Excess Volatility: Beyond Discount Rates*

Stefano Giglio Bryan Kelly

University of Chicago and NBER

April 14, 2015

Preliminary and Incomplete

Abstract

We document a form of excess volatility that is irreconcilable with standard models of prices, and in particular *cannot* be explained by variation in the discount rates of rational agents. We compare behavior of prices to claims on the same stream of cash flows but with different maturities. Prices of long-maturity claims are dramatically more variable than justified by the behavior of short maturity claims. Our analysis suggests that investors pervasively violate the "law of iterated values." The violations that we document are highly significant both statistically and economically, and are evident in all asset classes we study, including equity options, credit default swaps, volatility swaps, interest rate swaps, inflation swaps, and dividend futures.

*We are grateful to Drew Creal, Lloyd Han, Lars Hansen, and Stavros Panageas for many insightful comments.

1 Introduction

The field of modern financial economics is in large part organized around the notion of excess volatility in asset prices. As Shiller (1981) and others famously document, price fluctuations are "excessive" relative to predictions from the constant discount rate model. A potential resolution of the puzzle is to recognize that discount rates are variable. The leading frameworks of modern finance indeed center on descriptions of discount rate variation in models of rational expectations.

In this paper we document a form of excess volatility that is irreconcilable with standard models of prices, and that in particular *cannot* be explained by variation in the discount rates of rational agents. We compare behavior of prices to claims on the same stream of cash flows but with different maturities. Our analysis suggests that investors pervasively violate the "law of iterated values" (Anderson, Hansen, and Sargent (2003)). That is, the law of iterated expectations dictates that prices of long maturity claims reflect investors' expectations about the future value of short maturity claims. This imposes consistency requirements on the joint behavior of prices across the term structure.

We document an internal inconsistency in the price behavior of short and long maturity claims. In particular, prices on the long end of the curve are dramatically more variable than justified by the behavior of the short end. These violations are highly significant both statistically and economically. Excess volatility of long maturity prices is evident in all asset classes we study, including claims to equity volatility, sovereign and corporate credit default risk, interest rates, inflation, and corporate dividends.

A simple example illustrates the essence of our approach. Consider an asset that yields an uncertain cash coupon of x_t each period. At time t, an n-maturity claim receives the cash flow x_{t+n} in period n. In the absence of arbitrage, the prices of claims across all maturities are coupled by the dynamics of x_t.

No-arbitrage implies the existence of a pricing measure, $\mathcal{Q}$, that subsumes all risk premia, and their potentially time-varying dynamics, by construction. Under $\mathcal{Q}$, prices are expectations of future cash flows:[1]

$$p_t^n = E_t^{\mathcal{Q}}\left[x_{t+n}\right].$$

The convenience of representing prices as $\mathcal{Q}$-expectations is that any excess volatility that we find in prices *cannot* be due to a time-varying risk premium, by definition. If we can

[1]For exposition we assume here that the risk free r_t is 0, or, alternatively, that all cash flows in the contract are exchanged at maturity, as in the case of swaps. Later sections address the role of time varying interest rates in detail.

identify the dynamics of x_t under the $\mathcal{Q}$-measure, we can evaluate $E_t^{\mathcal{Q}}$ and test whether it is consistent with the price behavior observed in data.

In particular, suppose the $\mathcal{Q}$-dynamics of x_t obey a first order autoregression:

$$x_t = \bar{x} + \rho^{\mathcal{Q}} x_{t-1} + u_t$$

and we observe the prices of claims with maturities of one and two periods:

$$p_t^1 = E_t^{\mathcal{Q}}[x_{t+1}] = \bar{x} + \rho^{\mathcal{Q}} x_t$$

and

$$p_t^2 = E_t^{\mathcal{Q}}[x_{t+2}] = \bar{x} + \rho^{\mathcal{Q}} \bar{x} + (\rho^{\mathcal{Q}})^2 x_t$$

In this example, prices at all maturities share a one-factor structure.

Now, consider the comovement of these two prices, summarized by a regression of p_t^2 on p_t^1. The regression coefficient β is:

$$\beta = \frac{Cov(p_t^1, p_t^2)}{V(p_t^1)} = \frac{(\rho^{\mathcal{Q}})^3 \sigma_x^2}{(\rho^{\mathcal{Q}})^2 \sigma_x^2} = \rho^{\mathcal{Q}}$$

The comovement of the prices of the two maturities perfectly reveals the risk-neutral persistence implied by investors' valuation of the two claims. Intuitively, if investors price assets as though x_t is very persistent ($\rho^{\mathcal{Q}}$ close to 1), a rise in the short price p_t^1 will coincide with a rise in p_t^2 of nearly the same magnitude. That is, the stronger the comovement of prices at different maturities, the more persistent the dynamics of x_t under the pricing measure $\mathcal{Q}$.

