commodity futures markets at the 24 month maturity.

4 Potential Sources of Violation

In this section we explore potential explanations behind the pervasive evidence of excess price volatility relative to the affine-$\mathbb{Q}$ model. We classify possibilities into four forms of model misspecification, i) omitted factors, ii) long memory $\mathbb{Q}$-dynamics of cash flows, iii) non-linear $\mathbb{Q}$-dynamics, and iv) temporary mispricings along the term structure. 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 in this section is to provide the reader

with intuition for how certain affine model violations can impact the behavior of the variance ratio test.

We argue that missing factors are unlikely to generate patterns that we see in the data. We also show in simulations that non-linear dynamics and long-range dependence can potentially generate high variance ratios, but that incorporating just one or two additional factors in the test drops the variance ratios back to nearly one, and this is not the case in the data when we add an extra factor. Mispricing of long maturity claims due to overreaction emerges as a potential driver of high variance ratios, as we document a profitable trading strategy that exploits excessive fluctuations of long maturity prices.

4.1 When the Affine Model is Misspecified

We start with a general characterization of our tests under model misspecification. Our estimator assumes a K-factor affine-$\mathbb{Q}$ model of prices along the term structure. If this is not the true data generating process, then the population projection in Equation (7) becomes

$$f_{t,K+1} = b' F_{t,1:K} + u_t \tag{18}$$

or, in analogy to the matrix recursion in (8),

$$F_{t,2:K+1} = B F_{t,1:K} + U_t, \tag{19}$$

with B taking the same structure as earlier and $U_t = (u_t, 0, ..., 0)'$. Equation (18) now contains a residual that is solely due to specification error.

Under misspecification, the coefficient B in (19) is no longer fixed and instead becomes specific to the maturities used in the projection. For other maturities, the projection coefficient generally takes a different value. This reflects the fact that cross-equation restrictions of the affine model in (9) are only satisfied when the model is correctly specified.

A key question is whether the violations of the cross-equation restrictions observed in the data can tell us anything about the nature of the model misspecification. We arrived at the no-arbitrage restrictions in (9) by iterating expectations in the price-on-price projection equation. Repeating this using the representation of Equation (18) and imposing the no-arbitrage condition that $E_t^{\mathbb{Q}}[f_{t+1,j}] = f_{t,j+1}$, we find for all $j > 1$ that

$$F_{t,j+1:K+j} = B^j F_{t,1:K} + \sum_{l=0}^{j} B^l E_t^{\mathbb{Q}}[U_{t+l}]. \tag{20}$$

Equation (20) is an exact representation of prices at all maturities regardless of misspecification (assuming there is no arbitrage). The first term on the right-hand side captures the variation in $F_{t,j+1:K+j}$ that is consistent with the affine model restrictions given projection (18). The second term captures the deviation from the model. We can decompose the behavior of this deviation by projecting it onto $F_{t,1:K}$. All elements of the vector U_{t+1} other than the first are zero, so we write this projection as

$$e_1 \sum_{l=0}^{j} B^l E_t^{\mathbb{Q}}[u_{t+l}] = \gamma_{K+j} F_{t,1:K} + \zeta_{t,K+j},$$

where γ_{K+j} is a K-vector and $\zeta_{t,K+j}$ is scalar. This decomposition allows us to write (20) as

$$F_{t,j+1:K+j} = (B^j + \gamma_{K+j}) F_{t,1:K} + \zeta_{t,K+j} \tag{21}$$

where the projection residual $\zeta_{t,K+j}$ is orthogonal to the first K prices, $F_{t,1:K}$. When testing model restrictions, we estimate the unrestricted linear projection of $F_{t,j+1:K+j}$ on to $F_{t,1:K}$ in (21) and compare the estimated projection coefficient, $(B^j + \gamma_{K+j})$, to the affine-model-restricted coefficient, B^j.

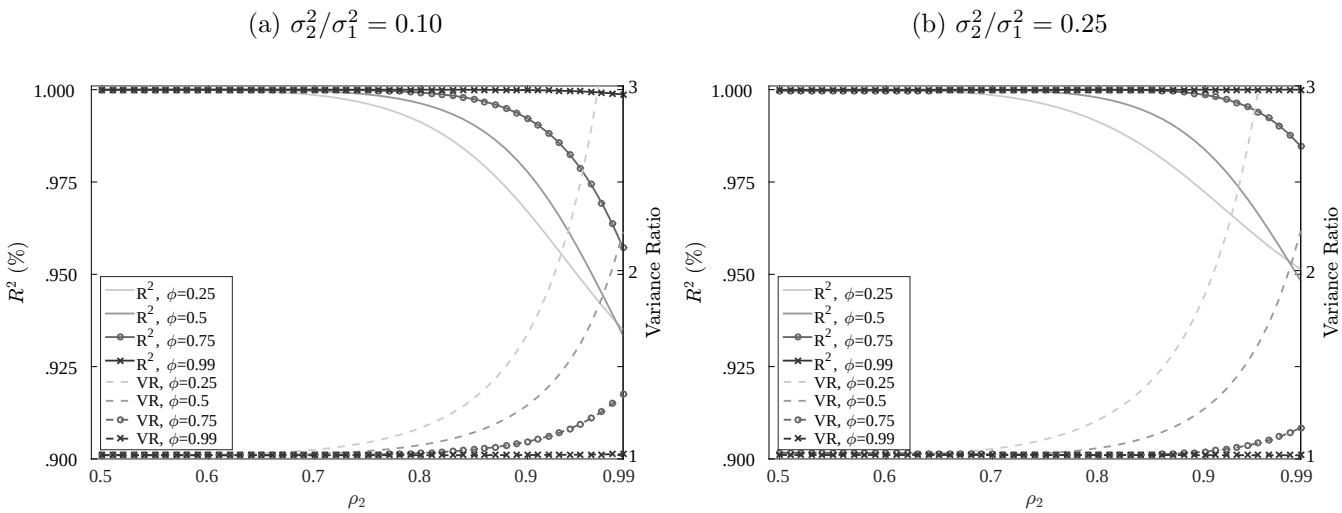
The behavior of the unrestricted projection is informative about the nature of the misspecification. Two stark empirical facts emerge uniformly from data in all asset classes. First, the unrestricted linear factor model (21) provides an excellent fit of the data, with R^2 approaching 100%. Second, variance ratios are significantly greater than one.

Together, these facts provide insight about behavior of the specification error term, $\sum_{l=0}^{j} B^l E_t^{\mathbb{Q}}[U_{t+l}]$. High variance ratios tell us that the total variation of the specification error, $Var(\sum_{l=0}^{j} B^l E_t^{\mathbb{Q}}[U_{t+l}])$, must be large. At the same time, an unrestricted R^2 approaching 100% means that the portion of the specification error that is uncorrelated with the short maturity prices, $Var(\zeta_{t,K+j})$, must be tiny. In other words, the specification error must be nearly perfectly correlated with the factors from the short end. This is evidently the case, as high variance ratios are equivalent to the unrestricted projection coefficients being significantly larger in magnitude than the model restriction allows—the γ_{K+j} coefficients are far from zero (as found in Figure 2). This is evidence of investor overreaction relative to the affine model.

4.2 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. Two pieces of evidence

Figure 11: Effects of Missing Factors



Note. The figure reports the variance ratios and R^2 obtained when the true model is a two-factor model but only one factor is used in estimating the $\mathbb{Q}$ dynamics. Results are for various combinations of the variance of the second factor relative to the first (σ_2^2/σ_1^2) and persistence of the second factor (ρ_2).

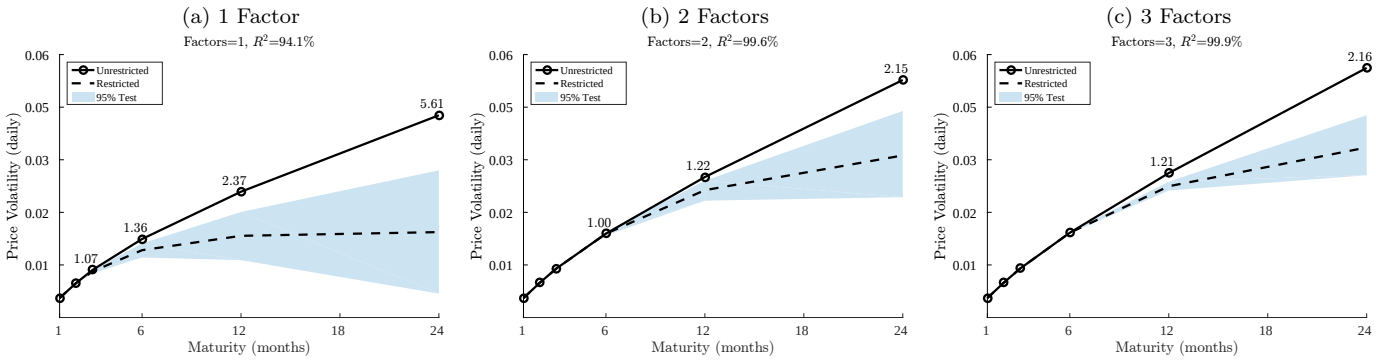
indicate that omitted factors are unlikely to explain our findings.

An omitted factor that is consistent with our estimation results must have a particular set of traits described in Section 4.1. It must be volatile and persistent enough to generate high variance ratio at long maturities. Yet it must also be highly correlated with the other factors, as too much unique variation in the factor will pull the R^2 below the values found in the data.

A calibration shows that such a factor is essentially infeasible from a quantitative standpoint. Consider a term structure whose data-generating process is a two-factor model of cash flows. Factor i has variance σ_i^2 and persistence of ρ_i, $i = 1, 2$, and the factors have a correlation of ϕ. What happens when we estimate an affine model with $K = 1$, thereby misspecifying the model to have too few factors? Figure 11 shows the possible scenarios for the (population) R^2 and variance ratio statistic at a maturity of 24 periods. The calculations are based on a range of values for the persistence of the second factor (ρ_2) and how correlated the factors are (ϕ). We fix the monthly persistence of the first factor to $\rho_1 = 0.5$, and fix the variability of the second factor relative to the first at either $\sigma_2^2/\sigma_1^2 = 0.10$ (left panel) or 0.25 (right panel).

We find no combination of parameters that can simultaneously generate an R^2 over 99% and a variance ratio that is meaningfully greater than one. The best chance comes when

Figure 12: Variance Swaps: Varying the Number of Factors



Note. See Figure 1.

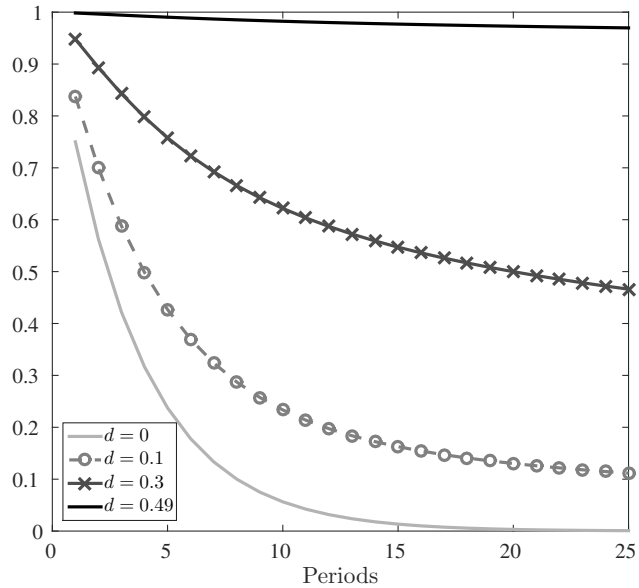
the second factor is extremely persistent ($\rho_2 \to 1$) and highly correlated with the first factor ($\phi \to 1$). This rather strange second factor influences long-end variances due to its strong serial correlation, yet it is masked by the first factor due to their high correlation, which allows the model to achieve a very high R^2 with a single factor. However, even in this "best" case, the variance ratios from the misspecified model rise only a few percentage points above one so long as the R^2 is near 99%.[28]

Second, if a missing factor were driving our results, we can account for it in our empirical analysis with a simple robustness check that allows for additional factors in the model. Figure 12 shows a sequence of variance swap test plots with the number of factors increasing from one to three. With one factor, the model R^2 is 94.1%, and the variance ratio at 24 months is 5.61. The two-factor case is the main result reported in Figure 1, which has an R^2 of 99.6% and a long-end variance ratio of 2.15. Finally, with three factors, the R^2 exceeds 99.9%, and continues to produce large economic and statistical rejections of the affine model ($VR_{24} = 2.16$). We see this type of behavior throughout the asset classes we study. Table A7 in Appendix F documents similarly high and significant variance ratios as we gradually expand the number of factors beyond that of our benchmark analysis in Section 3. Very broadly, our conclusions are unaffected by adding factors beyond those needed to explain at least 99% of the total term structure variation.[29]

[28] The quantitative results in this example are quite general; the figure shows a conservative set of parameter values. Considering more factors, allowing for a higher ratio of σ_2^2/σ_1^2, or allowing for greater persistence in the first factor typically make it even less likely that a missing factor can explain our findings.

[29] There is of course always a factor model that delivers variance ratios equal to one—it is a model with the number of factors equal to the number of maturities observed in the term structure. This extreme model is a reminder that the modeler's objective is to maximize the variety of phenomena explained by a model 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. Besides the evident inability of additional factors to reconcile the data with affine models, resorting to richer parameterizations when a great majority of data variation is already explained is scientifically unsatisfying. Duffee (2010) raises an

Figure 13: Long Memory Mean Reversion



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

4.3 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 evident from the concave shape of price volatility versus maturity. If the model is integrated, the transition matrix ρ in Equation (4) will have an eigenvalue of one (our methodology allows for non-stationary roots, see Appendix A). In all term structures, we estimate eigenvalues of ρ that are below one in absolute value.

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

additional concern about using too many factors. He shows that overfitting the interest rate term structure with more than three factors leads to implausibly high Sharpe ratios for some fixed-income portfolios.

Table 1: Effects of Long Memory

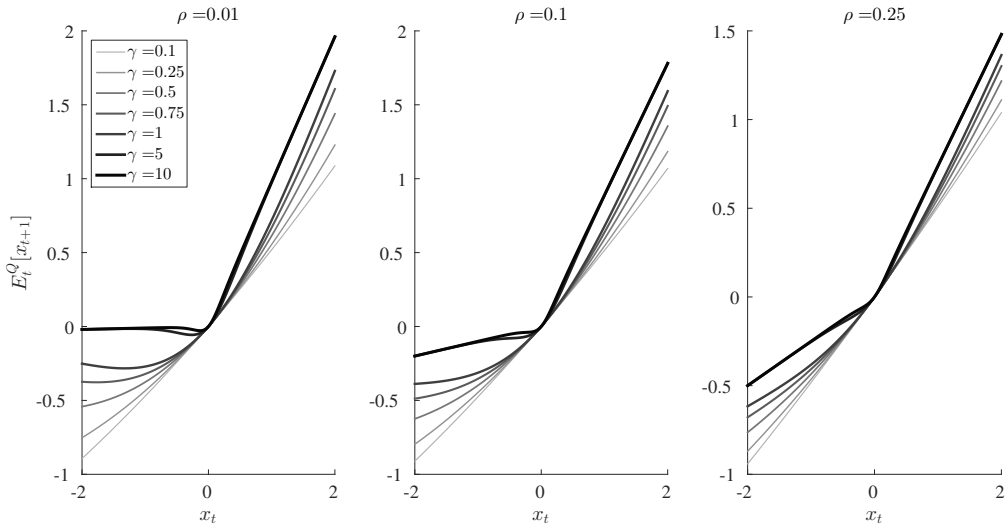
		AR(1)=0.25			AR(1)=0.50			AR(1)=0.75		
d	K	R^2	VR_{12}	VR_{24}	R^2	VR_{12}	VR_{24}	R^2	VR_{12}	VR_{24}
0.10	1	96.8	2.0	2.9	99.1	1.3	1.7	99.9	1.0	1.1
0.10	2	100.0	1.0	1.2	100.0	1.0	1.0	100.0	1.1	1.2
0.10	3	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0
0.20	1	97.1	2.4	4.1	98.9	1.5	2.2	99.9	0.9	0.9
0.20	2	100.0	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

Note. 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.

is 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. Figure 13 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 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 1 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 non-stationary. In these cases, inclusion of an "extra" factor (beyond the 99% R^2 requirement) brings variance ratios close to one. Evidently, despite its incorrect specification, the affine model with two or three factors is an accurate enough approximation

Figure 14: Non-linear Cash Flow Dynamics



Note. 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.

of the ARFIMA process that the misspecification can go undetected. The ability of one or two additional factors (beyond the 99% R^2 requirement) to drive variance ratios toward one in simulations is an important difference versus the data. In the missing factor robustness checks of Table A7, we find variance ratios in the data that remain well above one despite inclusion of an extraneous factor.

4.4 Non-linearities

A third potential explanation of our findings is that cash flows evolve non-linearly. In this section, we explore the effects of estimating and testing restrictions of a misspecified affine model when the true cash flow process has non-linear dynamics.

We study a class of processes known as smooth transition autoregressive (STAR) models.[30] As emphasized by Granger and Terasvirta (1993), STAR models encompass a broad variety of non-linear dynamics that have proven successful in modeling economic time series. While far from exhaustive, they allow us to gain some insight into the role that non-linearities play in our empirical results.

[30] Teräsvirta (1994) provides an excellent econometric treatment of STAR models.

Table 2: Effects of Non-linearity

		ρ=0.01			ρ=0.10			ρ=0.25		
γ	K	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

Note. 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.

We assume that cash flows evolve according the one-factor non-linear process

$$x_t = \rho x_{t-1}\left(1 - (1 + e^{-\gamma(x_{t-1}-c)})^{-1}\right) + (1-\rho)x_{t-1}(1 + e^{-\gamma(x_{t-1}-c)})^{-1} + \epsilon_t. \tag{22}$$

Equation (22) 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 for the process to transition between high and low serial correlation depending on the state of the process.[31] The degree of non-linearity is governed both by ρ and γ.

Figure 14 plots the model-implied relationship between x_t and $E_t^{\mathbb{Q}}[x_{t+1}]$, illustrating the extent of non-linearity 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 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.

Term structure prices are analytically intractable for STAR models, but are easy to calculate via simulation. We calculate 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.

[31] By incorporating time variation in autocorrelation, the STAR model's non-linearities accommodate parameter instability that may arise, for example, from investors learning about ρ.

The results are reported in Table 2. In this large family of non-linear models (including rather extreme non-linearities 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 non-linear $\mathbb{Q}$-dynamics and the variance ratio does not detect significant violations of cross-equation restrictions.

In Appendix G we explore more complex non-affine specifications, including heteroskedastic STAR models, mixture STAR/long memory models, and multifractal models. The behavior of variance ratio statistics in these simulated settings is similar to those in Tables 1 and 2.