In this one factor example, we can infer $\rho^{\mathcal{Q}}$ by observing at least two points in the term structure. From here, the price variance at any other maturity $n > 2$ is entirely determined by $\rho^{\mathcal{Q}}$ and the variance of the short maturity claim:

$$V(p_t^n) = V(p_t^1)(\rho^{\mathcal{Q}})^{2(n-1)} \tag{1}$$

Furthermore, once we infer $\rho^{\mathcal{Q}}$ from the short end of the term structure we can test for excess volatility at all other maturities by comparing the observed price variance, $V(p_t^n)$, to the predicted variance $(\rho^{\mathcal{Q}})^{2(n-1)} V(p_t^1)$. This is the key insight: In order to describe the variance of prices any maturity $n > 2$, one only needs to know the variance and covariance of p_t^1 and p_t^2 and a specification of dynamics under $\mathcal{Q}$. Second moment estimates from short maturity prices are readily available, and the model dynamics that we consider are sufficiently general to subsume all major asset pricing paradigms.

Importantly, a failure of the $\mathcal{Q}$-dynamics extracted from the short end of the curve to explain prices at the long end of the curve can be directly linked to a violation of the law of iterated values. Under $\mathcal{Q}$, iterated expectations bind together prices of claims across maturities. For example, the price of the n-period claim is coupled with the price of 1-period claim:

$$p_t^n = E_t^{\mathcal{Q}}[x_{t+n}] = E_t^{\mathcal{Q}}[E_{t+n-1}^{\mathcal{Q}}[x_{t+n}]] = E_t^{\mathcal{Q}}[p_{t+n-1}^1] = c + (\rho^{\mathcal{Q}})^{n-1} x_t$$

where

$$c = \overline{x} \sum_{j=0}^{n-1} (\rho^{\mathcal{Q}})^j.$$

A violation of this equation – for example, by inconsistent variances of the left-hand and right-hand sides – constitutes a failure of the law of iterated values.

We develop a general methodology for measuring and testing excess volatility that requires minimal modeling assumptions and exploits the information contained in the term structure of cash flow claims on any asset. Our methodology extends the preceding example to any setting in which the cash-flow variable x_t follows a factor structure with linear dynamics under $\mathcal{Q}$ and an arbitrary number of factors. This assumption is valid in standard term structure models, where it is typically derived from a linear factor structure under the physical measure together with an affine stochastic discount factor. It also describes many general-equilibrium asset pricing models, such as the long-run risks model of Bansal and Yaron (2004), the rare disaster model of Wachter (2013), and the cash flow duration model of Lettau and Wachter (2006). Furthermore, our setting allows for time-varying risk prices and stochastic volatility.

Our tests of excess volatility are remarkable for what they do *not* require. We require no data other than asset prices, obviating the need for data on the underlying cash flows x_t. Nor do we require a model of discount rates. By definition, the behavior of x_t under the pricing measure implicitly captures all relevant discount rate variation. In our approach, any surplus (or deficit) of p_t^n variance relative to its predicted value already accounts for discount rate effects. Finally, we require a minimal amount of time series information (used to compute the covariances of prices at different maturities). It is well known from the term structure literature that the dynamics under $\mathcal{Q}$ are precisely estimated with short time series via cross-sectional regressions of prices along the term structure. We leverage this fact and focus exclusively on $\mathcal{Q}$ dynamics, avoiding the difficulties of estimating physical dynamics and stochastic discount factors.

We consider a number of potential explanations for the excess volatility of long maturity claims such as liquidity and omitted factors. The evidence is inconsistent with both of these explanations. For most asset classes, we have detailed liquidity information across the term structure, we study only those maturities that transact regularly each day, and we demonstrate that observed price volatility is unassociated with bid-ask bounce or stale prices. Nor are omitted factors likely to explain our findings. First, parsimonious linear factor structures provide an extremely accurate description of every term structure that we study, with R^2 values generally exceeding 98% based on one to three factors. Second, we project long maturity prices onto the short end of the curve, and show that even the variance of the projected prices (which are 100% explained by short prices) remain highly significantly excessive. Finally, we perform a range of robustness tests allowing for richer factor structures, and our findings remain essentially unchanged.

Our paper is related to Stein (1989) in terms of economic intuition and implementation, who studies the pricing of volatility of the S&P 100 using the term structure of implied volatility of options. Stein compares $\rho^{\mathcal{Q}}$ with the persistence of volatility under the physical measure ρ. He finds that $\rho^{\mathcal{Q}} > \rho$, and interprets it as evidence of overreaction. We build on Stein (1989) in three ways. First and foremost, we do not compare physical and risk-neutral dynamics of the underlying cash flow process (ρ and $\rho^{\mathcal{Q}}$). This is a conceptually crucial difference. As the term structure literature (developed largely subsequent to Stein's analysis) has pointed out, $\rho^{\mathcal{Q}}$ will generally be different from ρ in the presence of time-varying discount rates, even in a rational model. In fact, in standard affine models $\rho^{\mathcal{Q}} - \rho$ exactly measures the time-varying component of risk premia. In other words, time varying discount rates provide a natural potential explanation for Stein's facts, in the same way that discount rate variation helps resolve Shiller's original excess volatility puzzle. In contrast, we use only information in the cross-section of prices, solely focusing on dynamics under the pricing measure. In essence, we compare the estimates of $\rho^{\mathcal{Q}}$ from the short end of the curve to $\rho^{\mathcal{Q}}$ implied by the long end. Because our analysis is *entirely* conducted under the pricing measure, time-variation in risk premia cannot play a role in our analysis. Any overreaction (excess volatility) that we find is not mechanically resolved by a time-varying discount rate explanation. The second contribution of our paper is to show that this excess volatility phenomenon is not merely a feature of the options market. It holds across diverse asset classes and across countries. Finally, we propose a general methodology that allows for an arbitrary factor structure (as opposed to the single factor model in Stein) and derive a test statistic for excess volatility.