4.5 Overreaction and Other Expectation Errors

A fourth possibility for explaining variance ratios greater than one is that the affine model is indeed an accurate description of the true value of claims, but that some of these claims are subject to temporary mispricing.

We can characterize what *any* mispricing must look like if prices along the term structure follow an (unrestricted) linear factor model. For illustration, suppose that a term structure of forward prices is exactly described by a one-factor model, so that (suppressing constants)

$$f_{t,1} = b_1 x_t, \quad f_{t,2} = b_2 x_t, \quad ..., \quad f_{t,n} = b_n x_t. \tag{23}$$

In order for this term structure to satisfy no-arbitrage, it *must* be the case that the loadings follow a geometric progression in some constant ρ:

$$b_1 = \rho, \quad b_2 = \rho^2, \quad ..., \quad b_n = \rho^n. \tag{24}$$

If the factor loadings behave in any other way, they violate the law of iterated expectations.

To prove this, first note that forward prices must be linear in the factor by assumption, meaning that $f_{t,j} = E_t^{\mathbb{Q}}[x_{t+j}] = b_j x_t$. Second, by the law of iterated expectations, any forward price today also represents an expectation of tomorrow's price of a forward with shorter maturity.

$$f_{t,j} = E_t^{\mathbb{Q}}[x_{t+j}] = E_t^{\mathbb{Q}}\left[E_{t+1}^{\mathbb{Q}}[x_{t+j}]\right] = E_t^{\mathbb{Q}}[f_{t+1,j-1}].$$

If we fix the initial coefficient to $b_1 = \rho$, these two properties together imply that $b_2 = b_1^2 = \rho^2$, which in turn implies $b_3 = \rho^3$, and so forth. This argument establishes the following proposition in the one-factor case, and easily generalizes to the case of multiple factors.

Proposition. *If a term structure of prices obeys an exact affine factor model, then mispric-*

ings exist along the term structure if and only if factor loadings have non-geometric decay.

This simple result is powerful for understanding the nature of affine model violations documented above. The unrestricted model in our variance ratio tests takes a linear factor form much like (23), and we find that this model provides an excellent description of the data with R^2 values near 100%. At the same time, the high variance ratios reveal that the b_j coefficients decay at a less than geometric rate. This violates the structure in (24), suggesting that the law of iterated expectations may be violated, which can in turn lead to mispricings. The empirical fact that the loadings decay more *slowly* than the affine model allows tells us that the nature of the model violation is one of overreaction at the long end of the term structure.

An important caveat is that the term structure R^2 for the unrestricted linear model is not *identically* 100%, which means that the conditions of the proposition are not exactly satisfied, and thus the slow decay in coefficients detected by high variance ratios is potentially due to the affine model being misspecified. 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? This issue makes it difficult to discern whether the affine model is violated due to misspecification, or due to mispricings arising from investor behaviors (such as a tendency to commit errors when iterating expectations).

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

4.5.1 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 hypothesis of a K-factor affine model, we can determine at any point in time whether a long maturity claim is overpriced or underpriced relative to the model by comparing traded versus fitted prices (where fits are an estimated function of the first K short maturity prices, as in Equation (10)). 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. This amounts to temporary over or undervaluation of long claims (relative to the model).

The logic of the strategy presented below begins with the presumption 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 claim with maturity $N + n > K$ and holding this position for n periods. 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

$$p_{t+n,N} = a_N + (b_N)' P_{t+n,1:K} \tag{25}$$

where a_N and b_N are model-implied coefficients as in Equation (10). And, over the n-period investment period, the claim has paid out cash flows of $x_{t+1}, ..., x_{t+n}$.

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 (25) 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 $x_{t+1}, ..., x_{t+n}$ and has ongoing value of $p_{t+n,N}$.

Portfolio $\mathcal{B}$ is designed to replicate the right-hand-side of Equation (25). First, it invests the present value of a_N in the n-maturity risk-free bond (for simplicity let us assume that the risk-free rate is zero). Next, it buys all claims with maturities of $n + 1, ..., n + K$, 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 b_N. Third, it buys $(1 - (b_N)'\mathbf{1})$ shares of an n maturity claim with price $p_{t,n}$.

After n periods, the risk-free bond has matured with a value of a_N and the position $(b_N)' P_{t,n+1:n+K}$ has ongoing value of $(b_N)' P_{t+n,1:K}$. 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 $x_{t+1}, ..., x_{t+n}$. 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 3.

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

Table 3: 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	$b_N' P_{t,1+n:K+n} + (1 - b_N'\mathbf{1})P_{t,n}$	0
$t+1$	$p_{t+1,N+n-1}$	x_{t+1}	$b_N' P_{t,1+n-1:K+n-1} + (1 - b_N'\mathbf{1})P_{t,n-1}$	x_{t+1}
$\vdots$				
$t+n$	$p_{t+n,N}$	x_{t+n}	$b_N' P_{t,1:K} + 0$	x_{t+n}

Note. 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 a_N in the n-maturity risk-free bond (we simplify with a risk-free rate of zero), buying all claims with maturities of $n+1, ..., n+K$ with the number of shares in each claim given by the vector b_N, and buying $(1 - (b_N)'\mathbf{1})$ shares of an n-maturity claim.

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.[32] 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 compute the return to this strategy taking into account realistic constraints on capital and margining of positions. In particular, we assume that each trade must be fully collateralized on both the long position and short position (an initial margin requirement of 100%). 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

$$C \geq qZ_L + qZ_S$$

or therefore

$$q \leq \frac{C}{2Z_S - \Pi}.$$

[32] In practice, the liquidation equation (25) does not hold exactly. To minimize the liquidation risk, a_N and b_N are based on unrestricted regressions of N-maturity prices on prices for maturities 1 through K. This minimizes the squared liquidation error.

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 implement the trading strategy in the variance swap market. We recreate a purely out-of-sample execution of the strategy. That is, when deciding on a trade at time t, estimated model parameters (particularly those of a_N and b_N) and position choices only use data that an investor would have access to in real time (the history of term structure prices through date t). We re-estimate 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 examine thresholds based on the historical distribution of Π during the rolling estimation window. Therefore, at each date t, the initial profit is being compared only with backward looking information and the trading choice genuinely preserves the out-of-sample character of the trade.

The "Variance Swaps" column in Table 4 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 ($N + n = 4$, 6, 12, 18, or 24 months), and with various thresholds for trade initiation (equal to the 50th, 75th, or 90th historical percentile for Π).[33]

We obtain consistently high Sharpe ratios in all cases, often above 1, 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 in Sections 4.3 and 4.4, 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 (e.g. one with non-linearities or long-range dependence), 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 models. We compare against three models in which long maturity variance ratios are greater than one because the estimated affine model is misspecified, but in which the simulated claims are always correctly priced. These include

1. the two factor affine model with $\rho_1 = 0.9$ and $\rho_2 = 0.5$, but estimated assuming a single factor structure

[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 one day in ten.

Table 4: Trading Strategy Sharpe Ratios

Mispricing Threshold	Longest Maturity Traded	Variance Swaps	Simulations			
			Missing Factor	Long Memory	Non-linear	Arbitrage
50	4	0.79	0.07	-0.34	-0.01	0.07
50	6	1.63	0.08	-0.24	0.06	-0.14
50	12	0.78	0.09	0.30	0.04	3.92
50	18	1.18	0.10	0.40	0.04	4.08
50	24	0.66	0.10	0.37	0.06	4.34
75	4	1.06	-0.13	-0.23	-0.16	0.21
75	6	1.99	-0.12	-0.40	-0.14	-0.10
75	12	0.81	-0.11	0.26	-0.06	3.89
75	18	1.27	-0.11	0.37	-0.01	4.07
75	24	0.83	-0.11	0.41	0.02	4.37
90	4	1.52	0.15	-0.26	-0.06	0.17
90	6	1.45	0.15	-0.36	-0.10	0.01
90	12	1.09	0.15	0.18	0.02	3.49
90	18	2.03	0.15	0.38	0.04	3.69
90	24	0.38	0.14	0.42	0.09	4.04
	Average	1.17	0.04	0.08	-0.01	2.41

Note. 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 datasets. Columns 4-6 feature models that are not exactly represented as an affine-$\mathbb{Q}$ model and therefore are misspecified by the investor operating the trading strategy. The last column is a dataset that has been simulated to incorporate a genuine mispricing.

2. the long memory ARFIMA model with $d = 0.49$ and AR(1) coefficient 0.75
3. the non-linear logistic STAR model with parameters $\rho = 0.01$ and $\gamma = 0.5$.

In each of these cases, we simulate a single sample of 1,000 daily term structure observations, and run the same out-of-sample trading strategy used for the variance swap data. As expected, Sharpe ratios in these cases are close to zero.

Lastly, we simulate a term structure that admits mispricing. In particular, the term structure exactly obeys an unrestricted linear factor structure but the loadings are not a consistent geometric progression. We assume that prices on the short end of the curve have factor loadings of $(\sum_{i=1}^{n} \rho_S^i)$ for $n = 1, ..., 11$ while loadings on the long end are $(\sum_{i=1}^{n} \rho_L^i)$ for $n = 12, ..., 24$. We set $\rho_S = 0.95$ and $\rho_L = 0.975$. By the proposition above, this represents

a bona fide mispricing of long maturity prices relative to the short end. In simulation, exploiting the mispricing delivers an out-of-sample Sharpe ratio as high as 4.4.[34] On the short end of the curve, prices are internally consistent, so Sharpe ratios are close to zero when trading at maturities less than 12 months.

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 scale the trading strategy to have a yearly standard deviation of 20%, comparable with the market, and focus on the 18-month case with a mispricing threshold of 50%. The average annual return of this strategy is 24% with an annual Sharpe ratio of 1.18. The alpha relative to the Fama and French (1993) three-factor model is 23% per annum and is highly statistically significant, meaning that essentially none of the strategy's performance is captured by exposure to the Fama-French factors. We obtain nearly identical results (alpha of 24%) when we add two more factors representing shocks to the level and slope of the variance swap curve.[35] The Sharpe ratios associated with this trading strategy thus do not seem explained by exposure to risk factors.

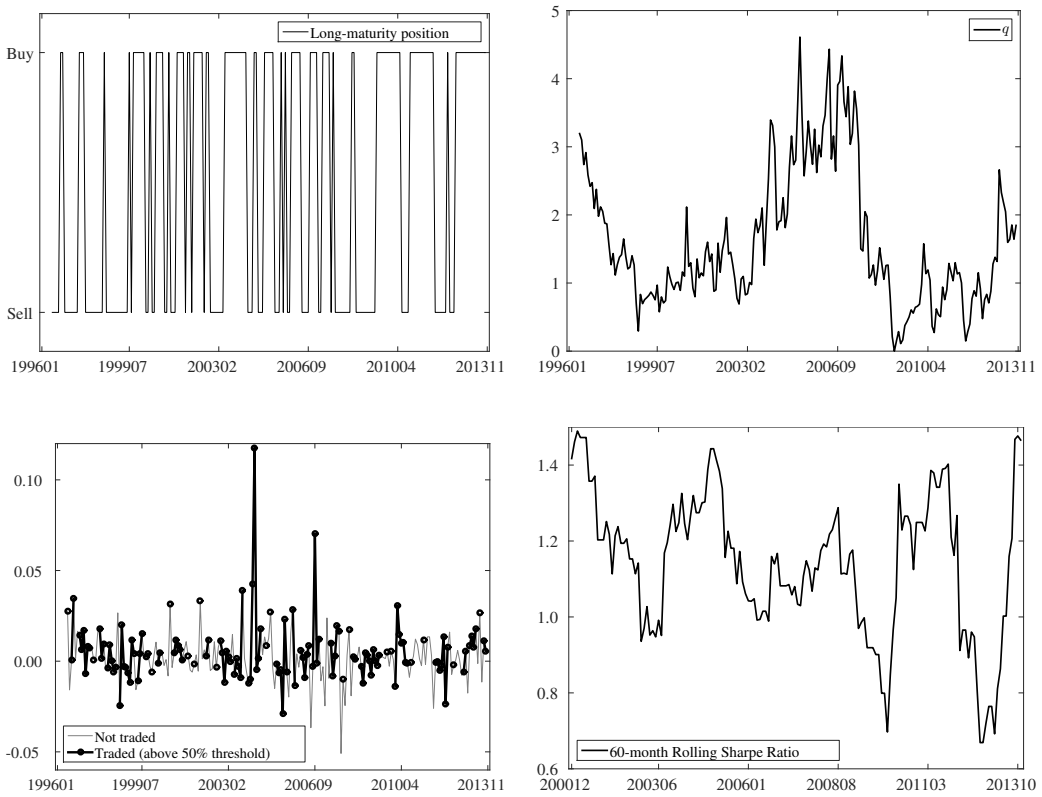
Figure 15 further details the performance of the trading strategy. The upper left panel shows when the strategy calls for a buy or a sell position in the long maturity swap. The strategy frequently changes the direction of the trade. In the average month, the long maturity claim is 27% likely to be traded in the opposite direction that it was the previous month. It is likely that frequent sign switching is the reason why the strategy's returns are essentially uncorrelated with standard risk factors.

The upper right panel shows the number of units that the strategy trades each month (corresponding to q in the trading strategy description, standardized to have a variance of one and a minimum of zero). This panel shows that the strategy does not scale up in riskier times and in fact puts on smaller positions during recessions. This is due to our assumption of fixed capital available for trading. In recessions the price of a variance swap rises so the strategy trades fewer swaps.

[34] Sharpe ratios are not infinite because the size of the profit each period is random and the recursive nature of the trading strategy realistically incorporates estimation noise in model parameters.

[35] 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.

Figure 15: Variance Swap Trading Strategy Performance



Note. 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, scaled number of swaps traded, rolling 60-month Sharpe ratio, and monthly realized returns.

The lower left panel shows the time series of returns to the strategy. The strategy only trades when the signal is sufficiently strong (when the deviation from the model price is greater than the median historical mispricing, given real-time information). Returns during traded months are shown by the heavy black line and marked by circles, and returns in non-traded (weak signal) months are shown in gray. Returns to the strategy are positively skewed, and do not suffer major losses during the 2007-2009 financial crisis. Its largest losses do occur during other risky episodes, however. The worst loss of 3% in May 2005 coincides with the so-called "correlation crisis," triggered by Ford and General Motors losing their investment grade status. The second worst loss is in August 2008 amid the Russian default and LTCM crisis. 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

strategies overall performance, and the rolling Sharpe ratio never falls below 0.65.

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 raises the possibility that the 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.

4.5.2 Extrapolation and Mispricing

In this subsection, we address the question of "What type of investor behavior might lead to mispricing?" by presenting a specific example of a model that results in mispricing of long maturity claims relative to short claims. The example is motivated by a foundational assumption in behavioral economics that investors over-extrapolate when forming expectations. Barberis (2013) explains,

> This assumption is usually motivated by Kahneman and Tversky (1974)'s representativeness heuristic. According to this heuristic, people expect even small samples of data to reflect the properties of the parent population. As a result, they draw overly strong inferences from these small samples, and this can lead to over-extrapolation.

A number of recent models explore the usefulness of extrapolative expectations in matching a variety of asset pricing phenomena, including excess price volatility in equity and credit markets.[36] 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 price volatility along a term structure. 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.[37]

In our stylized example, investor behavior implies an exact affine factor model for term structure prices, but with factor loadings that decay non-geometrically. Suppose that the

[36] For example, Barberis and Shleifer (2003), Greenwood and Shleifer (2014), Barberis et al. (2015*b*), Barberis et al. (2015*a*), Bordalo, Gennaioli and Shleifer (2015), and Gennaioli, Shleifer and Ma (2015).

[37] Furthermore, assets markets that have typically been modeled using extrapolation, such as stocks, mortgages, and corporate bonds, are long-duration assets. Thus, excess volatility in these markets is likely to be a similar phenomenon to the long maturity excess volatility that we document in many other markets.

cash flow process that establishes "correct" prices is a first order autoregression:

$$x_{t+1} = (1-\rho)\mu + \rho x_t + \epsilon_{t+1}.$$

We assume that investors, however, form biased expectations due to extrapolation. Their extrapolative expectations are summarized by replacing the long run mean of cash flows, μ, with a distorted mean,

$$\mu_t^\theta = \mu + \theta(x_t - \mu).$$

The distortion represents the investor's tendency to over-emphasize recent data when contemplating the cash flow distribution. If recent cash flows exceed the long run mean, investors believe that this mean is higher than in fact it is, and vice versa when x_t is below μ. In each period t, μ_t^θ stands in as investors' belief for the long run mean as they value cash flows at all future horizons. This leads to a term structure of forward prices that violates the law of iterated expectations:

$$f_{t,n} = E_t^\theta[x_{t+n}] = E_t[\mu_t^\theta + \rho(x_{t+n-1} - \mu_t^\theta)] = \mu(1-\rho)(1-\theta) + x_t(\rho^n[1-\theta] + \theta).$$

This form of expectation error produces a term structure of prices that is exactly described by an affine one-factor model, but with factor loadings that decay slower than geometrically with maturity. So, by the proposition above, this term structure admits mispricing.

5 Robustness

In this section we examine alternative formulations for the test of cross-equation restrictions. We show that our results are insensitive to measurement error in prices, and we provide additional evidence of excess volatility in subsample analysis.

5.1 Misspecification vs. Measurement Error

In the model setting of Section 2, we assume that prices are observed perfectly. Models of the bond term structure often consider prices that are observed with measurement error (see, for example, Joslin, Singleton and Zhu, 2011). In the presence of measurement error, estimates of ρ can suffer from attenuation bias, and this potentially biases long maturity variance ratios.

While measurement error in prices may impact our tests in theory, Appendix D.4 shows that this is a quantitatively remote possibility in practice for several reasons. First, we

provide simulation-based evidence that measurement error as large as the observed bid-ask spread has essentially no effect on our variance ratio statistic. Second, we show that in order to generate variance ratios in line with those in the data, the standard deviation of measurement error would need to be many times larger than observed bid-ask spreads (seven times larger in the variance swap market, and 10 times in the Treasury market). Third, we show that an instrumental variables correction for measurement error produces variance ratios nearly identical to our baseline results. Fourth, estimates of ρ extracted from various points on the term structure (shown in Figure 3) are gradually increasing in maturity, a pattern at odds with measurement error explanations. Measurement error is likely to be most severe at long maturities where liquidity is lower, which predicts that ρ estimates would decline with maturity, and which is the opposite of what we find in the data.

5.2 Why Test Long Maturities?

Are long maturity claims excessively volatile relative to the affine model, or are short maturity claims not volatile enough? In this subsection we discuss our choice to focus our tests on long maturity price volatility.

The first reason for our emphasis on long maturity excess volatility is that prices on the short end of the term structure do not appear to fluctuate excessively when compared to underlying physical cash flows, while long maturity prices do. While we conduct our main tests using only price data, the affine model also has implications about the comovement of prices with their underlying cash flows, as we analyze in this section.

Variance swaps provide a valuable case study because the underlying cash flow process is observable.[38] Payoffs to these securities are determined by the variance of S&P 500 index returns that is realized over the life of the contract. That is, realized variance (RV_t) corresponds to the cash flow variable x_t in our model. Because realized variance is public information, it serves as a natural anchor for understanding potential over or underreaction of swap prices. For illustration, suppose that the $\mathbb{Q}$-dynamics of realized variance are described by a one-factor model[39]

$$RV_{t+1} = c + \rho RV_t + \epsilon_{t+1}.$$

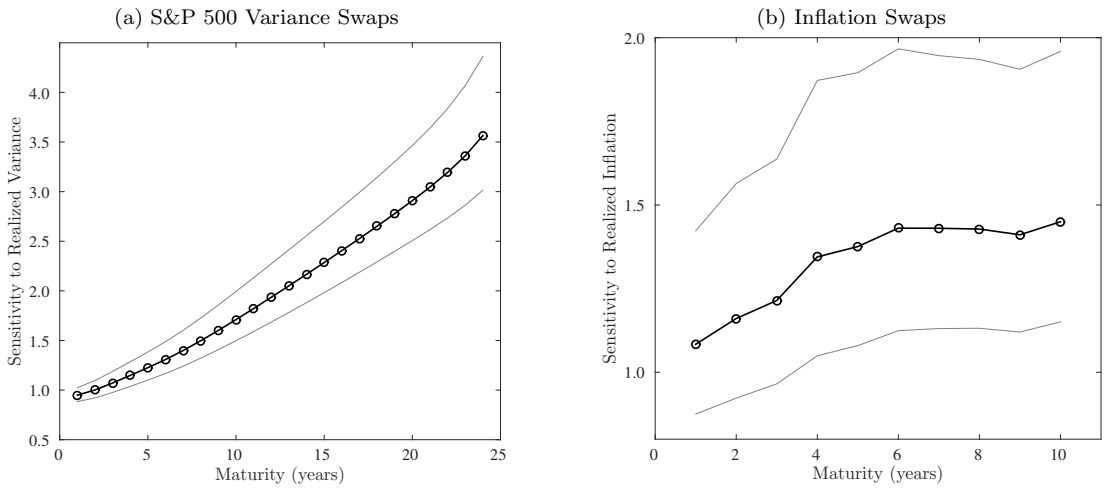
A regression of the two month swap on the one month swap implies a persistence estimate of $\hat{\rho} = 0.83$.

To understand the sensitivity of prices to fluctuations in realized variance, we can scale

[38] Contrast this with, for example, CDS term structures for which the underlying x_t corresponds to an unobservable default intensity.

[39] An unrestricted one-factor model explains 94% of the variation in the variance swap term structure.

Figure 16: Price Sensitivity to Underlying Physical Cash Flows



Note. The figure reports the regression coefficients of cumulative prices (scaled by the model-predicted loadings) on the underlying cash flow, for variance swaps (left) and inflation swaps (right).

the price of the j-maturity claim by the model-predicted loading, $p_{t,j}/\sum_{i=1}^{j}\hat{\rho}^i$, and regress this on RV_t. Following Equation 1, if the model is correctly specified, the scaled price should equal RV_t (plus a constant), and therefore this regression coefficient should equal one.

The left panel of Figure 16 plots the results of these sensitivity regressions. At the short end of the curve, the estimated sensitivity coefficient is 0.95, and the 95% confidence interval includes 1.0, indicating that the one month swap price reacts to realized variance in a manner entirely consistent with the one-factor model. At longer maturities, sensitivities rise sharply above one, suggesting that long maturity prices overreact to fluctuations in realized S&P 500 return variance given a one-factor model.

Another asset class with an observable underlying cash flow process is the inflation swap term structure. These claims pay off realized CPI inflation over the life of the contract. Regressing the scaled inflation swap price,[40] $p_{t,j}/\sum_{i=1}^{j}\hat{\rho}^i$, on realized inflation delivers a sensitivity coefficient of 1.08 for the one year contract and a 95% confidence interval that includes 1.0. The sensitivity estimates increase with maturity and the confidence intervals beyond four years no longer include one.

In summary, short maturity prices comove with current cash flows precisely as the model would predict. On the other hand, the long end of the curve is not only comoving too strongly with short-term prices, but overreacting to realized cash flows as well. This observation

[40] In a one-factor affine model for inflation swaps, the estimated $\mathbb{Q}$-persistence parameter for annual inflation is $\hat{\rho} = 0.46$.

leads us to interpret our results as overreaction at the long end of the curve rather than underreaction at the short end.

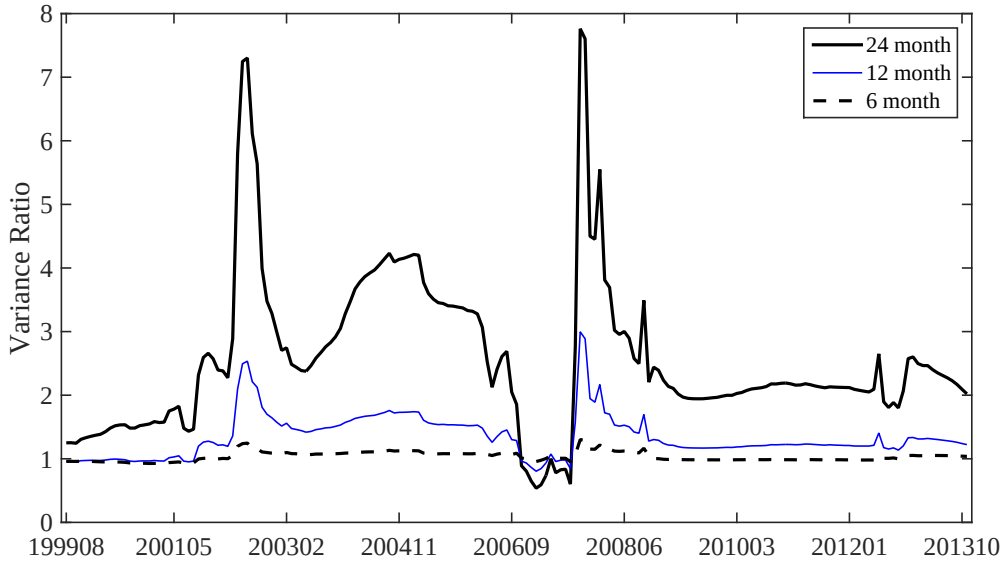
Appendix B.3 provides additional arguments in favor of our approach to testing (and provides supporting estimation and simulation evidence). There, we argue that testing the model using the long or the short end of the curve will yield asymptotically identical test results. We prove this in the analytically tractable one-factor case.

We also compare our variance ratio test with alternative specification tests (based on likelihood ratios) that extract factors from the entire term structure via principal components and use these to assess the overall fit of the model. Tests that extract factors from the entire term structure can lack power to detect the type of model violation that we document in the data. To show this, we simulate data sets that feature excess volatility only at the long end of the curve, and find that full term structure likelihood-based criteria reject the misspecified model less frequently than our variance ratio tests. These simulations capture the intuition that likelihood-based comparisons of the full term structure lack power because they average over model errors at all maturities. But prices of some claims (especially at the short half of the curve) are consistent with the null model. Averaging errors from the short end together with errors from the long end dulls the test's ability to detect the model violation at the long end. In contrast, variance ratios deliver a pointed evaluation of the consistenty between prices at short maturities and *each specific* long maturity. Furthermore, even if a likelihood-based test rejects a model, it is generally silent on the *reason* for the rejection, while our variance ratio statistics clearly illustrate which parts of the term structure are inconsistent with each other.

5.3 Stability in Subsamples

One advantage of our test is that it only uses comovement among prices to estimate and test the model. These covariances are precisely estimated even when a short time series is available, therefore our test can be conducted within short rolling subsamples. In Figure 17 we report variance ratio estimates in the variance swap market for a four-year rolling window. Variance ratios at 24 months are far above one for the majority of the sample, and reach peaks of nearly 8 in some subperiods. This demonstrates first that our findings are robust to alternative data samples. Perhaps more importantly, it illustrates that our main results are unlikely to be driven by instability of the affine model. For rolling estimation windows as short as six months we find results quantitatively similar to our full sample estimates.

Figure 17: Variance Swaps: rolling estimates and tests



Note. The figure plots the VR at maturities 6 months, 12 months, and 24 months obtained using a rolling 4-year window for estimation and testing.

6 Conclusion

We find that prices of long maturity claims are dramatically 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 expectations 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, 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 $\mathbb{Q}$ risk-neutral model dynamics. Only for the term structure of Treasuries is the degree of excess volatility economically small, consistent with the historical success of affine models for describing interest rates.

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 non-affine models, none of which appear

to capture the behavior of long maturity claims in the data. Lastly, we show that trading against long maturity excess volatility appears profitable after adjusting for exposure to standard risk factors. But our analysis into the sources of excess volatility is by no means exhaustive and calls for deeper investigation in future research. Our findings also call for more research into how agents form expectations over multiple horizons and the extent to which investor behavior is consistent with the law of iterated values.

References

Ait-Sahalia, Yacine, Mustafa Karaman, and Loriano Mancini. 2014. "The Term Structure of Variance Swaps, Risk Premia and the Expectations Hypothesis." Working paper.

Anderson, Evan W, Lars Peter Hansen, and Thomas J Sargent. 2003. "A quartet of semigroups for model specification, robustness, prices of risk, and model detection." *Journal of the European Economic Association*, 68–123.

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

Backus, David, Nina Boyarchenko, and Mikhail Chernov. 2015. "Term structures of asset prices and returns."

Bansal, Ravi, and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance*, 59(4): 1481–1509.

Barberis, Nicholas, and Andrei Shleifer. 2003. "Style investing." *Journal of Financial Economics*, 68(2): 161–199.

Barberis, Nicholas C. 2013. "Psychology and the Financial Crisis of 2007–2008." *Financial Innovation: Too Much Or Too Little?*, 15.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. 2015*a*. "Extrapolation and Bubbles." *Harvard and Yale unpublished manuscript*.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. 2015*b*. "X-CAPM: An extrapolative capital asset pricing model." *Journal of Financial Economics*, 115(1): 1–24.

Barsky, Robert B, and J Bradford De Long. 1993. "Why Does the Stock Market Fluctuate?" *The Quarterly Journal of Economics*, 108(2): 291–311.

Bekaert, Geert, and Robert J Hodrick. 2001. "Expectations hypotheses tests." *The journal of finance*, 56(4): 1357–1394.

Bikbov, Ruslan, and Mikhail Chernov. 2009. "Unspanned stochastic volatility in affine models: evidence from eurodollar futures and options." *Management Science*, 55(8): 1292–1305.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2015. "Investor Psychology and Credit Cycles Pedro Bordalo, Nicola Gennaioli, Andrei Shleifer October 2015." *Harvard working paper*.

Britten-Jones, Mark, and Anthony Neuberger. 2000. "Option prices, implied price processes, and stochastic volatility." *The Journal of Finance*, 55(2): 839–866.

Calvet, Laurent E, and Adlai J Fisher. 2004. "How to forecast long-run volatility: regime switching and the estimation of multifractal processes." *Journal of Financial Econometrics*, 2(1): 49–83.

Campbell, John Y. 1987. "Stock Returns and the Term Structure." *Journal of Financial Economics*, 18(2): 373–399.

Campbell, John Y. 1991. "A Variance Decomposition for Stock Returns." *The Economic Journal*, 101(405): 157–179.

Campbell, John Y. 1995. "Some Lessons from the Yield Curve." *Journal of Economic*

Perspectives, 9: 129–152.

Campbell, John Y., and John H. Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy*, 107(2): 205–251.

Campbell, John Y., and Robert J. Shiller. 1987. "Cointegration and Tests of Present Value Models." *Journal of Political Economy*, 95(5): 1062–1088.

Campbell, John Y., and Robert J. Shiller. 1988*a*. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies*, 1(3)(3): 195–228.

Campbell, John Y., and Robert J. Shiller. 1988*b*. "Stock Prices, Earnings, and Expected Dividends." *Journal of Finance*, XLIII: 661–676.

Campbell, John Y., and Robert J. Shiller. 1991. "Yield Spreads and Interest Rate Movements: A Bird's Eye View." *Review of Economic Studies*, 58(3): 495–514.

Carr, Peter, and Roger Lee. 2009. "Volatility derivatives." *Annu. Rev. Financ. Econ.*, 1(1): 319–339.

Casassus, Jaime, and Pierre Collin-Dufresne. 2005. "Stochastic convenience yield implied from commodity futures and interest rates." *The Journal of Finance*, 60(5): 2283–2331.

Cochrane, John H. 1992. "Explaining the variance of price-dividend ratios." *Review of Financial Studies*, 5(2): 243–280.

Cochrane, John H. 2008. "The Dog That Did Not Bark: A Defense of Return Predictability." *Review of Financial Studies*, 21(4): 1533–1575.

Cochrane, John H. 2011. "Discount Rates." *Journal of Finance*, 66(4): 1047–1108. AFA Presidential Address.

Cochrane, John H., and Monika Piazzesi. 2005. "Bond Risk Premia." *American Economic Review*, 95(1): 138–160.

Cochrane, John H, and Monika Piazzesi. 2009. "Decomposing the yield curve."

Collin-Dufresne, Pierre, and Robert Goldstein. 2002. "Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility." *Journal of Finance*, 57(4): 1685–1730.

Creal, Drew, and Jing Cynthia Wu. 2015. "Estimation of Affine Term Structure Models with Spanned or Unspanned Stochastic Volatility." *Journal of Econometrics*, 185(1): 60–81.

Dai, Qiang, and Kenneth J. Singleton. 2002. "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure." *Journal of Financial Economics*, 63(3): 415–441.

Dai, Qiang, and Kenneth Singleton. 2003. "Term structure dynamics in theory and reality." *Review of Financial Studies*, 16(3): 631–678.

Davidson, Russell, and James G MacKinnon. 2004. *Econometric theory and methods.* Vol. 5, Oxford University Press New York.

Dew-Becker, Ian, and Stefano Giglio. 2013. "Asset Pricing in the Frequency Domain: Theory and Empirics." Working paper.

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

Drechsler, Itamar, and Amir Yaron. 2011. "What's Vol Got to Do with it?" *The Review of Financial Studies*, 24(1): 1–45.

Duffee, Gregory R. 2002. "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance*, 57(1): 405–443.

Duffee, Gregory R. 2010. "Sharpe ratios in term structure models." Working paper.

Duffie, Darrell. 2010. *Dynamic asset pricing theory.* Princeton University Press.

Duffie, Darrell, and Kenneth J Singleton. 1999. "Modeling term structures of defaultable bonds." *Review of Financial studies*, 12(4): 687–720.

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

Duffie, Darrell, Rui Kan, et al. 1996. "A yield-factor model of interest rates." *Mathematical finance*, 6: 379–406.

Egloff, Daniel, Madrkus Leippold, and Liuren Wu. 2010. "The Term Structure of Variance Swap Rates and Optimal Variance Swap Investments." *Journal of Financial and Quantitative Analysis*, 45(5): 1279–1310.

Fama, Eugene F. 1970. "Efficient capital markets: A review of theory and empirical work*." *The journal of Finance*, 25(2): 383–417.

Fama, Eugene F. 1991. "Efficient capital markets: II." *The journal of finance*, 46(5): 1575–1617.

Fama, Eugene F, and Kenneth R French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of financial economics*, 33(1): 3–56.

Fama, Eugene F., and Robert R. Bliss. 1987. "The Information in Long-Maturity Forward Rates." *The American Economic Review*, 77(4): 680–692.

Fleckenstein, Matthias, Francis A Longstaff, and Hanno Lustig. 2013. "Deflation risk." National Bureau of Economic Research.

Fleming, Michael J, and John Sporn. 2013. "Trading activity and price transparency in the inflation swap market." *Economic Policy Review*, 19(1).

Gabaix, Xavier. 2012. "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance." *Quarterly Journal of Economics*, 127(2): 645–700.

Gennaioli, Nicola, Andrei Shleifer, and Yueran Ma. 2015. "Expectations and Investment."

Giglio, Stefano, Matteo Maggiori, and Johannes Stroebel. 2015*a*. "No-Bubble Condition: Model-free Tests in Housing Markets." *Econometrica*, forthcoming.

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

Granger, Clive WJ, and Roselyne Joyeux. 1980. "An introduction to long-memory time series models and fractional differencing." *Journal of time series analysis*, 1(1): 15–29.

Granger, Clive WJ, and Timo Terasvirta. 1993. "Modelling non-linear economic relationships." *OUP Catalogue.*

Greenwood, Robin, and Andrei Shleifer. 2014. "Expectations of returns and expected returns." *Review of Financial Studies*, hht082.

Gurkaynak, Refet S., Brian Sack, and Eric Swanson. 2005. "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models." *American Economic Review*, 95(1): 425–436.

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

Hamilton, James D., and Cynthia Wu. 2012. "Identification and Estimation of Gaussian Affine-Term-Structure Models." *Journal of Econometrics*, 168(2): 315–331.

Hansen, Lars P., and Jose A. Scheinkman. 2009. "Long-term risk: an operator approach." *Econometrica*, 77: 177–234.

Hansen, Lars Peter. 2012. "Dynamic valuation decomposition within stochastic economies." *Econometrica*, 80(3): 911–967.

Hansen, Lars Peter. 2014. "Nobel Lecture: Uncertainty Outside and Inside Economic Models." *Journal of Political Economy*, 122(5): 945–987.

Hansen, Lars Peter, and Scott F Richard. 1987. "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models." *Econometrica: Journal of the Econometric Society*, 587–613.

Hansen, Lars Peter, and Thomas J Sargent. 1980. "Formulating and estimating dynamic linear rational expectations models." *Journal of Economic Dynamics and Control*, 2: 7–46.

Hanson, Samuel G, and Jeremy C Stein. 2015. "Monetary policy and long-term real rates." *Journal of Financial Economics*, 115(3): 429–448.

Herskovic, Bernard, Bryan T Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh. 2014. "The common factor in idiosyncratic volatility: Quantitative asset pricing implications." National Bureau of Economic Research.

Jiang, George J., and Yisong S. Tian. 2005. "The Model-Free Implied Volatility and its Information Content." *Review of Financial Studies*, 18(4): 1305–1342.

Joslin, Scott. 2006. "Can unspanned stochastic volatility models explain the cross section of bond volatilities."

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu. 2011. "A New Perspective on Gaussian Dynamic Term Structure Models." *The Review of Financial Studies*, 24(3): 926–970.

Kahneman, Daniel, and Amos Tversky. 1974. "Subjective probability: A judgment of representativeness." In *The Concept of Probability in Psychological Experiments*. 25–48. Springer.