In Section **??** we present our general modeling framework, of which the preceding one-factor example is a special case. We present our approach to estimation and inference in Section **??**. In Section **??** documents the our central empirical facts for excess volatility across many asset classes. Section **??** discusses a number of extensions and Section **??** concludes.

2 Affine Term Structures and Pricing Under $\mathcal{Q}$

In this section we specify the dynamic structure of the economy under a probability measure denoted $\mathcal{Q}$, the pricing measure. [2]The state space under Q is described by a vector autoregressive process for the factors, F_t:

$$F_t = c^{\mathcal{Q}} + \rho^{\mathcal{Q}} F_{t-1} + \Sigma e_t \tag{2}$$

For simplicity, we assume that the innovation vector, e_t, is standard Gaussian and discuss the implications of this assumption below.

We consider an asset whose scalar log cash flow growth process, x_t, depends on the state vector according to

$$x_t = \delta_0 + \delta_1' F_t. \tag{3}$$

Assuming an investor who is risk neutral under the measure $\mathcal{Q}$, the price of an exponential claim on x_t with maturity n (for example a bond) is given by

$$\begin{aligned} P_{t,n} &= E^{\mathcal{Q}}\left[\exp\left(\sum_{j=0}^{n-1} x_{t+j}\right)\right] && (4) \\ &= \exp(a_n + b_n F_t) && (5) \end{aligned}$$

where the factor loading depends on the maturity of the claim according to

$$b_n = \delta_1'(I + \rho^{\mathcal{Q}} + ... + (\rho^{\mathcal{Q}})^{n-1}). \tag{6}$$

The intercept is an inconsequential constant function of remaining model parameters that

[2]The pricing measure is a transformation of the objective statistical measure that scales physical probabilities by investors' marginal utilities state-by-state. This carries the implication that asset prices are martingales under Q, a feature that we exploit in our development. Such a measure is guaranteed to exist under the minimal assumption of no-arbitrage (Harrison and Kreps (1979), Harrison and Pliska (1981)).

drops out from all variance calculations. Note also that no risk-free rate adjustment appears in equation (**??**), a point we discuss below. All assumptions are discussed briefly in the next section and in detail, asset class by asset class, in the appendix.

How variable are prices in this model? What drives the comovement across claims of different maturity? Price fluctuations for all claims to x_t are entirely driven by fluctuations in the state factors, but they differ with maturity depending on the state persistence matrix, $\rho^{\mathbb{Q}}$. Consider the stationary case in which all eigenvalue of $\rho^{\mathbb{Q}}$ are strictly less than one in modulus. At the short end, the claim receives a single cash flow, and has sensitivity to state fluctuations given by $\delta_1'(I + \rho^{\mathbb{Q}})$. As the maturity rises to n, the claim receives a total of n cash flows. Because x_t is persistent, this claim is more sensitive to the state, represented by the additional $\rho^{\mathbb{Q}}$ term in (**??**). The increasing powers of $\rho^{\mathbb{Q}}$ also indicate that more distant cash flows matter less for price fluctuations today. Thus, price volatility is an increasing but *concave* function of maturity.

Moreover, the model structure places restrictions on the exact price variances that are admissible. If there are K factors, but $n > K$ maturities, then the variance of any $n - K$ prices in the term structure are entirely pinned down by the other K price variances, through $\rho^{\mathbb{Q}}$.

2.1 Discussion of the main assumptions

Given that the results of the paper are derived under the assumptions about the term structure described above, it is important to underline which are the key assumptions in that setup, and which assumptions play instead a minor role.

The two fundamental assumptions of the paper are:

1. Cash flows x_t follow a factor structure, and

2. Factors obey linear dynamics.

The first assumption of the model can be easily verified in the data. The term structure model we described above has an extraordinarily high degree of explanatory power for asset prices in a wide variety of asset classes. For traded claims with a term structure of maturities, we typically find that a small number of latent factors (typically one to three, many fewer than the number of traded maturities) explains close to 100% of the price variation throughout the term structure. The most well known example is the US treasury bond term structure, though we find the same feature among derivatives in credit, equity, currency, and other markets. This strongly supports the assumption of a factor structure.

The second main assumption of the paper is that state dynamics are linear under $\mathcal{Q}$. The vast majority of models in the asset pricing literature assume linear dynamics for the underlying state variables, both in the structural general equilibrium asset pricing models and in the reduced-form term structure literature. In addition, the Wold decomposition applies in our setting as long as the factors are stationary under $\mathcal{Q}$ (stationarity is typically assumed in term structure models and is supported by our empirical evidence). This implies that state dynamics can be represented as a vector moving average process (with possible infinite lags), for which our flexible finite-order VAR can be viewed as an approximation. Finally, Le, Singleton and Dai (2010) discuss conditions for a model to posses a linear affine structure under $\mathcal{Q}$ even when objective dynamics are non-affine.