Kelly, Bryan, Hanno Lustig, and Stijn Van Nieuwerburgh. 2013. "Firm volatility in granular networks." National Bureau of Economic Research.

Le, Anh, Kenneth J Singleton, and Qiang Dai. 2010. "Discrete-Time AffineQ Term Structure Models with Generalized Market Prices of Risk." *Review of Financial Studies*, 23(5): 2184–2227.

LeRoy, Stephen F., and Richard D. Porter. 1981. "The Present-Value Relation: Tests Based on Implied Variance Bounds." *Econometrica*, 49(3): 555–574.

Pástor, Ľuboš, and Pietro Veronesi. 2003. "Stock Valuation and Learning about Profitability." *The Journal of Finance*, 58(5): 1749–1790.

Pástor, Ľuboš, and Pietro Veronesi. 2009*a*. "Learning in Financial Markets." *Annu. Rev. Financ. Econ.*, 1(1): 361–381.

Pástor, Ľuboš, and Pietro Veronesi. 2009*b*. "Technological Revolutions and Stock

Prices." *The American Economic Review*, 1451–1483.

Piazzesi, Monika. 2010. "Affine Term Structure Models." In *Handbook of Financial Econometrics*. Elsevier.

Pontiff, Jeffrey. 1997. "Excess volatility and closed-end funds." *The American Economic Review*, 155–169.

Samuelson, Paul A. 1965. "Proof that properly anticipated prices fluctuate randomly." *Industrial management review*, 6(2): 41–49.

Shiller, Robert J. 1979. "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure." *Journal of Political Economy*, 87(6): 1190–1219.

Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71(3): 421–436.

Singleton, Kenneth J. 1980. "Expectations models of the term structure and implied variance bounds." *The Journal of Political Economy*, 1159–1176.

Stein, Jeremy. 1989. "Overreactions in the Options Market." *Journal of Finance*, 44(4): 1011–1023.

Teräsvirta, Timo. 1994. "Specification, estimation, and evaluation of smooth transition autoregressive models." *Journal of the american Statistical association*, 89(425): 208–218.

Timmermann, Allan G. 1993. "How learning in financial markets generates excess volatility and predictability in stock prices." *The Quarterly Journal of Economics*, 1135–1145.

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

van Binsbergen, Jules, Wouter Hueskes, Ralph Koijen, and Evert Vrugt. 2013. "Equity yields." *Journal of Financial Economics*, 110(3): 503–519.

Vasicek, Oldrich. 1977. "An equilibrium characterization of the term structure." *Journal of financial economics*, 5(2): 177–188.

Veronesi, Pietro. 1999. "Stock market overreactions to bad news in good times: a rational expectations equilibrium model." *Review of Financial Studies*, 12(5): 975–1007.

Wachter, Jessica A. 2005. "Solving models with external habit." *Finance Research Letters*, 2(4): 210–226.

Wachter, Jessica A. 2013. "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?" *Journal of Finance*, 68(3): 987–1035.

Internet Appendix

A Model Identification and Estimation

In this appendix we show how to estimate the matrix B of size $K \times K$ in a setting in which: 1) any $G \geq K$ maturities at the short end of the curve are observed and are used to construct the K factors, and 2) the G maturities observed are not necessarily consecutive (for example, one wants to extract $K = 2$ factor using maturities 1, 2, 4 or 1, 2, 4, 6). If $G > K$, the first K principal components of the G observed maturities are used as factors H_t (which will still be a K-element vector). We proceed with the derivation assuming cumulative claims are used, but an equivalent derivation holds when using forwards.

We refer to the G maturities observed at the short end of the curve as $n_1, ..., n_G$, and to the vector of those prices as $P_{t,G}$. We assume that each individual observed price in the term structure has potential measurement error:

$$p_{t,n} = \mathbf{1}' \left[\rho + \rho^2 + ... + \rho^n\right] H_t + u_{t,n}$$

while the first K principal components of $P_{t,G}$ are observed without error:

$$\overline{P}_t = f \cdot P_{t,G}$$

where f is an $K \times G$ matrix selecting the first K principal components of $P_{t,G}$ (referred to as $\overline{P}_t$). Naturally, this nests the case (studied in the paper) in which exactly the first K maturities are observed without error and used as factors ($K = G$): in that case, f is simply the identity matrix.

We can prove the following Proposition, that shows how to recover the matrix ρ in this setting (and therefore in turn recover all loadings of long-term prices onto the short-end factors $\overline{P}_t$ under the model).

Proposition 1. *Consider the regression of a price $p_{t,n_{G+1}}$ onto the factors $\overline{P}_t$:*

$$p_{t,n_{G+1}} = d + c'\overline{P}_t + u_{t,n_{G+1}}$$

All eigenvalues ρ_i of ρ are among the roots of the polynomial equation

$$\left[1 + \rho_i + ... + (\rho_i)^{n_{G+1}-1}\right] = \tilde{c}_1 \left[1 + \rho_i + ... + (\rho_i)^{n_1-1}\right] + ... + \tilde{c}_G \left[1 + \rho_i + ... + (\rho_i)^{n_G-1}\right]$$

where all of the coefficients $\tilde{c}$ depend exclusively on the factor loadings f and on the regression coefficients c.

Proof. Start by defining

$$S^n \equiv \mathbf{1}'(\rho + \rho^2 + ... + \rho^n)$$

S^n is a $1 \times K$ vector that depends *only* on the diagonal matrix ρ. We can therefore write for each price:

$$p_{t,n} = S^n H_t + u_{t,n}$$

Note that calling ρ_i the ith element of the diagonal of ρ, we can rewrite S^n as:

$$S^n = \begin{bmatrix} \rho_1 + ... + \rho_1^n \\ \rho_2 + ... + \rho_2^n \\ ... \\ \rho_K + ... + \rho_K^n \end{bmatrix}'$$

The assumption that the principal components $\overline{P}_t$ are observed without error yields:

$$\overline{P}_t = f \cdot \begin{bmatrix} S^{n_1} \\ S^{n_2} \\ ... \\ S^{n_G} \end{bmatrix} H_t$$

where n_1, n_2, ... are the observed maturities. Consider now the regression (allowing for sample error):

$$p_{t,n_{G+1}} = d + c'\overline{P}_t + u_{t,n_{G+1}}$$

and project each side of the equation on H_t (noting that $u_{t,n_{G+1}}$ is orthogonal to H_t). The loadings on H_t on the two sides of the equation must match. Therefore:

$$S^{n_{G+1}} = c'f \begin{bmatrix} S^{n_1} \\ S^{n_2} \\ ... \\ S^{n_G} \end{bmatrix} = \tilde{c} \begin{bmatrix} S^{n_1} \\ S^{n_2} \\ ... \\ S^{n_G} \end{bmatrix}$$

where the last equaility is obtained by defining $\tilde{c} = c'f$, a $1 \times H$ vector that depends only the factor loadings f and the regression coefficients c. We can then write:

$$S^{n_{G+1}} = \tilde{c}_1 S^{n_1} + \tilde{c}_2 S^{n_2} + ... + \tilde{c}_G S^{n_G}$$

where $\tilde{c}_i$ is the i-th element of $\tilde{c}$. Now, given that as shown above each element i of S^n depends only on element i of the diagonal of ρ, this is a system of K independent equations, each of the form:

$$\left[\rho_i + \rho_i + ... + \rho_i^{n_{G+1}-1}\right] = \tilde{c}_1 \left[\rho_i + \rho_i + ... + \rho_i^{n_1-1}\right] + ... + \tilde{c}_G \left[\rho_i + \rho_i + ... + \rho_i^{n_G-1}\right]$$

Finally, we can divide by ρ_i throughout (assuming $\rho_i \neq 0$) and obtain:

$$\left[1 + \rho_i + ... + (\rho_i)^{n_{G+1}-1}\right] = \tilde{c}_1 \left[1 + \rho_i + ... + (\rho_i)^{n_1-1}\right] + ... + \tilde{c}_G \left[1 + \rho_i + ... + (\rho_i)^{n_G-1}\right]$$

□

Note that each element i of ρ needs to satisfy this equation: the matrix ρ can therefore be computed by finding the roots of this polynomial equation. This structure has the convenient feature that we can estimate state dynamics from the yields without any maximization (as is typical in term structure models).

Once ρ has been recovered, we can construct S^n for each maturity n. Since

$$\overline{P}_t = \overline{S} H_t$$

where

$$\overline{S} = f \begin{bmatrix} S^{n_1} \\ S^{n_2} \\ \dots \\ S^{n_G} \end{bmatrix}$$

is a $K \times K$ matrix, we can write:

$$H_t = \overline{S}^{-1} \overline{P}_t$$

Therefore, we can also write

$$p_{t,n} = S^n \overline{S}^{-1} \overline{P}_t + u_{t,n}$$

The matrix of loadings on the "observable factors" $\overline{P}_t$ is therefore $S^n \overline{S}^{-1}$. These factors can be used to construct a variance ratio test that compares the variance of the component of $p_{t,n}$ predicted (in unrestricted regressions) by the factors $\overline{P}_t$ to the variance predicted under the model (with coefficients $S^n \overline{S}^{-1}$).

One final consideration is that there will generally be $n_{G+1} - 1$ roots of this polynomial (some of them potentially complex or explosive), while we only seek K parameters. This equation shows that the $\mathbb{Q}$ dynamics and the comovements of prices only identify the eigenvalues of ρ up to the set of roots of this polynomial. It does not tell us which roots to choose, as they imply the *same* covariance among prices (while a full MLE procedure that exploits both information about the $\mathbb{P}$ and the $\mathbb{Q}$ dynamics will be able to choose among them). Of course, in our baseline case, where we only select the first K prices as factors, we will always have as many roots as parameters (K).

We use the following selection procedure for the roots. First, we only consider non-explosive roots. This is motivated by the unambiguous empirical fact that price variances are concave in maturity for all the markets we study, especially at the short end of the curve where our estimation is coming from. If prices rise less than linearly with horizon, the system is best described by stationary dynamics. Second, among the non-explosive roots, we select the K most persistent ones. This ensures that our excess volatility findings will be the most conservative (they will suggest the least excess volatility) of all of the covariance-equivalent roots we could have reported. Finally, following the term structure literature, we choose real roots whenever possible.

B Model Testing

B.1 Bootstrap Inference

We obtain bootstrap standard errors using the semiparametric bootstrap procedure described in Davidson and MacKinnon (2004). Bootstrap standard errors are used to test the null hypothesis that the variance ratio at a given maturity $n > K$ is equal to one, or equivalently

that the covariance of prices at maturity $n > K$ are consistent with the model estimated from the vector of prices at maturities $1{:}K+1$.

The bootstrap proceeds as follows. First, we construct fitted errors under the null for each time t and all maturities n as:

$$\hat{\epsilon}_{t,n} = p_{t,n} - \hat{p}_{t,n}$$

where $\hat{p}_{t,n}$ is the price predicted by the model under the null (and relying on $\hat{\rho}$ estimated from the regression of $p_{t,K+1}$ on $P_{t,1:K}$). Next, an AR(1) for the errors is estimated for each maturity:

$$\hat{\epsilon}_{t,n} = \gamma_n \hat{\epsilon}_{t-1,n} + \hat{u}_{t,n}$$

This step allows us to explicitly account for the time-series correlation properties of the errors.

Each bootstrap sample is obtained by jointly resampling the error innovations $\hat{u}_{t,n}$ across maturities. Denote with tildes the quantities that are generated in each bootstrap sample; for example, the resampled error innovations u are denoted $\tilde{u}_{t,n}$. Using the estimated persistence $\hat{\gamma}_n$ for each maturity, together with the resampled error innovations $\tilde{u}_{t,n}$, we generate a panel of resampled errors $\tilde{\epsilon}_{t,n}$. The bootstrapped prices are then constructed as:

$$\tilde{p}_{t,n} = \hat{p}_{t,n} + \tilde{\epsilon}_{t,n}$$

Using the resample term structure of prices constructed the bootstrap sample prices, $\tilde{p}_{t,n}$, we re-run our entire analysis. Importantly, we re-estimate the matrix ρ in the bootstrap sample (obtaining $\tilde{\rho}$) and obtain the variance ratio test statistic. Because we re-estimate ρ in each bootstrap sample, our procedure takes into account sampling uncertainty regarding the decay rate under $\mathbb{Q}$. We conduct all of bootstrap inference using 1,000 bootstrap samples.

We have derived the analytical asymptotic distribution of the variance ratio statistic and compared this with the finite sample bootstrap-based inference, and they behave similarly in moderately sized samples. We find that bootstrap standard errors are more conservative in small samples and thus base our main analysis on these. Details for the derivation of the asymptotic distribution and its comparison with the bootstrap distribution are available upon request.

B.2 Finite Sample Simulations

Our approach to inference for variance ratios relies on factor persistences estimated from prices on the short end of the term structure. A natural concern is that it may be hard to estimate the behavior of a small but very persistent factor from the short end alone. In other words, even when the model is correctly specified, one may be concerned that short end prices are not informative enough about the $\mathbb{Q}$-dynamics of low volatility/high persistence factors, and that this may lead to inappropriate inference. In this appendix, we show that this is not the case. Short maturity prices are sufficiently informative about low frequency $\mathbb{Q}$-dynamics so that our variance ratio tests always retain correct size. That is, when the null hypothesis is true, we reject the null approximately 5% of the time when we use a 5%

Table A5: Simulated Variance Ratio Tests Under Correct Specification

	$\sigma_2^2/\sigma_1^2 = 0.25$			$\sigma_2^2/\sigma_1^2 = 0.10$			$\sigma_2^2/\sigma_1^2 = 0.05$			$\sigma_2^2/\sigma_1^2 = 0.01$		
ρ_2	5%	10%	$\frac{\text{Std}(VR)}{\text{BSE}(VR)}$	5%	10%	$\frac{\text{Std}(VR)}{\text{BSE}(VR)}$	5%	10%	$\frac{\text{Std}(VR)}{\text{BSE}(VR)}$	5%	10%	$\frac{\text{Std}(VR)}{\text{BSE}(VR)}$
0.9000	0.103	0.149	0.896	0.129	0.169	0.873	0.113	0.162	0.835	0.030	0.052	1.836
0.9500	0.043	0.097	0.876	0.078	0.123	0.871	0.094	0.130	0.862	0.136	0.167	0.786
0.9900	0.009	0.041	0.911	0.023	0.061	0.853	0.032	0.075	0.871	0.067	0.105	0.885
0.9990	0.019	0.037	1.182	0.024	0.054	1.052	0.033	0.069	0.951	0.071	0.116	0.942
0.9999	0.057	0.091	1.447	0.060	0.104	1.070	0.067	0.109	0.978	0.099	0.152	0.958

Note. Realized rejection rates across 5,000 simulations at 5% and 10% bootstrap critical values. $\frac{\text{Std}(VR)}{\text{BSE}(VR)}$ is the ratio of the standard deviation of 24-month variance ratio statistics to the median bootstrap standard error across simulations.

critical value, we reject the null approximately 10% of the time when we use a 10% critical value, and so forth. In other words, our rejection of the affine model is not driven by our choice to estimate model parameters using short end prices.

Our estimation and inference procedure is well behaved because our bootstrap distribution for the variance ratio statistic takes into account sampling variation in the parameters estimated from the short end. If there are some parameters that are hard to accurately estimate (for example, the persistence parameter for a low variance factor), the variation in bootstrap samples fully accounts for this.

To understand the performance of our inference approach we conduct simulations. We generate term structures of prices with maturities up to 24 periods assuming a two-factor model. The first factor is the dominant factor and has variance $\sigma_1^2 = 1$ and persistence $\rho_1 = 0.75$. For the weaker second factor, we consider a gradually decreasing range of variance ($\sigma_2^2/\sigma_1^2 \to 0$) and a gradually increasing range of persistence ($\rho_2 \to 1$). Based on 1,000 periods of simulated term structure prices, we estimate the model using the shortest maturities (1,2, and 3) and calculate the variance ratio statistic, its standard error, and its p-value for the 24-month claim. We generate 5,000 such samples at each set of parameters, and report summary statistics across simulations. We report the realized rejection rates based on 5% and 10% critical values of the test. We also report the ratio of the standard deviation of the variance ratio statistic to the median bootstrap standard error across simulation; which should be near one if the test is behaving appropriately. Results are shown in Table A5.

Overall, finite sample inference behaves reasonably. The test seems to reject too *infrequently*, and the realized standard deviation of the variance ratio statistic tends to be slightly smaller than the asymptotic standard error. These facts indicate that the critical values that we use in our empirical analysis are slightly conservative.

B.3 Long Maturities vs. Short Maturities

In this section we provide additional motivational evidence for our choice of testing overreaction of the long end of the curve relative to the short end.

B.3.1 Alternative Formulations of the Test

While Figure 16 motivates our test's emphasis on long maturity excess volatility, there are a multitude of ways to formulate tests of cross-equation restrictions. One natural alternative to our approach is to estimate model parameters from the long end of the term structure, and perform variance ratio tests on the short end. This alternative is statistically equivalent to the test that we propose, but can conceal important model violations.

For illustration, consider a setting where prices in fact obey a strict one-factor structure, but where the no-arbitrage cross-equation restrictions are violated. In particular, suppose that prices on the short end of the term structure (j=1,2) behave according to $p_{t,j} = (\rho_S + ... + \rho_S^j)x_t$ while prices on the long end (j=N-1,N) are $p_{t,j} = (\rho_L + ... + \rho_L^j)x_t$, with $\rho_L > \rho_S > 0$.

In the population version of our baseline test, we estimate the model parameter from a regression of $p_{t,2}$ on $p_{t,1}$ and therefore recover ρ_S, which we use to impose model restrictions. Next we estimate an unrestricted regression of long maturity price $p_{t,N}$ on $p_{t,1}$, which has a coefficient of $Cov(p_{t,N}, p_{t,1})/Var(p_{t,1}) = (\rho_L + ... + \rho_L^N)/\rho_S$. We compare this to the restricted regression of $p_{t,N}$ on $p_{t,1}$ imposing $\rho_L = \rho_S$, which implies a coefficient of $(\rho_S + ... + \rho_S^N)/\rho_S$. The variance ratio for the long maturity test is therefore

$$VR_N = \left(\frac{\rho_L + ... + \rho_L^N}{\rho_S + ... + \rho_S^N}\right)^2.$$

In the alternative approach of estimating from the long end and testing on the short end, the model parameter is derived from regressing $p_{t,N}$ on $p_{t,N-1}$, yielding an estimate equal to ρ_L. The unrestricted regression coefficient of the short maturity price $p_{t,1}$ on $p_{t,N}$ is $\rho_S/(\rho_L + ... + \rho_L^N)$ and the restricted coefficient is $\rho_L/(\rho_L + ... + \rho_L^N)$. The variance ratio for the short maturity test is therefore

$$VR_1 = \left(\frac{\rho_S}{\rho_L}\right)^2.$$

Clearly, tests based on VR_1 and VR_N are equivalent as deviations from unity occur in both cases if and only if $\rho_L \neq \rho_S$. An important difference between the two tests is how they aggregate specification errors along the term structure. A value of VR_1 near to but just below one may indicate an important model violation. For example, if $\rho_S = 0.97$ and $\rho_L = 0.99$ and we are considering maturities up to 24 periods, then $VR_1 = 0.92$ and $VR_{24} = 2.49$. In this example, the model violation is one of high duration. Its impact on the behavior of short maturity claims is limited, as indicated by the small deviation of VR_1 from one, while it is a crucial violation for understanding the pricing of long maturity claims.

This example is representative of price behavior in all asset classes we study. Term structure data very broadly imply high cash flow persistence, so the most useful securities for identifying model violations are those with long maturities. Prices of these claims aggregate parameter discrepancies over long horizons, making it particularly easy to visualize the internal inconsistency of prices for a given model, as in Figure 1.

Another testing approach is to use a likelihood ratio or other distance metric to compare pricing errors between two models—one model that imposes pricing restrictions versus a

Table A6: Model Comparison Using Full Term Structure

Misspecification					
ρ_S	ρ_L	$\text{MSE}_1/\text{MSE}_2$	BIC_1	BIC_2	VR_{24}
0.75	0.80	1.020	-866	**-909**	1.76
0.75	0.85	1.057	-674	**-688**	3.43
0.75	0.89	1.078	**-580**	-558	6.43
0.85	0.90	1.011	**-795**	-768	2.23
0.85	0.95	1.024	**-634**	-551	5.87
0.85	0.99	1.023	**-575**	-453	14.61
0.90	0.95	1.006	**-796**	-707	2.64
0.90	0.99	1.010	**-667**	-544	6.56
0.95	0.99	1.002	**-847**	-724	2.49

Note. The table reports statistical tests of an affine model with violation of arbitrage. The model is a one factor model, but specified such that for maturities up to 12, Q-persistence is ρ_S, while for maturities above 12 it is ρ_L. The table reports the ratio of mean squared pricing errors for a 1 and 2 factor model as well as the BIC criterion for one and two factors. The last column reports the variance ratio at 24 months using the number of factors selected by the BIC criterion.

more general model with weaker restrictions—using prices from the entire term structure for estimation (as in Bekaert and Hodrick, 2001). In our setting, a natural implementation of this approach would estimate latent factors by extracting principal components from the panel of all maturities. Then, parameters of the null model (e.g., a one-factor affine no-arbitrage model) are estimated by minimizing pricing errors, and the overall fit is compared to that of a specific alternative (e.g., a two-factor model). This test has the benefit of using information throughout the term structure and generally has excellent power for distinguishing between alternative models.

We find, however, that this approach often lacks power to reject the null model in the presence of long maturity overreaction like that documented in Section 3. A simulation is helpful for understanding how standard model comparison tests can fail to detect overreaction. We generate data from a one-factor model with maturities of up to 24 periods. Simulated prices behave very similarly to an affine model except that long maturity prices overreact and therefore violate the no-arbitrage internal consistency conditions. In particular, for short maturities (j=1,...,12), factor loadings are given by $\sum_{i=1}^{j} \rho_S^i$, while for long maturities (j=13,...,24) the loadings are $\sum_{i=1}^{j} \rho_L^i$, where $\rho_L > \rho_S$.

In each simulation, we generate 10 years of monthly data. From simulated prices, we estimate a one-factor model that extracts a single component from the full panel of prices, then estimate the model's single model parameter ρ by minimizing the sum of squared pricing errors at all maturities. We compare this fit to an otherwise identical model that allows for two principal component factors, again estimating this model's two persistence parameters by minimizing pricing errors. We then compare the models in two ways. First, we calculate the ratio of mean squared pricing errors for the one-factor and two factor model ($\text{MSE}_1/\text{MSE}_2$). Adding factors can only improve the model's fit. High values indicate that moving to two factors produces a large improvement in fit. We also report the Bayesian

Information Criterion (BIC) for each model assuming that errors are normally distributed. The BIC trades off model fit versus parameterization, with lower values of BIC indicating a superior model. Because the BIC is based on log likelihoods of the estimated models, BIC comparison is conceptually similar to conducting a likelihood ratio test.

Table A6 shows simulation results. We consider various degrees of model misspecification described by a given combination of ρ_S and ρ_L, and report the average of each model statistic across 1,000 simulations. Overall, simulations show that it is difficult to reject the one-factor model based by comparing it with an encompassing two-factor model. The improvements in mean squared error are small, usually less than a few percent. And, in most of the cases we consider, the BIC prefers the one-factor model (superior BIC values are shown in bold).

For comparison, we also report our one-factor variance ratio test, which estimates the model's single parameter from the first two maturities and tests cross-equation restrictions for the longest maturity (24 months). In contrast to the full term structure model comparison approach, our variance ratio test easily detects the internal consistency violation with variance ratios above two in all cases but one (in all cases, the variance ratio is significantly greater than one at the 5% level or better). The reason for the discrepancy between the two approaches is that our test is explicitly designed to detect the type of overreaction found in the data and built into the data generating process (DGP) for these simulations. The information criterion, on the other hand, relies on assessing the one-factor model solely based on its performance relative to the two-factor model. But, in this example, both models are misspecified, so the likelihood of rejecting the null is small. Of course, if the alternative specification matched the DGP, the BIC would always select the alternative over the null. In reality, the exact nature of the misspecification is unknown, and the variance ratio test is well suited to detect overreaction without the need to specify a particular alternative.

C Affine Representation of Structural Models

The affine-$\mathbb{Q}$ representation is typically associated with reduced-form models, as in (Duffie, Pan and Singleton, 2000). However, many workhorse structural asset pricing models also feature affine $\mathbb{Q}$ dynamics. In this section we briefly review the $\mathbb{Q}$-dynamics of prevalent consumption-based models.

We begin with the long run risks models of Bansal and Yaron (2004), in which log consumption growth and its volatility follow linear dynamics. The log pricing kernel is approximately linear (the linearity is exact with unit EIS, and the linear approximation is extremely accurate, as shown in Dew-Becker and Giglio (2013)). In this model the log price and the log price-dividend ratio of all consumption or dividend strips are linear functions of the model's state variables (the persistent consumption growth term x_t and the conditional variance of consumption growth σ_t^2). Prices of consumption and dividend strips therefore follow an exponentially affine specification (with heteroskedasticity).

A related paper, Drechsler and Yaron (2011), extends the model to match the variance risk premium. Dew-Becker et al. (2015) solve for the term structure of variance swaps in that model. In that model the $\mathbb{Q}-$dynamics of variance are linear, and the log pricing kernel is linear (under the standard approximation), and thus variance swaps also follow an affine

structure. Note that in this paper the distribution of the shocks is not normal under $\mathbb{Q}$ (due to the presence of jumps), but this is irrelevant for the term structure of variance swaps because these are linear (not exponential) claims to future variance.

Next, we consider two time-varying rare disaster models, Gabaix (2012) and Wachter (2013). In Gabaix's model, the use of linearity-generating processes (LGP) implies that the (level) price-dividend ratio is linear for all dividend strips. While the LGP assumption buys tractability in modeling price-dividend ratios, the term structure of claims does not follow linear dynamics; the model therefore is not nested in the affine-$\mathbb{Q}$ class. In Wachter (2013), on the other hand, the prices and price-dividend ratios for consumption and dividend strips are loglinear in the disaster probability λ_t, which itself follows a (linear) square-root process. The model therefore follows in the category of exponential affine-Q models with heteroskedasticity, like long run risks.

Finally, the habit formation model of Campbell and Cochrane (1999) does not map directly into the affine specification, as discussed in Wachter (2005).

A number of papers have explored the relationship between learning and excess volatility, such as Timmermann (1993); Barsky and De Long (1993); Veronesi (1999); Pástor and Veronesi (2003, 2009*b*,*a*). In some (but not all) cases, such as in Barsky and De Long (1993), adding learning to the model preserves the affine-$\mathbb{Q}$ structure. In other cases, learning about model parameters induces non-linearities, with which we deal directly in Section 4.4.

D Risk-free Rate Variation, Heteroskedasticity and Other Considerations

In this section we consider in greater detail some additional theoretical consideration that may play a role in our analysis, including the role of interest rate variation, heteroskedasticity, and measurement error.

D.1 Stochastic Risk-free Rates

For many of the asset classes considered in this paper, time variation in the risk-free rate plays a minor role in determining the volatility of prices along the term structure, and is typically ignored in the literature (for example, Ait-Sahalia, Karaman and Mancini (2014) ignore risk-free rate variation when pricing variance swaps).

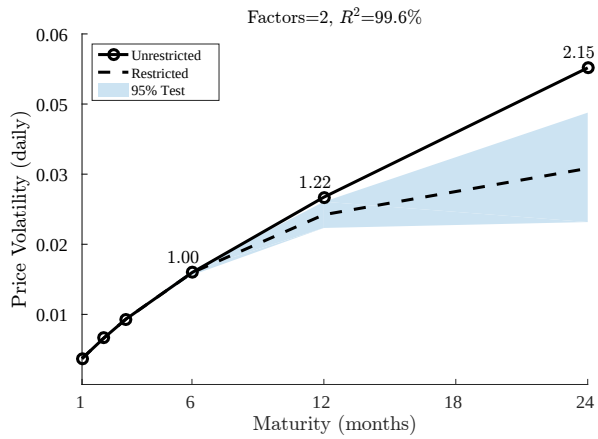
For other asset classes, interest rate variation plays a more important role. Here we show that in exponential-affine models where not only log cash flows x_t but also short-term rates r_t are linear functions of the factors, our test is valid even in the presence of (unmodeled) stochastic interest rates. Consider in particular a cumulative contract that pays all the cash flows at maturity, and has an upfront payment of the price. Then, we can write the price as:

$$p_{t,n} = E_t^{\mathbb{Q}}\left[\frac{e^{x_{t+1}+\ldots+x_{t+n}}}{e^{r_t+\ldots+r_{t+n-1}}}\right] = E_t^{\mathbb{Q}}\left[e^{y_{t+1}+\ldots+y_{t+n}}\right] \tag{26}$$

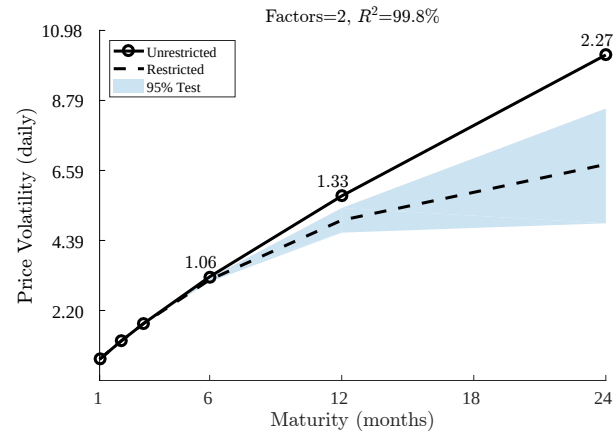
where $y_t = x_t - r_{t-1}$. If y_t is a linear function of the factors (for example because x_t and r_{t-1}

Figure A18: Variance Swap Tests: Linear vs. Exponential-Affine

(a) Affine model



(b) Exponential-affine model



Note. See Figure 1.

are driven by the same factors), we can simply see this price as a claim to risk-free-adjusted cash flows y_t. Finally, remember that none of our analysis requires us to actually *observe* the cash flow (in this case y_t): it is enough to know that the price is determined according to an exponential-affine model in *some* cash flow y_t.

The argument also holds when all payments are exchanged at maturity, since in that case

$$p_{t,n} = E_t^{\mathbb{Q}}\left[\frac{E_t^{\mathbb{Q}}[e^{r_t+\ldots+r_{t+n}}]}{e^{r_t+\ldots+r_{t+n}}} e^{x_{t+1}+\ldots+x_{t+n}}\right]$$

which means that we can construct the price $\tilde{p}_{t,n} = p_{t,n}\delta_{t,n}$, where $\delta_{t,n}$ is the price of a risk-free bond with maturity n, and the adjusted price $\tilde{p}_{t,n}$ will have the same form as (26).

D.2 Linear Versus Exponential Representations

In modeling the market for volatility claims we have followed the literature in writing the payoff as a linear function of underlying factors. We now explore the robustness of our results to a common non-linear functional form. We study an alternative model for volatility claims in which realized variance is assumed to be exponentially linear in the factors:

$$RV_t = \exp(x_t), \quad x_t = \delta_0 + \delta_1' H_t$$

with H_t conditionally normally distributed and homoskedastic (we treat the heteroskedastic case below). In this case, the log price of a forward claim to one period of variance at time $t+n$ is

$$\ln p_{n,t} = E_t^{\mathbb{Q}}[\exp(x_{t+n})] = \mathbf{1}'\rho^n H_t + \text{constant}.$$

We construct psuedo-cumulative claims whose prices are the sum of the log prices of the individual cash flows,

$$\tilde{p}_{t,n} = \sum_{j=1}^{n} \ln p_{t,n}$$

These do not correspond to log prices of tradable cumulative contracts, but instead are a way to aggregate the log forward prices into a form for which our variance ratio tests are applicable.

Figure A18 reports variance ratios for cumulative variance swap prices when realized variance is assumed to be affine in levels as in Figure 1 (left panel) or affine in logs (right panel).[41] The figure shows that there is little difference between variance ratio tests in the two contexts. In both cases, the null model is significantly rejected with variance ratios above 2.0 at 24 months.

D.3 Heteroskedasticity Adjustment in Exponential-affine Models

The exponential-affine model for volatility described in the previous section also allows us to understand the effects of stochastic volatility on the model-predicted factor loadings (remember that stochastic volatility is inconsequential for the test when modeling volatility in a linear framework).

Below we derive the model-predicted factor loadings in the exponential-affine model for volatility when the conditional variance of the factors is assumed to be proportional to the one-period price (the VIX), capturing the intuition that as the VIX increases, future fluctuations in variance will be more pronounced:

$$Var_t(H_{t+1}) = \Gamma_t \Gamma_t' = \Gamma\Gamma'\sigma_t^2$$

with:

$$\sigma_t^2 = a * f_{1,t}$$

for $a > 0$ a proportionality constant. In this model, the loadings of log forward prices $f_{t,n}$ on the factors follow:

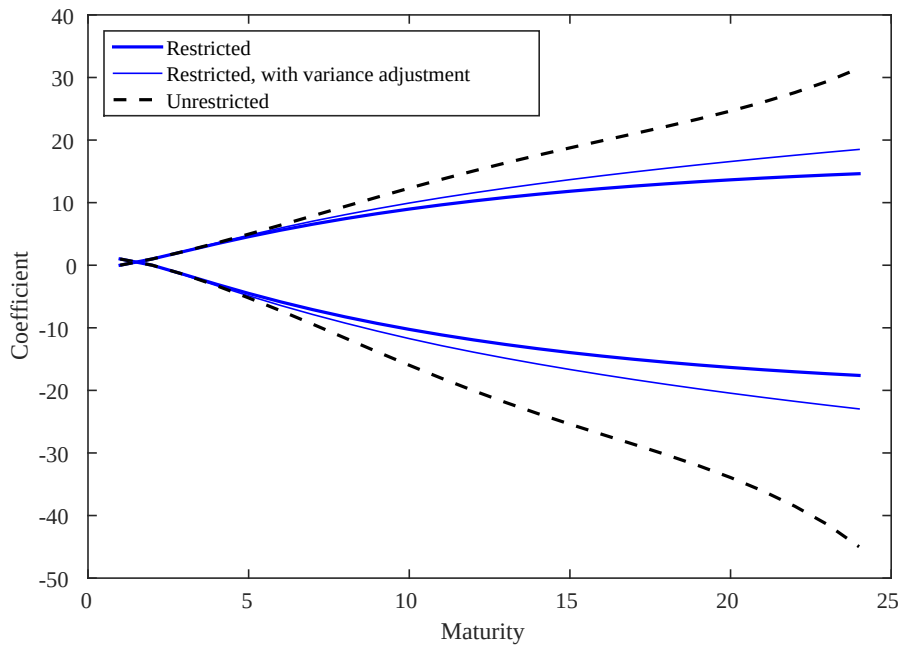
$$b_1 = \mathbf{1}'\rho + \frac{1}{2}a(\mathbf{1}'\Gamma\Gamma'\mathbf{1})\mathbf{1}'\rho$$

$$b_{n+1} = b_n'\rho + \frac{1}{2}a(b_n'\Gamma\Gamma' b_n)\mathbf{1}'\rho$$

Given that empirically $a > 0$ (the volatility of volatility increases when the VIX is high), it immediately follows that the heteroskedasticity adjustment will slow down the decay of factor loadings as the maturity increases. Potentially, this effect can generate higher factor loadings at higher maturities under the null, thus reducing the heteroskedasticity-corrected variance ratios. It is therefore important to quantify the magnitude of this adjustment.

[41] To construct the log prices of variance forwards, we interpolate the variance swap curve using a cubic spline. Our original test on variance swaps does not require any interpolation, though working with forwards does. The results of Figure A18 are robust to different methods of interpolation.

Figure A19: Variance Swap Loadings (Homoskedastic vs Heteroskedastic model)



Note. The figure plots the loadings of prices of each maturity on the two factors (1-month and 2-month price). Dashed lines indicate loadings in the unrestricted model, solid lines indicate loadings in the restricted model. The thick line reports the coefficients under the homoskedasticity assumption, the thin line adjusts for heteroskedasticity.

Below we discuss in detail how the adjustment term at each maturity n, $\frac{1}{2}a(b_n\Sigma\Sigma b_n')\mathbf{1}'\rho$, can be estimated by regressing the conditional *variance* of $f_{t,n}$ on $f_{1,t}$. We can then use these estimated adjustment terms to study how the factor loadings b_n change once we account for heteroskedasticity. Figure A19 reports the loadings of log cumulative variance swap prices, $p_{t,n}$, onto the first two prices, in the null model with and without heteroskedasticity adjustment, as well as the unrestricted loadings. The figure shows that quantitatively the heteroskedasticity adjustment has only a minor effect on the loadings on the two factors.

D.3.1 Derivation and Estimation

We consider here the case of forward claims on a cash flow $\exp\{x_t\}$, where x_t is linear in the factors. Assume that $\mathbb{P}$ dynamics follow:

$$H_{t+1} = c + \rho^{\mathbb{P}} H_t + \Gamma_t \epsilon_{t+1}$$

$$x_t = \delta_0 + \mathbf{1}' H_t$$

The one-period stochastic discount factor follows:

$$M_{t,t+1} = exp(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\epsilon_{t+1})$$

where the vector λ_t captures the time-varying prices of risk of the different shocks.