The specification above also imposes some other, non-critical assumptions that we briefly discuss here and explore more in detail in the Appendix.

Exponentially affine form. We assumed above that the term structures we consider have prices that depend exponentially on the cash flows x_t. This is a natural assumption for some asset classes, such as bonds, in which the prices are exponentially affine functions of $x_t = -r_t$. In other term structures, payoffs are linear in the cash flow variable:

$$P_{t,n} = E^{\mathcal{Q}}\left[\sum_{j=0}^{n-1} x_{t+j}\right]$$

For example, variance swaps are linear claims to realized variance $x_t = RV_t$. These models reflect a linear relation between prices and factors rather than log-linear, so that $P_{t,n} = a_n + b_n F_t$. It is easy to show that linear models imply the same relation between b_n and $\rho^{\mathcal{Q}}$: $b_n = \delta_1'(I + \rho^{\mathcal{Q}} + ... + (\rho^{\mathcal{Q}})^{n-1})$, so conceptually none of our results hinge on the exponentially affine form. For expositional purposes, we use the exponential framework throught the paper, but when testing for excess volatility we will model the different term structures using the appropriate functional form for prices, linear or log-linear, in each asset class.

Risk-free discounting. In this paper we consider two types of term structures of claims to x_t. The first includes contracts in which one party makes an upfront cash payment and then receives x_t over time, as in the case of bonds. The second includes contracts in which neither party has an upfront cash outlay, and payments are exchanged at maturity, as in the case of futures or swaps.[3] This case corresponds exactly to the price formula in equation (**??**). Because all payments happen at the same time, the valuation formula under

[3] In practice, the line between the two is blurred, since even future contracts and other swaps may require intermediate payments in the form of margin payments or intermediate settlements. We ignore these differences here.

$\mathcal{Q}$ holds without a risk-free rate adjustment.

The case with upfront payments can be mapped to eq. (**??**) under the assumpion that the short-term rate also depends on a finite number of factors (recall that our methodology include as many factors as needed according to the data). Consider a claim to the cash flow $X_{t+n} = \exp\left(\sum_{j=1}^{n} x_{t+j}\right)$ occurring at maturity $t+n$. Each claim should be discounted at the appropriate risk-free rate for that maturity:

$$P_{t,n} = \delta(t,n) E_t^{\mathcal{Q}} \left[X_{t+n} \right]$$

where $\delta(t,n)$ is the risk-free discount factor from t to $t+n$ or, equivalently, the price of a $n-$period risk-free bond at time t. Now, consider the short-term log risk-free proces r_t, and assume that it also is a function of the factors, $r_t = h_0 + h_1 F_t$. We have that

$$\delta(t,n) = exp\{a_n^r + b_n^r F_t\}$$

and also that

$$E_t^Q X_{t+n} = exp\{a_n + b_n F_t\}.$$

Combining these we obtain:

$$P_{t,n} = exp\{(a_n^r + a_n) + (b_n^r + b_n)F_t\} = exp\{\tilde{a}_n + \tilde{b}_n F_t\}$$

returning us to the original specification, and showing that our cash flow processes can always be recast as a claim on $\tilde{x}_t = x_t - r_t$ as opposed to a claim on x_t.

Heteroskedasticity. Two elements are important to potentially explain time variation in risk premia. One is the time variation in the price of risk, which is automatically taken into account here, as it affects the relationship between quantities under $\mathcal{Q}$ and under the physical measure, and which we need not explicitly model due to our methodology that works solely under Q. The second is the possibility that the volatility of factor innovations is time-varying, i.e. that shocks take the form $\Sigma_t e_{t+1}$. In exponentially-affine models (though *not* in linear models), time variation in Σ_t affects the coefficients b_n if this variation is correlated with the factors. This is due to a Jensen inequality effect caused by the exponential dependence of prices on future cash flows x_t. In that case, we cannot relate b_n only to the persistence matrix $\rho^{\mathcal{Q}}$, but also need to account for volatility.

In the Appendix we consider a number of robustness checks with regards to heteroskedasticity. First,we show that in the term structures in which it matters (those with exponential-affine form), the extent of excessive price volatility appears even larger when we carefully accounting for heteroskedasticity, compared to more conservative estimates when stochastic volatility is naively treated as a constant. Furthermore, the bond market term structure literature has noticed that the effects of this Jensen adjustment are typically negligible, and we confirm this in our setting. For reasonable calibrations of the volatility process, the effect on the coefficients b_n is nearly undetectable. In the main paper we therefore ignore heteroskedasticity and proceed with a fully homoskedastic model, where time variation in risk premia is entirely due to time-varying risk prices, and direct readers interested in the effects of heteroskedasticity to the appendix.

3 Estimation and Testing

In this section we develop a methodology for testing excess volatility across the term structure given the model specification of the Section **??**. We start by discussing parameter identification for factor dynamics under $\mathcal{Q}$. We then show how to infer dynamics under $\mathcal{Q}$ from the comovement of prices at the short end of the term structure. Finally, we propose a test for "excess volatility" of prices at the long end of the term structure.