The term Γ_t captures stochastic conditional volatility of the factors; we specify the exact assumptions about the dependence of Γ_t on time-t information below.

For any forward asset on a cash flow x_t, with maturity $n+1$, we have the recursive equation:

$$f_{t,n+1} = E_t[\exp\{-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\epsilon_{t+1}\}f_{t+1,n}]$$

(where the expectation E_t is under the physical measure).

Now, we conjecture that the forward price is an exponentially-affine function of the factors:

$$f_{t,n+1} = \exp\{a_{n+1} + b_{n+1}H_t\}$$

Taking logs:

$$a_{n+1} + b_{n+1}H_t = lnE_t[\exp\{-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\epsilon_{t+1} + a_n + b_nH_{t+1}\}]$$

$$= lnE_t[\exp\{-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\epsilon_{t+1} + a_n + b_n(c + \rho^{\mathbb{P}}H_t + \Gamma_t\epsilon_{t+1})\}]$$

$$= -r_t - \frac{1}{2}\lambda_t'\lambda_t + a_n + b_n(c + \rho^{\mathbb{P}}H_t) + \frac{1}{2}V_t((-\lambda_t' + b_n\Gamma_t)\epsilon_{t+1})$$

$$= -r_t - \frac{1}{2}\lambda_t'\lambda_t + a_n + b_n(c + \rho^{\mathbb{P}}H_t) + \frac{1}{2}\lambda_t'\lambda_t + \frac{1}{2}b_n\Gamma_t\Gamma_t b_n' - b_n'\Gamma_t\lambda_t$$

$$= -r_t + a_n + b_nc + b_n\rho^{\mathbb{P}}H_t + \frac{1}{2}b_n\Gamma_t\Gamma_t b_n' - b_n'\Gamma_t\lambda_t$$

For the very first maturity (i.e. $f_{t,1}$), we have:

$$a_1 + b_1H_t = lnE_t\exp\{-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\epsilon_{t+1} + x_{t+1}\}\}$$

$$= lnE_t\exp\{-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\epsilon_{t+1} + \delta_0 + \mathbf{1}'[c + \rho^{\mathbb{P}}H_t + \Gamma_t\epsilon_{t+1}]\}\}$$

$$= -r_t + \delta_0 + \mathbf{1}'c + \mathbf{1}'\rho^{\mathbb{P}}H_t + \frac{1}{2}\mathbf{1}'\Gamma_t\Gamma_t'\mathbf{1} - \mathbf{1}'\Gamma_t\lambda_t$$

In both expressions for $n = 1$ and for $n > 1$, we have the terms $\Gamma_t\Gamma_t'$ and $\Gamma\lambda_t$ that are functions of time-t information. To find an exponentially-affine solution, these terms need to be linear in the factors. Following the term structure literature, we assume that $\Gamma_t\Gamma_t'$ is linear in H_t (which makes the term $b_n\Gamma_t\Gamma_t b_n'$ also linear in H_t). In particular, we assume that:

$$V_t(H_{t+1}) = \Gamma_t\Gamma_t' = \Gamma\Gamma'\sigma_t^2$$

with:

$$\sigma_t^2 = a * f_{1,t}$$

for some $a > 0$.

λ_t is assumed to follow:

$$\lambda_t = \Gamma_t^{-1}\Gamma(\lambda + \Lambda H_t)$$

This makes the term $\Sigma_t \lambda_t$ also linear in H_t. In addition, if the risk-free rate is $r_t = a_0 + a_1 H_t$, the term a_1 would also enter the recursion for b_n. In what follows, we ignore risk-free rate variation as it plays a minor role in the pricing of variance swaps.

We can rewrite the expressions under $\mathbb{Q}$, using the same normalizations we have used in our main analysis: $\rho \equiv \rho^{\mathbb{P}} - \Gamma\Lambda$ (the VAR companion matrix under $\mathbb{Q}$) is diagonal, and $c^{\mathbb{Q}} \equiv c - \lambda\Gamma = 0$. We can then rewrite the expressions as:

$$a_{n+1} + b_{n+1}H_t = a_n + b_n c + b_n \rho^{\mathbb{P}} H_t + \frac{1}{2} b_n \Gamma\Gamma \sigma_t^2 b_n' - b_n'\Gamma(\lambda + \Lambda H_t)$$

$$a_1 + b_1 H_t = \delta_0 + \mathbf{1}'c + \mathbf{1}'\rho^{\mathbb{P}} H_t + \frac{1}{2}\mathbf{1}'\Gamma\Gamma'\sigma_t^2\mathbf{1} - \mathbf{1}'\Gamma(\lambda + \Lambda H_t)$$

or:

$$a_{n+1} + b_{n+1}H_t = a_n + b_n \rho H_t + \frac{1}{2} b_n \Gamma\Gamma b_n' \sigma_t^2$$

$$a_1 + b_1 H_t = \delta_0 + \mathbf{1}'\rho H_t + \frac{1}{2}\mathbf{1}'\Gamma\Gamma'\mathbf{1}\sigma_t^2$$

Now, recall that $\sigma_t^2 = a * f_{1,t} = a\mathbf{1}'\rho H_t$. The expressions then become:

$$a_{n+1} + b_{n+1}H_t = a_n + b_n \rho H_t + \frac{1}{2} b_n \Gamma\Gamma b_n' (a\mathbf{1}'\rho H_t)$$

$$a_1 + b_1 H_t = \delta_0 + \mathbf{1}'\rho H_t + \frac{1}{2}\mathbf{1}'\Gamma\Gamma'\mathbf{1}(a\mathbf{1}'\rho H_t)$$

We can now match coefficients on H_t, and obtain:

$$b_1 = \mathbf{1}\rho + \frac{1}{2}\mathbf{1}'\Gamma\Gamma'\mathbf{1}(a\mathbf{1}'\rho)$$

$$b_{n+1} = b_n \rho + \frac{1}{2} b_n' \Gamma\Gamma b_n (a\mathbf{1}'\rho)$$

To learn about the magnitude of the coefficient adjustments $\frac{1}{2}\mathbf{1}'\Gamma\Gamma'\mathbf{1}(a\mathbf{1}'\rho)$ and $\frac{1}{2}b_n'\Gamma\Gamma b_n(a\mathbf{1}'\rho)$, we proceed as follows. First, note that the conditional variance of the log cash flow in the model is (up to a constant):

$$V_t(x_{t+1}) = \mathbf{1}'\Gamma\Gamma'\mathbf{1}(a\mathbf{1}'\rho H_t) = \mathbf{1}'\Gamma\Gamma'\mathbf{1} a f_{t,1}$$

Therefore, regressing $V_t(x_{t+1})$ onto $f_{t,1}$ would yield an estimate of the term $\mathbf{1}'\Gamma\Gamma'\mathbf{1}a$. This would allow us to estimate the heteroskedasticity adjustment for b_1. Next, consider the

conditional variance of the first log price (from the left-hand side of the equations above):

$$V_t(f_{t+1,1}) = b_1'\Gamma_t\Gamma_t b_1 = b_1'\Gamma\Gamma b_1 a f_{t,1}$$

The regression coefficient of $V_t(f_{t+1,1})$ onto $f_{t,1}$ yields an estimate of $b_1'\Gamma\Gamma b_1 a$, which we can use to adjust the coefficient b_2 for the effects of conditional volatility. Continuing the recursion, this allows us to compute the adjustment for all maturities.

Two final notes on the implementation. First, the most natural way to implement the conditional variance regression is to regress the monthly realized volatility of each variable (x_t, $f_{t,1}$, $f_{t,2}$, and so on, computed as the sum of changes in log prices during the month). While we don't observe high-frequency data on realized volatility x_t within a month, we can use the realized volatility of $f_{t,1}$ as a proxy. Second, log realized volatilities for maturities above 12 are very noisy, due to the interpolation-induced errors. We therefore apply the regression coefficients estimated for maturity 12 to all higher maturities. This procedure is conservative because the coefficients of this regression appear to be strongly decreasing with maturity (so after maturity 12 they should be even lower than those observed at maturity 12); in addition, the theory predicts that they should be decreasing as maturity increases, since the overall volatility of forwards should converge to zero as maturities increase.

As a robustness test, we also compute the volatility adjustments by using the the month-to-month squared change in price as left-hand side variable as opposed to the within-month realized volatility. The results are essentially identical.

D.4 Measurement Error

In the theoretical setting of Section 2, the prices derived in Equations (6) and (9) show that the value of a claim at any maturity is representable as an exact, error-free linear function of prices of claims at other points on the term structure.

As Piazzesi (2010) notes, observed prices may not perfectly represent the theoretical expectations of investors, but instead may also include "measurement errors" that arise from data entry errors, building price series from multiple (and potentially asynchronous) data sources, vendors that interpolate data to fill in missing prices, etc. In the context of US treasury yields, Cochrane and Piazzesi (2005) find evidence indicating that the data indeed contain patterns that are a signature of measurement error.

Measurement error potentially influences our parameter estimates and test statistics. While in our setup measurement error at all maturities above K is explicitly incorporated (and thus the standard errors of our test correctly account for it), we assume that maturities $1, ..., K$ are observed without error. This is not an unreasonable assumption because for many term structures we consider, short maturity claims are indeed the most liquid.

If, however, maturities $1, ..., K$ are subject to measurement error, the regression that extracts the matrix ρ will suffer from attenuation bias, and this affects our variance ratio test. We address this in two ways. First, we show that for reasonable values for the magnitude of measurement error, the distortion relative to the model without measurement error is minimal. Second, we use errors-in-variables methods to conduct tests that are robust to measurement error. The resulting tests produce nearly identical findings to those in Section

3, indicating that measurement error is not responsible for the long maturity excess volatility that we document.

More formally, we use a tilde to represent error-ridden observable prices[42]

$$\tilde{f}_{t,j} = f_{t,j} + v_{t,j}, \quad j = 1, ..., N.$$

The variance ratio test in Equation (13) depends on coefficient estimates in restricted and unrestricted projections of the error-free long maturity price $f_{t,K+j}$ onto error-free short-maturity prices $F_{t,1:K}$.[43] When the error-free projections are infeasible due to noise in prices, consistency of the variance ratio test faces two obstacles. First, we require a consistent estimate of the unrestricted coefficient based on noisy data $\tilde{f}_{t,K+j}$ and $\tilde{F}_{1:K}$ (in analogy with Equation (11)) in order to calculate the numerator of the variance ratio. Second, we need a consistent estimate of the restricted coefficient to construct the denominator of the variance ratio (which relies on b in Equation (7)) but that is based on the noisy short-end prices $\tilde{f}_{t,K+1}$ and $\tilde{F}_{1:K}$.

To understand the magnitude of the measurement error problem, we look at term structures for which we have information about the bid-ask spread—for variance swaps and Treasuries. A reasonable calibration for the standard deviation of the measurement error is the bid-ask spread itself (0.3 annualized volatility points for short maturity variance swaps, and 1bp for short maturity Treasury bond yields). First, we add iid measurement error of this magnitude to the actual data, and recompute the variance ratios (top row of Figure A20). For both variance swaps and Treasuries, the addition of measurement error has tiny effects on the observed variance ratios.

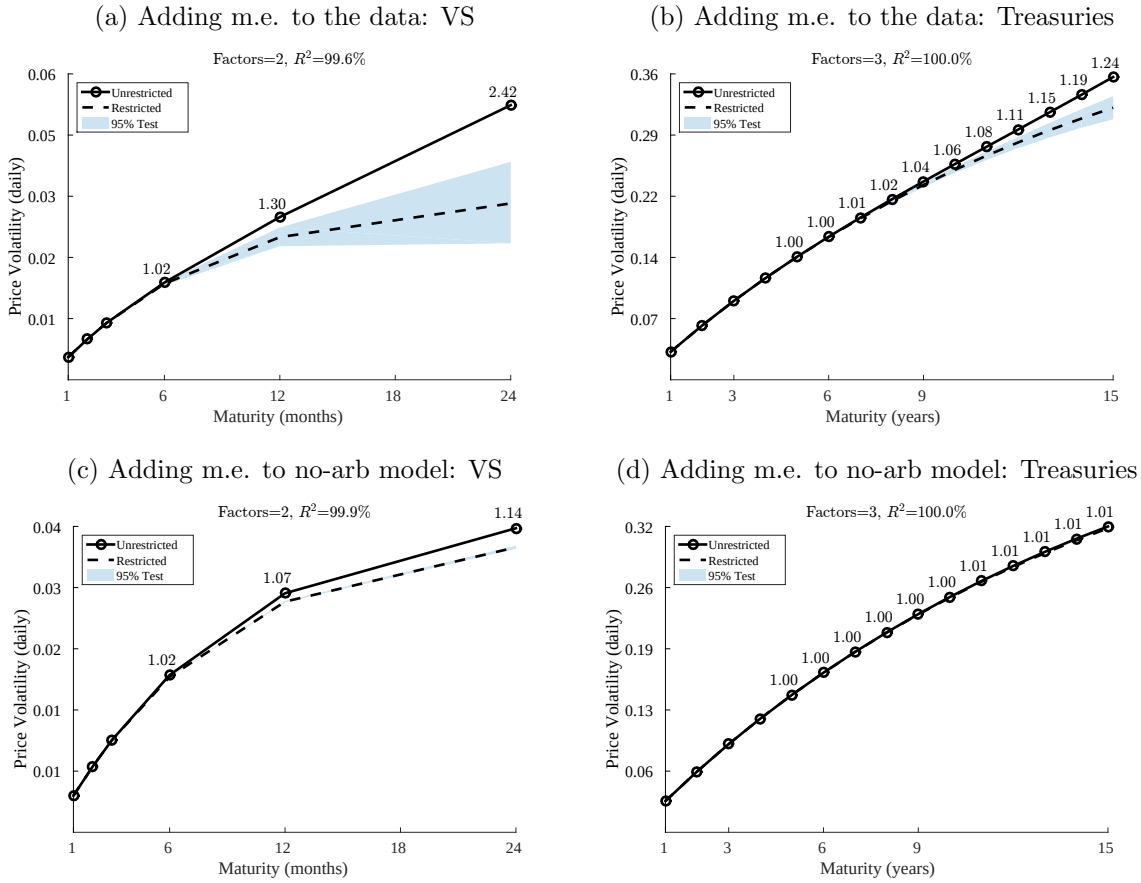
Second, we use the data to construct term structure prices that exactly satisfy the affine model form. To do this, we estimate the model from the short end of each curve and construct the new dataset using the *fitted* prices from the model. In this artificial dataset, prices at all maturities are fully consistent with the short end, and variance ratios are one at all maturities. Next, we add iid measurement error to this artificial dataset, re-estimate the model, and calculate variance ratios. Now, the extent to which variance ratios differ from one is due *entirely* to measurement error. The bottom row shows the results for variance swaps and Treasuries, again confirming the small effect of measurement error on our variance ratio test.

We can use the same procedure to calibrate how large measurement error is needed to generate the variance ratios we see in the data. For variance swaps, we need measurement error with standard deviation of more than 2 volatility points, i.e. 7 times the bid-ask spread at the short end of the curve. For Treasuries, we need measurement error with standard deviation of at least 10bp, 10 times the average bid-ask spread of short-maturity bonds. These results show that in both markets, we need unrealistic measurement error to

[42] Piazzesi (2010), in Section 6, gives an excellent overview of model specification choices when affine term structures are subject to measurement error. If we assume, as is often the case in affine models, that the first K prices are perfectly observed and only maturities $K+1$ to N are subject to errors, then our baseline estimator in Section 2 remains consistent, and p-values of our test retain appropriate size due to our bootstrap standard errors.

[43] The specific form of the weighting matrix, which take the value $\hat{\Sigma}_{1:K}$ Equation (13), appears in both the numerator and denominator and is thus not crucial for consistency of the test.

Figure A20: Calibration of Measurement error



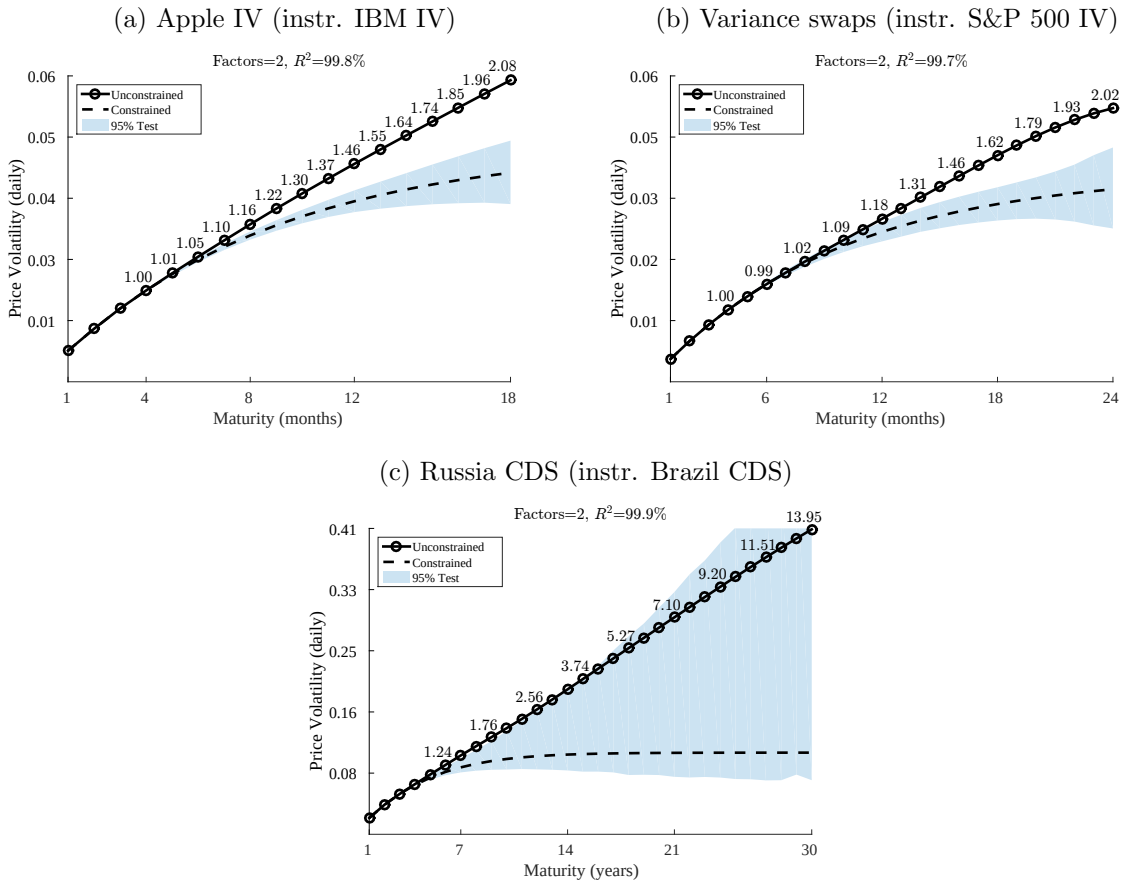
Note. See Figure 1.

produce variance ratios as high as we document.

Next, we describe an instrumental-variable correction for measurement error. Instrumental variables (IV) methods are a common means of consistently estimating a regression coefficient when the independent variable is observed with error. For example, suppose the affine model has a single factor and the errors are uncorrelated across maturities. In this case, b is consistently estimated by an IV regression of $\tilde{f}_{t,2}$ on $\tilde{f}_{t,1}$, using any other price $\tilde{f}_{t,j}$ at maturity $j > 2$. If the errors are uncorrelated across maturities, $\tilde{f}_{t,j}$ is a valid instrument for the noisy dependent variable $\tilde{f}_{t,1}$. By the same rationale, the unrestricted long-end projection coefficient can be consistently estimated as well. Given consistent estimates of b and the long maturity unrestricted coefficient, the variance ratio test will be consistent. The only qualitative difference versus Equation (13) is that the weighting matrix will be replaced with an estimate of short maturity noisy price variance, $Var(\tilde{F}_{t,1:K})$.

In practice, however, it is quite likely that measurement errors are correlated across maturities, so the strategy of instrumenting with other maturities in the *same* term structure fails to satisfy the exclusion restriction. It is less likely that measurement errors would be correlated across *different* term structures. We therefore use prices from different but related

Figure A21: Instrumental Variables Adjustment for Measurement Error



Note. See Figure 1.

term structures as instruments to help resolve potential inference problems due to errors-in-variables bias.

As a first example, we revisit the two-factor affine model for the term structure of Apple's variance claims studied in Figure 4. If prices are measured with error, then we must instrument the regressions of $\tilde{f}_{t,j}^{\text{Apple}}$ on $\tilde{F}_{t,1:2}^{\text{Apple}}$ (for $j = 3, ..., 24$). As instruments, we use short-end prices of claims to IBM variance, $\tilde{F}_{t,1:2}^{\text{IBM}}$. This approach is valid under the conditions that true, error-free short prices $F_{t,1:2}^{\text{Apple}}$ and $F_{t,1:2}^{\text{IBM}}$ are correlated between the different term structures but the errors $v_{t,1:2}^{\text{Apple}}$ and $v_{t,1:2}^{\text{IBM}}) = 0$ are not. Indeed, the volatility of individual stocks tend to exhibit strong cross correlation,[44] but there is no obvious reason to suspect that errors in the measurement of their prices are correlated.

The variance ratio test results for this example are plotted in the left panel of Figure A21. Test statistics based on the IV adjustment are nearly identical to those in the baseline estimation. The same is true if we instrument the variance swap term structure tests using implied volatilities of the S&P 500 options (second panel), and if we instrument the Rus-

[44]See Kelly, Lustig and Van Nieuwerburgh (2013) and Herskovic et al. (2014).

sian CDS term structure with Brazilian CDS spreads (third panel). In all cases, values of instrumented test statistics are quantitatively similar to those in Section 3, suggesting that our main findings cannot be explained by measurement error.[45]

E Data Details and Asset-specific Modeling Considerations

In this section we show how each asset class considered maps into our linear or loglinear framework.

E.1 Variance Swaps and Related Variance Derivatives

As discussed in the text, the price of a variance swap follows:[46]

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

We then model RV_t as a linear function of the factors, which immediately yields:

$$p_{t,n} = a_n + b_n' H_t \tag{27}$$

An attractive feature of the simple payoff structure of variance swaps is that dependence of prices on factors, $b_n' H_t$, is robust to many modifications of the factor model. For example, because the swap price is the expected value of the level of RV_t, having both prices and payoffs linear in the factors no longer requires Gaussianity. Any shock distribution with constant means implies the pricing structure in (27).

One important consideration to keep in mind is that because variances are non-negative, a homoskedastic linear Gaussian model is an imperfect description of RV_t. Stochastic variance is a standard feature in the bond and option pricing literatures, and a number of solutions exist that ensure positive variances. The most common solution is to use a CIR volatility process. In these models, the model innovations remain standard normal, but are multiplied by a volatility that scales with the factors (and hence with the level of volatility). The modified model takes the general form[47]

$$H_t = \rho H_{t-1} + \Sigma_{t-1} u_t$$

[45] In the case of Russian CDS, the standard errors of the instrumented statistics are much larger than our estimates based on OLS, which is likely due to the fact that Russian and Brazilian CDS spreads share a much lower correlation than, for example, Apple and IBM implied volatilities.

[46] We ignore risk-free rate variation, since its volatility and correlation with the variance swap payoff are small, following Ait-Sahalia, Karaman and Mancini (2014), Egloff, Leippold and Wu (2010), Dew-Becker et al. (2015).

[47] For infinitesimal time intervals, the variance may be constructed to maintain strictly positive variance while retaining the Gaussianity of factor innovations, u_t. In discrete time, this heteroskedastic Gaussian process does not perfectly rule out negative variances, but may be constructed to do so with probability arbitrarily close to one.

where Σ_{t-1} is a constant function of H_{t-1}. When the model is specified at a high enough frequency (going to continuous time in the limit), and assuming appropriate Feller conditions for the model parameters (see Dai and Singleton (2002)), the probability of variance going below zero tends to zero.

Note that this stochastic volatility case only affects the scale of the innovation u_t. Therefore, the expected *level* payoff in is unaffected, hence equation (27) is also unaffected. Different versions of this model are applied by Ait-Sahalia, Karaman and Mancini (2014), Egloff, Leippold and Wu (2010), Dew-Becker et al. (2015).

As discussed in the text, in some of our tests we take ATM implied variance as a proxy for the risk-neutral expected variance. This is motivated by the theoretical result of Carr and Lee (2009) who show that to a first-order approximation, ATM implied volatility corresponds to the price of a volatility swap (a claim to realized volatility). Perhaps more importantly, our use of ATM is also motivated by practical considerations. ATM volatility is more widely available, especially for long dated options, because it only requires one ATM option price to construct. The synthetic variance swap price, VIX^2, can be calculated for all of our option term structures but is less stable than ATM implied volatility due to its reliance on OTM option prices, of which fewer are available at long maturities.

Our variance swap data comes from two industry sources, both described in Dew-Becker et al. (2015). Our implied variance series are obtained from Optionmetrics (equity derivatives) and JP Morgan (currency IV).

Our analysis of the term structure of ATM implied variance uses the same model as for variance swaps, but sets $p_{t,n} = IV_t^n$, where IV_t^n is the n-maturity option-implied variance.

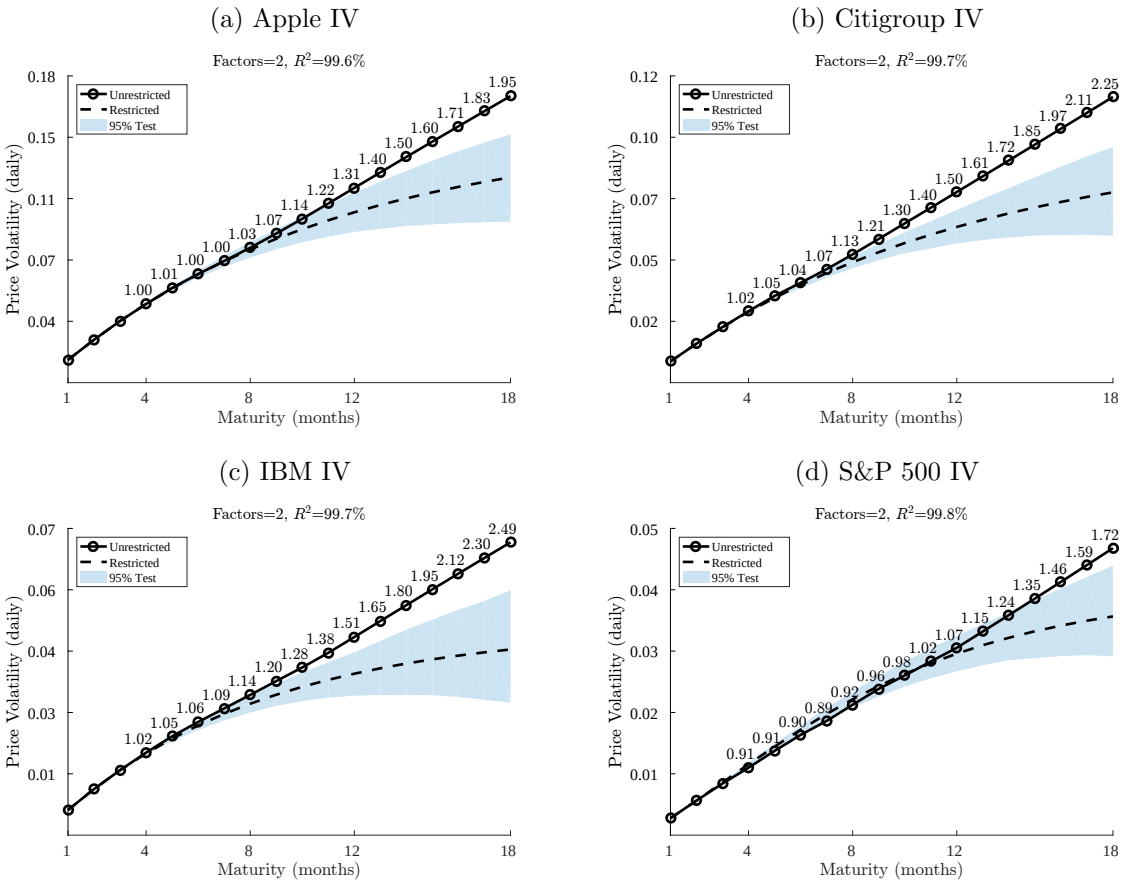
As a robustness check, we also construct the term structure of the VIX using option prices, following the SVI fitting procedure described in Dew-Becker et al. (2015). Note that we need to both interpolate and extrapolate the implied volatility curve (using the SVI model), and the relative scarcity of out-of-the-money options at long maturities can result in noisy VIX estimates. Also, for some of our options sample, there are not enough OTM options available to estimate the VIX at maturities above one year. We report the results using sample dates where the entire term structure up to 18 months is observed (for all contracts, we have between 1,000 and 2,000 days that can be used for estimation). Figure A22 shows that the variance ratios for the term structure of the VIX behave very similarly to the ones constructed for implied volatilities (Figure 4), though with larger confidence intervals.

E.2 Treasuries

Our development of the exponential-affine model for interest rates follows Hamilton and Wu (2012), who study the class of Gaussian affine term structure models developed by Vasicek (1977), Duffie, Kan et al. (1996), Dai and Singleton (2002), and Duffee (2002), and studied by many others.

In the Gaussian affine term structure model, bonds are claims on short-term interest rates. One-period log risk-free rate x_t is a linear function of the factors with factor dynamics under the pricing measure described by a VAR, just as in our main set-up. The price of a

Figure A22: VIX term structure



Note. See Figure 1.

risk-free bond that pays \$1 after n periods is

$$P_{t,n} = E^{\mathbb{Q}}\left[\exp\left(-\sum_{j=1}^{n} x_{t+j}\right)\right]. \tag{28}$$

We assume that factor shocks are homoskedastic $\Sigma_t = \Sigma$ following Hamilton and Wu (2012), which implies that the log bond price is

$$p_{t,n} \equiv \log P_{t,n} = a_n + b_n H_t.$$

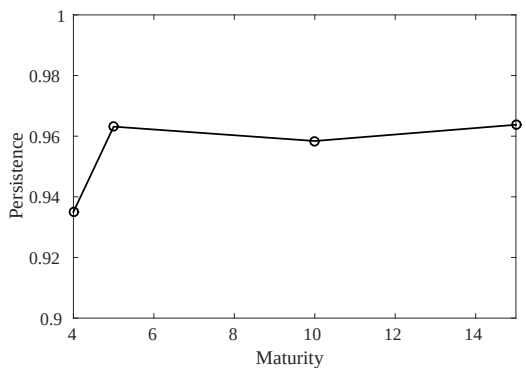
The factor loading depends only on the persistence of the factors:

$$b_n = \mathbf{1}'(I + \rho + ... + (\rho)^{n-1}). \tag{29}$$

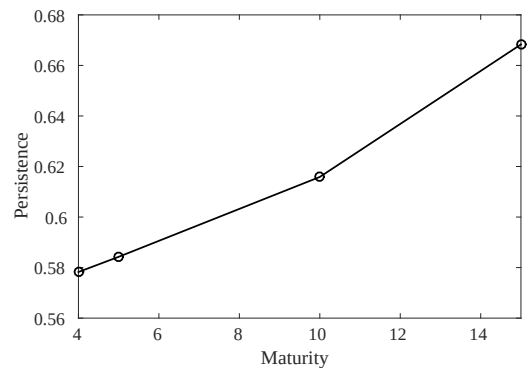
The intercept is an inconsequential constant function of remaining model parameters, and drops out from all variance calculations.

Figure A23: $\mathbb{Q}$-persistence Estimated Along the Term Structure of Treasuries

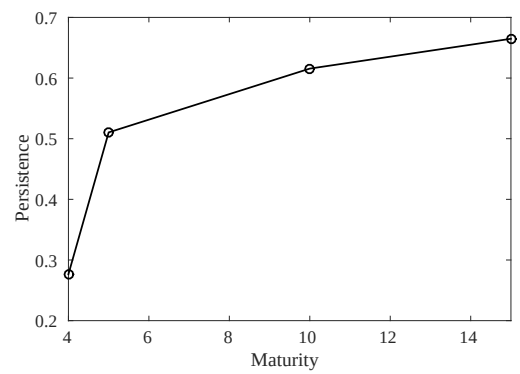
(a) First factor



(b) Second factor



(c) Third factor



Note. See Figure 3.

We conclude by reporting the estimates of ρ obtained by regressions of prices for contiguous maturities at different points in the term structure. The figure shows that for U.S. treasuries,the estimated persistence is stable along the entire term structure.

E.3 Credit Default Swaps

To model CDS spreads, we apply the reduced-form modeling of Duffie and Singleton (1999), in which the price of a defaultable bond is written in terms of a default intensity process λ_t and a process of loss given default L_t. The precise relationship between the price of the bond at time t, P_t, and the processes for λ_t and L_t does not directly map into our general framework of Section 2.

However, Duffie and Singleton (1999) show that under the assumption of fractional recovery of market value in case of default, the price of a defaultable zero-coupon bond can be written as:

$$P_{t,n} = E_t^{\mathbb{Q}}\left[exp(-\int_t^n R_s ds)\right]$$

with

$$R_s = r_s + \lambda_s L_s$$

where λ_t is the default intensity and L_t the loss given default. The defaultable bond can be modeled as a default-free bond with a default-adjusted interest rate. We assume that: 1. r_s and $\lambda_s L_s$ are linear in the factors; 2. underlying factors are homoskedastic; and 3. coupons on the underlying defaultable bonds are small enough (relative to the default-adjusted interest rate) so that the yield of an n-maturity defaultable bond with coupon is close to an n-maturity zero-coupon defaultable bond. We can then write:

$$p_{t,n} = log(P_{t,n}) = -ny_t^n = (a_r^n + a_{\lambda L}^n) + (b_r^n + b_{\lambda L}^n)H_t$$

while for the default-free bond (with log yield y_F) we have:

$$-ny_{F,t}^n = a_r^n + b_r^n H_t$$

To link the bond price to the observed CDS spread, we start from the approximate bond-CDS basis relation, that states

$$Z_t^n \simeq Y_t^n - Y_{F,t}^n$$

i.e. the CDS spread Z_t^n with maturity n is approximately equal to the yield of the bond Y_t^n of that maturity in excess of the corresponding risk-free rate Y_F^n with the same maturity.

Given that both Y_t^n and $Y_{F,t}^n$ are close to zero, we can write the yield spread to a first-order approximation as:

$$Y_t^n - Y_{F,t}^n \simeq log(1 + Y_t^n) - log(1 + Y_{F,t}^n) = y_t^n - y_{F,t}^n$$

so that:

$$nZ_t^n \simeq n(y_t^n - y_{F,t}^n) = -a_{\lambda L}^n - b_{\lambda L}^n H_t$$

This representation links the observed CDS spread allows us to focus on the cross-section of CDS spreads stripped of the risk-free rate dynamics, which will highlight the factor structure in default risk.

E.4 Inflation Swaps

Inflation swaps are claims to future inflation where the the buyer commits to pay a predetermined amount $(1 + p_{t,n})^n - 1$ and receives $[I(t+n)/I(t)] - 1$, where $I(t)$ is the price level index. Risk-neutral pricing implies:

$$(1 + p_{t,n})^n - 1 = E_t^{\mathbb{Q}} \left[\frac{I(t+n)}{I(t)} - 1 \right]$$

Calling $\pi_t = \Delta lnI(t)$, and moving to continuous time, we can write:

$$P_{t,n} = e^{p_{t,n}n} = E_t^{\mathbb{Q}}\left[exp(\int_t^{t+n} \pi_s ds)\right]$$

Just as in the case of bonds, we will have that log cumulative prices $n \cdot p_{t,n}$ will be linear in the factors:

$$n \cdot p_{t,n} = a_n + b_n H_t$$

E.5 Commodity Futures

Call $F_{t,n}$ the price of a future with maturity n. As in Duffie (2010) and Casassus and Collin-Dufresne (2005), if S_t is the value of the underlying at time t, we have:

$$F_{t,n} = E_t^Q[S_{t+n}]$$

Now, if $X_t = log(S_t)$, then we have:

$$F_{t,n} = E_t^Q[\exp\{X_{t+n}\}]$$

We can rewrite X_t as:

$$X_{t+n} = X_t + \sum_{s=1}^{n} x_{t+s}$$

with

$$x_t = \Delta X_t$$

We may model these growth rates as functions of latent factors, so that:

$$x_t = \delta_1' H_t$$

$$F_{t,n} = E_t^Q[\exp\{X_t + \sum_{s=1}^{n} x_{t+s}\}]$$

We can therefore rewrite:

$$\frac{F_{t,n}}{S_t} = E_t^Q[\exp\{\sum_{s=1}^{n} x_{t+s}\}]$$

which therefore has the standard affine form. Note also that we can rewrite the expression for the futures without reference to the underlying, rescaling each future by the price of the first-maturity future:

$$F_{t,1} = S_t E_t^Q[\exp\{x_{t+1}\}]$$

so that:

$$\frac{F_{t,n}}{F_{t,1}} = \frac{E_t^Q[\exp\{\sum_{s=1}^{n} x_{t+s}\}]}{E_t^Q[\exp\{x_{t+1}\}]} \simeq E_t^Q[\exp\{\sum_{s=2}^{n} x_{t+s}\}]$$

This expression maps directly into our exponential-affine framework.