3.1 Theoretical identification of the parameters

Suppose that a term structure is driven by K factors, and we observe N prices along the term structure. The factor follow the dynamics reported in eq. **??**. Note that the dynamics of x_t under $\mathcal{Q}$ depend on a large number of parameters: δ_0 (1 parameter), δ_1 (K parameters), $c^{\mathcal{Q}}$ (K parameters), $\rho^{\mathcal{Q}}$ (K^2 parameters), and Σ ($\frac{K(K-1)}{2}$ parameters). Not all these are separately identified, as pointed out by Joslin, Singleton and Zhu (2011): we can restrict some of the parameters to achieve identification. We choose the particular restrictions of Joslin, Singleton and Zhu (2011), that are without loss of generality and have a particularly clear interpretation in our setting.

The identification restrictions of Joslin, Singleton and Zhu (2011) are as follows. We assume δ_1 to be a column of ones. We assume the vector $c^{\mathcal{Q}}$ to be a column of zeros, and the matrix $\rho^{\mathcal{Q}}$ to be diagonal. Finally we impose that Σ is lower triangular. The most interesting identification restriction we impose is that $\rho^{\mathcal{Q}}$ is diagonal: it therefore captures the eigenvalues of the transition matrix of the factors, which in turn summarizes the persistence of the shocks to the factors under the $\mathcal{Q}$ measure. Note that in theory $\rho^{\mathcal{Q}}$ can have both explosive dynamics (when the elements of $\rho^{\mathcal{Q}}$ are greater than 1 in absolute value) and could be complex (implying cycles of the covariances across matirities).

Following the term structure literature, we exclude both cases whenever fully real, nonexplosive dynamics are compatible with the data: both cases have clearly counterfactual predictions for the term structure, and are easy to reject empirically (as for example shown in Hamilton and Wu (2010)).

3.2 Estimation of the dynamics $\rho^{\mathcal{Q}}$

We now show that the covariance of prices of claims along the term structure fully pins down the vector $\rho^{\mathcal{Q}}$. We show that the elements of $\rho^{\mathcal{Q}}$ (recall that $\rho^{\mathcal{Q}}$ is a diagonal matrix) can be recovered as the roots of a polynomial, whose coefficients depend on the covariance of prices.

From equation (**??**), log prices are linear in the state factors

$$p_t^n = a_n + b_n' F_t + u_t^n \tag{7}$$

where the term u_t^n is a typically very small iid measurement error term that is added to the model to avoid stochastic singularity.

Given that the purpose of our analysis is to identify the dynamics of K factors from the "short end" of the term structure, it is convenient to subdivide prices into the short and the long end of the term structure. Let us call the vector of prices of H claims at the short end of the yield curve p_t^{short}. The only requirement we impose is that $H \geq K$, so that the prices at the short end of the curve contain enough information to identify the $\mathcal{Q}$ dynamics of the factors. The claims at the short end of the curve will have maturities $n_1, ..., n_H$.

Finally, analogously to standard term structure models, we assume that while each individual price p_t^n is observed with error u_t^n, the first K principal components of the *short* set of prices are observed without measurement error. Call the vector of K principal components

$$\overline{p}_t = f \cdot p_t^{short}$$

where f is the $K \times H$ matrix of loadings that can be estimated via principal component analysis.

Given this standard setup, we can prove the following Proposition.

Proposition 1. *Consider the regression of a price $p_t^{n_{H+1}}$ onto the factors $\overline{p}_t$:*

$$p_t^{n_{H+1}} = d + c'\overline{p}_t + u_t^{n_{H+1}}$$

All eigenvalues $\rho_i^{\mathcal{Q}}$ of $\rho^{\mathcal{Q}}$ are among the roots of the polynomial equation

$$\left[1+\rho_i^{\mathcal{Q}}+...+(\rho_i^{\mathcal{Q}})^{n_{H+1}-1}\right]=\tilde{c}_1\left[1+\rho_i^{\mathcal{Q}}+...+(\rho_i^{\mathcal{Q}})^{n_1-1}\right]+...+\tilde{c}_H\left[1+\rho_i^{\mathcal{Q}}+...+(\rho_i^{\mathcal{Q}})^{n_H-1}\right]$$

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

Proof. Start by defining

$$S^n \equiv \delta_1'(I+\rho^{\mathcal{Q}}+...+(\rho^{\mathcal{Q}})^{n-1})$$

where δ_1' is a vector of 1s. S^n is a $1\times K$ vector that depends *only* on the diagonal matrix $\rho^{\mathcal{Q}}$. Note that calling $\rho_i^{\mathcal{Q}}$ the ith element of the diagonal of $\rho^{\mathcal{Q}}$, we can rewrite S^n as:

$$S^n=\begin{bmatrix} 1+\rho_1^{\mathcal{Q}}+...+(\rho_1^{\mathcal{Q}})^{n-1} \\ 1+\rho_2^{\mathcal{Q}}+...+(\rho_2^{\mathcal{Q}})^{n-1} \\ ... \\ 1+\rho_K^{\mathcal{Q}}+...+(\rho_K^{\mathcal{Q}})^{n-1} \end{bmatrix}'$$

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

$$\overline{p}_t=f\cdot\begin{bmatrix} a_{n_1} \\ a_{n_2} \\ ... \\ a_{n_H} \end{bmatrix}+f\cdot\begin{bmatrix} S^{n_1} \\ S^{n_2} \\ ... \\ S^{n_H} \end{bmatrix}F_t$$

Consider now the regression:

$$p_t^{n_{H+1}}=d+c'\overline{p}_t+u_t^{n_{H+1}}$$

Substituting for the prices on each side of the equation we obtain:

$$a_n+S^nF_t+u_t^{n_{H+1}}=d+c'f\begin{bmatrix} a_{n_1} \\ a_{n_2} \\ ... \\ a_{n_H} \end{bmatrix}+c'f\begin{bmatrix} S^{n_1} \\ S^{n_2} \\ ... \\ S^{n_H} \end{bmatrix}F_t+u_t^{n_{H+1}}$$

Matching the coefficients on F_t we derive the equation:

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

or:

$$S^{n_{H+1}} = \tilde{c}_1 S^{n_1} + \tilde{c}_2 S^{n_2} + ... + \tilde{c}_H S^{n_H}$$

where the $1xH$ vector $\tilde{c} = c'f$ depends only the factor loadings f and the regression coefficients c. Now, given that as shown above each element i of S^n depends only on element i of the diagonal of $\rho^{\mathcal{Q}}$, this is a system of K independent equations, each of the form:

$$\left[1 + \rho_i^{\mathcal{Q}} + ... + (\rho_i^{\mathcal{Q}})^{n_{H+1}-1}\right] = \tilde{c}_1 \left[1 + \rho_i^{\mathcal{Q}} + ... + (\rho_i^{\mathcal{Q}})^{n_1-1}\right] + ... + \tilde{c}_H \left[1 + \rho_i^{\mathcal{Q}} + ... + (\rho_i^{\mathcal{Q}})^{n_H-1}\right]$$

□

Note that each element i of $\rho^{\mathcal{Q}}$ needs to satisfy this equation: $\rho^{\mathcal{Q}}$ 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).

One consideration is that there will generally be n_{H+1} roots of this polynomial (some of them complex), while we only seek K parameters. This equation shows that the $\mathcal{Q}$-measure dynamics and the comovements of prices only identify the eigenvalues of $\rho^{\mathcal{Q}}$ 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 P and the Q dynamics will be able to choose among them). We use the following selection procedure. 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, we choose real roots whenever possible, since complex roots imply regression coefficients of prices on the factors $\overline{p}_t$ at maturities above n_{H+1} that display cycles across maturities, an implication that is strongly counterfactual.

3.3 The Test for "Excess Volatility"

Thus far, we have derived a regression equation whose coefficients recover the $\mathcal{Q}$-measure state persistence matrix, $\rho^{\mathcal{Q}}$, relying on $H+1$ prices at the short end of the curve. This, together with the term structure pricing model, pins down the variance of prices at *all* remaining maturities (up to the variance of the "pricing errors" u_t^n).

Consider the price of a "long" claim, that with maturity $L > n_{H+1}$, described by

$$p_t^L = a_L + b_L F_t + u_t^L$$

We will refer to the error-free price, $a_L + b_L F_t$, and the "clean" price. We are particularly interested in the clean price variance at the long end of the maturity curve, i.e. the variance of the price movements at the long end that are explained by the movements of the short end of the curve.

While we do not observe the factors, it is easy to see from the derivations in the previous section that, up to a constant, they can be expressed as a linear function of the K principal components $\overline{p}_t$:

$$F_t = R\overline{p}_t$$

where the $K \times K$ matrix R depends only on the matrix $\rho^{\mathcal{Q}}$ and the vector of principal component loadings f. In addition, the regression coefficient b_L is precisely equal to:

$$b_L = \delta_1'(I + \rho^{\mathcal{Q}} + ... + (\rho^{\mathcal{Q}})^{L-1})$$

and therefore it also depends directly on $\rho^{\mathcal{Q}}$. The model estimated from the short end of the curve therefore completely pins down the variance of the clean price at the long end:

$$V^{restricted}(p_t^L) = b_L' R' V(\overline{p}_t) R b_L$$

We can however an empirical version of that variance by estimating the regression:

$$p_t^L = \alpha_L + \beta_L \overline{p}_t + u_t^L$$

therefore actually using information on the empirical comovement of prices at the long and the short end of the curve. We obtain

$$V^{unrestricted}(p_t^L) = \beta_L' V(\overline{p}_t) \beta_L$$

Our test proceeds by comparing the two variances. Note that in both calculations, the variance of the pricing error is removed from consideration. Our test statistic is the ratio of unrestricted to restricted variance:

$$VR_L = \frac{V^{unrestricted}(p_t^L)}{V^{restricted}(p_t^L)}. \tag{8}$$

A variance ratio significantly greater than one indicates excess volatility relative to the model.