F Missing Factors: Empirical Evidence

Table A7 reports robustness checks varying the number of factors, K. For each term structure, the middle number of factors is the number used in our baseline analysis. We compare these results to tests that include one additional or one less factor. For each choice of K we report the term structure panel R^2 along with variance ratios and their bootstrap p-values at various maturities.

Table A7: Robustness to Varying the Number of Factors

K	Mat.	VR	p-value	R^2	K	Mat.	VR	p-value	R^2
Variance Swaps					Apple IV				
2	24	4.30	0.00	0.988	1	3	1.10	0.00	0.967
2	48	9.83	0.00	0.988	1	6	1.55	0.00	0.967
2	60	11.68	0.00	0.988	1	12	3.17	0.00	0.967
2	120	11.24	0.04	0.988	1	18	5.44	0.00	0.967
3	24	3.67	0.02	0.999	2	3	1.00	0.33	0.994
3	48	8.39	0.01	0.999	2	6	1.09	0.00	0.994
3	60	9.99	0.02	0.999	2	12	1.60	0.00	0.994
3	120	10.03	0.09	0.999	2	18	2.37	0.00	0.994
4	24	1.24	0.76	0.999	3	3	1.00	1.00	0.999
4	48	1.97	0.60	0.999	3	6	1.16	0.00	0.999
4	60	2.26	0.55	0.999	3	12	1.91	0.00	0.999
4	120	2.88	0.49	0.999	3	18	3.04	0.00	0.999
Citigroup IV					IBM IV				
1	3	0.95	1.00	0.982	1	3	1.15	0.00	0.973
1	6	0.88	1.00	0.982	1	6	1.87	0.00	0.973
1	12	1.33	0.00	0.982	1	12	4.52	0.00	0.973
1	18	2.18	0.00	0.982	1	18	8.05	0.00	0.973
2	3	1.00	0.62	0.998	2	3	1.00	0.94	0.997
2	6	1.08	0.00	0.998	2	6	1.08	0.00	0.997
2	12	1.98	0.00	0.998	2	12	1.60	0.00	0.997
2	18	3.51	0.00	0.998	2	18	2.34	0.00	0.997
3	3	1.00	0.00	0.999	3	3	1.00	0.00	0.999
3	6	1.02	0.00	0.999	3	6	1.05	0.00	0.999
3	12	1.61	0.00	0.999	3	12	1.48	0.00	0.999
3	18	2.64	0.00	0.999	3	18	2.10	0.00	0.999
S&P 500					NASDAQ IV				
1	3	1.21	0.00	0.961	1	3	1.10	0.00	0.986
1	6	2.05	0.00	0.961	1	6	1.48	0.00	0.986
1	12	5.07	0.00	0.961	1	12	2.70	0.00	0.986
1	18	9.03	0.00	0.961	1	18	4.46	0.00	0.986

2	3	1.00	0.18	0.995
2	6	0.97	0.95	0.995
2	12	1.20	0.00	0.995
2	18	1.60	0.00	0.995
3	3	1.00	1.00	0.998
3	6	0.98	0.99	0.998
3	12	1.23	0.03	0.998
3	18	1.68	0.01	0.998

2	3	1.00	0.96	0.998
2	6	0.98	0.97	0.998
2	12	1.01	0.41	0.998
2	18	1.10	0.04	0.998
3	3	1.00	0.00	1.000
3	6	1.03	0.00	1.000
3	12	1.22	0.00	1.000
3	18	1.51	0.00	1.000

Stoxx 50 IV

1	12	1.24	0.00	0.964
1	18	1.65	0.00	0.964
1	24	1.99	0.00	0.964
1	30	2.67	0.00	0.964
2	12	1.01	0.39	0.998
2	18	1.22	0.00	0.998
2	24	1.42	0.00	0.998
2	30	1.84	0.00	0.998
3	12	0.92	0.97	0.999
3	18	1.07	0.09	0.999
3	24	1.26	0.00	0.999
3	30	1.68	0.00	0.999

FTSE 100 IV

1	12	1.06	0.09	0.970
1	18	1.23	0.00	0.970
1	24	1.47	0.00	0.970
1	30	1.87	0.00	0.970
2	12	1.58	0.00	0.998
2	18	2.40	0.00	0.998
2	24	3.42	0.00	0.998
2	30	4.81	0.00	0.998
3	12	1.39	0.00	0.999
3	18	1.60	0.00	0.999
3	24	1.89	0.00	0.999
3	30	2.41	0.00	0.999

DAX IV

1	12	1.22	0.00	0.976
1	18	1.64	0.00	0.976
1	24	2.08	0.00	0.976
1	30	2.79	0.00	0.976
2	12	1.05	0.10	0.999
2	18	1.30	0.00	0.999
2	24	1.57	0.00	0.999
2	30	2.03	0.00	0.999
3	12	1.04	0.12	0.999
3	18	1.35	0.00	0.999
3	24	1.73	0.00	0.999
3	30	2.36	0.00	0.999

GBP/USD IV

1	3	1.04	0.08	0.970
1	6	1.35	0.00	0.970
1	12	2.59	0.00	0.970
1	24	8.04	0.00	0.970
2	3	1.00	0.62	0.998
2	6	1.10	0.00	0.998
2	12	1.54	0.00	0.998
2	24	3.34	0.00	0.998
3	3	1.00	1.00	1.000
3	6	1.01	0.00	1.000
3	12	1.11	0.00	1.000
3	24	1.55	0.00	1.000

GBP/JPY IV

1	3	1.07	0.02	0.952
1	6	1.70	0.00	0.952
1	12	4.14	0.00	0.952
1	24	13.02	0.00	0.952
2	3	1.00	0.81	0.997
2	6	1.11	0.00	0.997

USD/CHF IV

1	3	1.01	0.30	0.956
1	6	1.19	0.01	0.956
1	12	1.91	0.00	0.956
1	24	4.65	0.00	0.956
2	3	1.00	0.66	0.996
2	6	1.12	0.00	0.996

2	12	1.49	0.00	0.997	2	12	1.68	0.00	0.996
2	24	2.83	0.00	0.997	2	24	3.68	0.00	0.996
3	3	1.00	0.00	1.000	3	3	1.00	1.00	0.999
3	6	1.02	0.00	1.000	3	6	1.01	0.00	0.999
3	12	1.10	0.00	1.000	3	12	1.18	0.00	0.999
3	24	1.42	0.00	1.000	3	24	1.82	0.00	0.999

Treasuries					BofA CDS				
1	5	0.94	0.75	0.985	1	5	1.09	0.25	0.948
1	7	0.92	0.69	0.985	1	7	1.21	0.18	0.948
1	10	0.93	0.60	0.985	1	10	1.51	0.09	0.948
1	15	1.05	0.45	0.985	1	15	2.70	0.01	0.948
2	5	1.02	0.00	0.999	2	5	1.06	0.00	0.995
2	7	1.08	0.00	0.999	2	7	1.11	0.00	0.995
2	10	1.23	0.00	0.999	2	10	1.32	0.00	0.995
2	15	1.66	0.00	0.999	2	15	2.12	0.00	0.995
3	5	1.00	0.09	1.000	3	5	0.99	0.91	0.999
3	7	1.01	0.02	1.000	3	7	0.96	0.98	0.999
3	10	1.05	0.01	1.000	3	10	0.97	0.73	0.999
3	15	1.22	0.00	1.000	3	15	1.19	0.06	0.999

Morgan Stanley CDS					JPM CDS				
1	5	1.70	0.00	0.960	2	5	0.99	0.57	0.985
1	7	2.32	0.00	0.960	2	7	0.83	1.00	0.985
1	10	3.51	0.00	0.960	2	10	0.82	0.93	0.985
1	15	7.12	0.00	0.960	2	15	1.06	0.38	0.985
2	5	1.10	0.00	0.995	3	5	0.95	0.98	0.997
2	7	1.23	0.00	0.995	3	7	0.74	1.00	0.997
2	10	1.56	0.00	0.995	3	10	0.67	1.00	0.997
2	15	2.76	0.00	0.995	3	15	0.79	0.67	0.997
3	5	0.99	0.93	0.999	4	5	0.97	0.96	0.999
3	7	0.99	0.83	0.999	4	7	0.82	0.99	0.999
3	10	1.07	0.06	0.999	4	10	0.85	0.58	0.999
3	15	1.55	0.00	0.999	4	15	1.27	0.19	0.999

Italy CDS					Russia CDS				
1	10	0.79	0.94	0.972	1	10	1.77	0.00	0.981
1	15	0.75	0.97	0.972	1	15	3.22	0.00	0.981
1	20	0.73	0.98	0.972	1	20	5.49	0.00	0.981
1	30	0.71	0.98	0.972	1	30	11.77	0.00	0.981
2	10	1.20	0.00	0.998	2	10	2.03	0.00	0.998
2	15	1.90	0.00	0.998	2	15	3.88	0.00	0.998
2	20	2.99	0.00	0.998	2	20	6.79	0.00	0.998
2	30	6.21	0.00	0.998	2	30	14.78	0.00	0.998

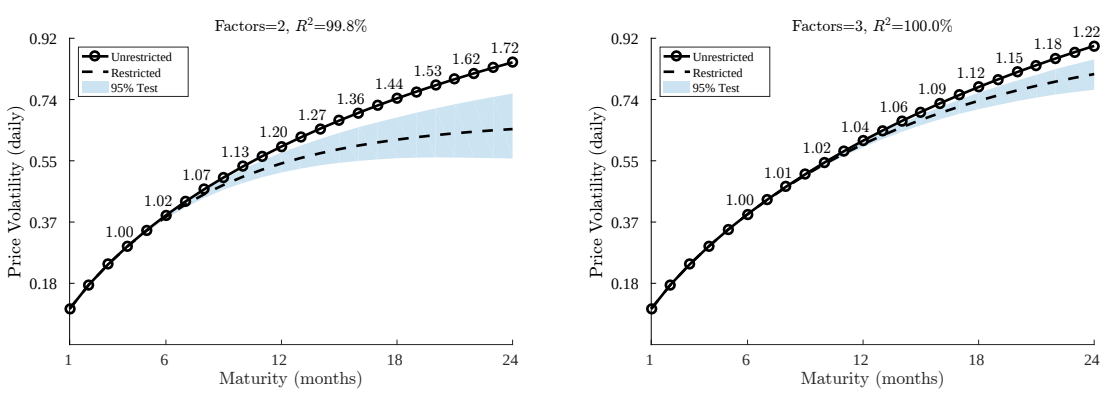
3	10	1.64	0.01	1.000	3	10	1.78	0.00	1.000
3	15	2.80	0.01	1.000	3	15	3.13	0.00	1.000
3	20	4.48	0.01	1.000	3	20	5.25	0.00	1.000
3	30	9.34	0.00	1.000	3	30	11.12	0.00	1.000

Brazil CDS					Inflation Swaps				
1	10	1.14	0.70	0.963	3	10	1.14	0.11	0.988
1	15	1.13	0.73	0.963	3	15	2.01	0.00	0.988
1	20	1.14	0.71	0.963	3	20	3.73	0.00	0.988
1	30	1.16	0.69	0.963	3	30	8.17	0.00	0.988
2	10	2.28	0.00	0.997	4	10	0.88	0.99	0.994
2	15	4.49	0.00	0.997	4	15	1.51	0.00	0.994
2	20	7.78	0.00	0.997	4	20	2.79	0.00	0.994
2	30	17.62	0.00	0.997	4	30	6.10	0.00	0.994
3	10	1.11	0.09	0.999	5	10	1.00	0.62	0.997
3	15	1.37	0.05	0.999	5	15	1.77	0.00	0.997
3	20	1.72	0.03	0.999	5	20	3.27	0.00	0.997
3	30	2.73	0.01	0.999	5	30	7.12	0.00	0.997

Note. For each term structure, the table reports the variance ratio test and a p-value for the one-sided test that variance ratio is greater than 1. Each panel reports the test for a different number of factors (first column) and at a different maturity (second column). The table also reports the panel R2, computed as the fraction of the total variation explained by the first K principal components.

G Additional Simulations of Non-affine Models

Figure A24: Multifractal Variance Model



Note. Simulation of the multifractal volatility model as in Calvet and Fisher (2004), and variance ratio test with 2 factors (left panel) or 3 factors (right panel). See also Figure 1.

This appendix present results when from applying our variance ratio tests to additional non-affine term structure. Table A8 extends the analysis of non-linear logistic STAR model to allow for heteroskedastic shocks. The specifications are identical to those in Table 2 and

Table A8: Non-linear Specification with Heteroskedasticity

		ρ=0.01			ρ=0.10			ρ=0.25		
γ	K	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	99.0	1.10	1.21	99.8	1.03	1.04	100.0	1.00	1.00
0.5	2.0	100.0	0.97	0.94	100.0	1.00	1.00	100.0	1.00	1.00
0.5	3.0	100.0	0.99	0.99	100.0	1.01	1.01	100.0	1.00	1.00
1.0	1.0	99.8	1.02	1.05	99.7	1.05	1.07	99.9	1.00	1.00
1.0	2.0	100.0	1.00	1.01	100.0	0.99	0.98	100.0	1.00	1.00
1.0	3.0	100.0	0.99	0.98	100.0	0.99	0.99	100.0	1.00	1.00
5.0	1.0	99.9	1.01	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	0.99	0.98	100.0	1.00	1.00
5.0	3.0	100.0	0.99	0.97	100.0	0.99	0.98	100.0	1.00	1.00

Note. Variance ratios and R^2 computed in simulations of a logistic STAR model with parameters γ and ρ. Shocks are GARCH(1,1) with parameters $\alpha = 0.05$ and $\beta = 0.90$, and with an unconditional standard deviation of one. 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.

the shocks share the same the unconditional shock variance. In Table A8, however, the shocks follow a GARCH(1,1) process with parameters of $\alpha = 0.05$ and $\beta = 0.90$. The results and conclusions from Table 2 are unaffected by the presence of heteroskedasticity.

Next, we analyze the behavior of the variance ratio test for models with more complicated $\mathbb{Q}$-dynamics. In particular, Table A9 reports results for various processes that additively combine a non-linear logistic STAR component (as in Section 4.4) and an ARFIMA component (as in Section 4.3). Because these specifications involve richer driving processes that the individual STAR and ARFIMA analyses in the main text, we allow the estimated affine model to have up to four factors. Again, this extended analysis does not change our conclusions from the main text.

Finally, we simulate the multifractal model of Calvet and Fisher (2004) for variance, and study the term structure of variance claims with up to 24 months maturity. We use the same parameterization of the variance process as in Calvet and Fisher (2004). Figure A24 shows that a 2-factor affine model generates a variance ratio of 1.7 at 24, and adding a third factor brings the variance ratio down to 1.2.

Table A9: Non-linear and Long Memory Mixture Models

		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}
				Panel A: Non-linear component $\rho = 0.01$, $\gamma = 0.1$						
0.10	2	100.0	1.10	1.22	100.0	1.11	1.31	100.0	1.06	1.22
0.10	3	100.0	1.00	1.04	100.0	1.01	1.09	100.0	1.02	1.14
0.10	4	100.0	1.00	1.02	100.0	1.00	1.00	100.0	1.00	1.00
0.20	2	100.0	1.02	1.26	100.0	1.22	1.69	100.0	1.10	1.37
0.20	3	100.0	1.00	1.09	100.0	1.01	1.14	100.0	1.03	1.20
0.20	4	100.0	1.00	1.05	100.0	1.00	0.99	100.0	1.00	1.01
0.40	2	100.0	0.95	0.98	100.0	1.36	2.32	100.0	1.08	1.37
0.40	3	100.0	1.01	1.14	100.0	1.01	1.16	100.0	1.02	1.21
0.40	4	100.0	1.03	1.25	100.0	1.00	0.99	100.0	1.01	1.07
0.49	2	100.0	1.06	1.34	100.0	1.38	2.47	100.0	1.03	1.22
0.49	3	100.0	1.01	1.14	100.0	1.01	1.17	100.0	1.01	1.16
0.49	4	100.0	1.03	1.27	100.0	1.00	0.99	100.0	1.02	1.18
				Panel B: Non-linear component $\rho = 0.01$, $\gamma = 0.5$						
0.10	2	99.9	1.04	1.16	100.0	1.04	1.16	100.0	1.06	1.22
0.10	3	100.0	1.01	1.09	100.0	1.01	1.07	100.0	1.01	1.10
0.10	4	100.0	1.02	1.13	100.0	1.02	1.13	100.0	1.02	1.11
0.20	2	99.9	1.03	1.14	99.9	1.06	1.21	100.0	1.08	1.28
0.20	3	100.0	1.01	1.10	100.0	1.00	1.07	100.0	1.02	1.13
0.20	4	100.0	1.02	1.13	100.0	1.02	1.13	100.0	1.02	1.13
0.40	2	99.9	1.03	1.17	99.9	1.16	1.53	100.0	1.11	1.47
0.40	3	100.0	1.01	1.13	100.0	1.01	1.13	100.0	1.03	1.24
0.40	4	100.0	1.03	1.19	100.0	1.02	1.16	100.0	1.02	1.18
0.49	2	99.8	1.05	1.25	99.9	1.21	1.74	100.0	1.12	1.55
0.49	3	100.0	1.01	1.15	100.0	1.02	1.19	100.0	1.04	1.27
0.49	4	100.0	1.03	1.24	100.0	1.02	1.13	100.0	1.02	1.19
				Panel C: Non-linear component $\rho = 0.01$, $\gamma = 5.0$						
0.10	2	100.0	1.00	1.00	100.0	1.00	1.00	100.0	1.01	1.02
0.10	3	100.0	1.00	0.99	100.0	1.00	0.99	100.0	1.00	0.99
0.10	4	100.0	1.00	1.01	100.0	1.00	1.01	100.0	1.00	1.01
0.20	2	100.0	1.00	0.99	100.0	1.00	1.00	100.0	1.02	1.09
0.20	3	100.0	1.00	0.99	100.0	1.00	0.99	100.0	1.00	0.99
0.20	4	100.0	1.00	1.01	100.0	1.00	1.01	100.0	1.00	1.00
0.40	2	100.0	1.00	0.99	100.0	1.01	1.03	100.0	1.01	1.05
0.40	3	100.0	1.00	0.99	100.0	1.00	0.99	100.0	1.00	1.00
0.40	4	100.0	1.00	1.01	100.0	1.00	1.01	100.0	1.00	1.01
0.49	2	100.0	1.00	0.99	100.0	1.03	1.10	100.0	1.00	1.01
0.49	3	100.0	1.00	0.99	100.0	1.00	0.99	100.0	1.00	1.01
0.49	4	100.0	1.00	1.01	100.0	1.00	1.01	100.0	1.00	1.01

Note. Variance ratios and R^2 computed in simulations of a mixture model that is the sum of ARFIMA(1,d,0) and logistic STAR(ρ, γ) processes. 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.