We conduct statistical inference for the hypothesis $H_0 : VR_L = 1$. In Appendix XXX, we derive a test statistic t_{VR} that is asymptotically distributed $N(0,1)$ under the null that claims of all maturities are priced according to the K-factor model in equation (**??**), allowing for pricing error at maturities above n_H that is uncorrelated with the factors. To establish the robustness of this test, we also conduct inference via the bootstrap procedure described in Appendix XXX.

The practical implementation of our tests proceeds as follows. For any term structure of claims with N different maturities, we must first select the number of factors $K < N$. Next, we select H prices at the short end of the curve that will be used to estimate the K factors $\overline{p}_t$. H needs to be at least as large as K, while still capturing dynamics implied by the short end of the curve, to allow for a meaningful comparison between the behavior of the "long" and the "short" ends of the curve. We then estimate the K principal components of the H short yields and obtain $\overline{p}_t$ and the matrix of loadings f.

Then, we estimate the contemporaneous regression (**??**), using on the left-hand side the $H+1$ available maturity (the first H were used to construct the factors $\overline{p}_t$). From the regression coefficients, we solve for the K estimated diagonal elements of $\hat{\rho}_Q$ that determine factor persistence under the pricing measure.

The final step is to estimate the actual and predicted variances of prices at remaining maturities ($L = n_{H+1}, ..., n_N$). For each long maturity series, we re-estimate regression (**??**) replacing the L-maturity price as the dependent variable. The explained variance in this regression, defined in equation (**??**), serves as our estimate of the actual price variance, $\hat{V}^{unrestricted}$. By focusing on the fitted prices and discarding the regression error, we impose that 100% of our "actual" price variance is explained by the K prices at the short end of the curve. Any uncorrelated variation due to measurement error or omitted factors is not allowed to enter into our excess variance calculation.

The predicted variance under the model, $\hat{V}^{restricted}$, is calculated using $\hat{\rho}_Q$ as shown in (**??**). The test statistic is calculated as $\hat{VR} = \hat{V}^{unrestricted} / \hat{V}^{restricted}$, and we conduct inference using the bootstrap procedure described in Appendix XXX.

4 Empirical Findings

4.1 Term structures across asset classes: data and models

In this section we briefly describe the asset classes for which we test for excess volatility. In each case, the pricing and factor structure described in section XXX applies with small modifications, all of which are reviewed in detail in the Appendix. We also leave for the appendix a more in-depth description of the data.

4.1.1 Interest Rates

US government bond prices are among the most well studied data in all of economics. Our data US bond data comes from Guryanak, Sack and Wright (2006). The data consists of zero-coupon nominal rates with maturities of 1 to 30 years for the period 1985 to 2014, and is available at the daily frequency. The term structure is bootstrapped from coupon bonds and uses only interpolation, not extrapolation (so that a maturity will only be present if enough coupon bonds are available for interpolation at that maturity). Given the high liquidity of the Treasuries market, we use all available maturities starting in 1985.

4.1.2 Credit default swaps

Credit default swaps (CDS) are the primary security used to trade and hedging default risk of corporations and sovereigns. As of December 2014, the notional value of single-name CDS outstanding was $10.8 trillion. Our CDS data is from MarkIt. The CDS data includes maturities of 1, 3, 5, 7, 10, 15, 20, and 30 years for the period 2001 to 2013 and is available at the daily frequency. While not all maturities are equally traded, we focus on the most liquid single-name and sovereign CDS. It is useful to remember that our test will be conducted and reported separately at all available maturities; so it will be easy to assess to what extent the results are driven by maturities for which liquidity is high or low. Among the different CDS contracts written on the same reference entity, we choose those with highest liquidity. In particular, we choose CDS written on senior bonds, with modified-restructuring (MR) clause, and denominated in US dollars.[4] Since there was little CDS activity before the financial crisis and most of these contracts had low liquidity, we focus 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) during 2008[5], a year in which CDS trading volume and

[4]For sovereigns, we use contracts with the CR clause, as more data is available than for the MR contracts.
[5]See Fitch (2009).

CDS spread variability were particularly high.

The Appendix describes how CDS prices can be represented in the framework of Section XXX.

4.1.3 Inflation swaps

We obtain inflation swaps data from JP Morgan. We observe the full term structure between 1 and 30 years, at the daily frequency, between 2004 and 2014. As reported in Fleming and Sporn (2013), “Despite a low level of activity and its over-the-counter nature, 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.”. In addition, there is significant trading voume at all maturities, including the very top ones (20 to 30 years).

The term structure model for inflation swaps follows closely the benchmark of Section XXX. We report the details in the appendix.

4.1.4 Variance claims

Markets for financial volatility, including options and variance swap markets, possess a rich term structure of claims. These markets allow participants to trade and hedge the price volatility of effectively any financial security, including equity indices and individual stocks, currencies, government and corporate bonds, etc.

The first market we study is that for variance swaps on the S&P 500 index, claims to future realized variance (the sum of squared daily returns of the index). The price of a variance swap corresponds on the expectation (under Q) of future realized variance:

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

and therefore it fits directly into our framework (with the only difference that the price depends on the payoff variable x_t in levels, not in logs). As discussed in detail in Dew-Becker et al. (2015), the variance swap market is an over-the-counter market with a total outstanding notional of around \$4bn vega at the end of 2013 (meaning that a movement of one point in volatility would result in \$4bn changing hands). More importantly, the price of a variance swap is anchored to the price of a syntetic swap that can be constructed from option prices. Dew-Becker et al. (2015) show that the term structure of variance swap prices matches very closely the term structure of synthetic claims constructed from option (typically known as the VIX).

We also the term structure of at-the-money implied volatilities extracted from options in a variety of asset classes. In theory, synthetic prices of variance swaps (which follow exactly equation XXX) can be constructed in any market by combining the prices of puts and calls at different strikes: the price of these synthetic portfolio, commonly known as the VIX, is tied by arbitrage to the price of a variance swap. This would theoretically allow us to study the term structure of variance swap prices in any markets in which we can observe put and call prices, even if actual variance swaps are not trade. Unfortunately, for many asset classes not enough strikes are available to reliably construct the term structure of the VIX. We therefore rely on at-the-money (ATM) implied variances as a proxy for the VIX. This brings us closer to the original setup of Stein (1989), who was working with ATM implied volatilities, and is in part justified by the observation that ATM implied volatilities correspond – up to a first-order approximation – to the prices of claims to realized volatility ($\sqrt{RV}$), as demonstrated by Carr and Lee (2009). Using ATM implied variances allows us to study a large number of markets.

In addition to variance swap on the S&P 500, we study the term structure of ATM implied volatilities for the most liquid options available in OptionMetrics: three domestic indices (S&P 500, Nasdaq, Dow Jones), three international indices (Stoxx 50, FTSE, DAX), three individual names (Apple, Citigroup, IBM), and three currency options,

4.1.5 Currencies

Currency forwards are the primary contracts (along with currency swaps) used to trade and hedge exchange rate risk. As of December 2014, the notional value outstanding of currency forwards and swaps was over $60 trillion. Our currency data is from JPMorgan Dataquery. We study six different currencies (versus the US dollar). For four of these we have maturities of 1, 3, 6, 9, and 12 months. For the Euro and the Mexican Peso we have maturities up to 15 years. Some data are available from 1996 to 2014, and all series have data at least as far back as 2002. We have daily data for all currencies.

We also study currency options, which allow investors to trade and hedge exchange rate volatility. Our data are from JPMorgan Dataquery and have maturities of XXX for the period 1990 to 2014 at the monthly frequency. These data focus on the term structure of Black-Scholes implied volatility.

4.1.6 Dividend Claims

We obtain Stoxx 50 dividend futures prices from Bloomberg and Eurex (using the latter prices whenever available). We obtain weekly series from May 2009 to January 2015 (we use weekly data to reduce the impact of noise). Dividend futures data are obtained by interpolating contracts with fixed expiration dates in the December of each year. Since part of the dividends expiring in the first year are already accrued at the time of the transaction, we exclude the first maturity from our analysis. We therefore obtain contracts of maturity 2 to 7 years. Finally, we adjust all contracts by the risk-free rate to obtain spot prices rather than futures. This step is useful in comparing the prices of the dividend strips to those of the stock market.

The small set of maturities available does not allow us to really compare the short and the long end of the curve. However, in the case of dividends we actually observe the price of a claim to the whole infinite-horizon set of dividends: the Stoxx 50 itself. We can therefore perform the following exercise: we extract the $\rho^{\mathcal{Q}}$ matrix using the time series of dividend strips, and compare them with the volatility of the price of the stock market.

In particular, from the dividend futures we obtain finite-maturity claims to dividend growth of the form:

$$\frac{P_t^n}{D_t} = E^{\mathcal{Q}}[exp\{\sum_{j=0}^{n-1} x_{t+j+1}\}]$$

where $x_{t+j} = \Delta d_{t+j+1} - r_{t+j}$. Using the Campbell-Shiller loglinearization, the price-dividend ratio of the stock market can be approximated as:

$$\frac{P_t}{D_t} = E_t^{\mathcal{Q}}[exp\{k + \sum_{j=0}^{\infty} \theta^j x_{t+j+1}\}]$$

for θ close to 1 (typically calibrated around 0.975).

4.2 Implementation

Before describing the empirical results, we discuss here the procedure we used to identify the number of factors K, the number of prices at the short end of the curve that are used to extract $\rho^{\mathcal{Q}}$.

To choose the number of factors K, we use the panel of prices for all N available maturities to calculate the number of principal components necessary to explain at least 99% of the variance in the panel. This serves as our estimate of K.

We choose H (the prices that define the "short-end" of the term structure) for each asset based on the available maturities at which that cash flow is traded. In particular, for term structures for which claims are typically traded (and data are available) up to 24 months, we define the short end of the curve as composed of maturities up to 6 months. For term structure where the maturities traded are as high as 10 years, we define the short end as 3 years. Finally, for term structures that extend up to 30 years, we define the short and as maturities up to 5 years. The main theoretical justification for linking the definition of the "short end" of the term structure to the set of maturities traded is the following. The set of maturities $n_1, ..., n_N$ we observe to actually trade presumably span the maturities for which investors believe there would be significant price variation. Therefore, we define the short end of the term structure relative to the set of maturities the investors choose to trade.

4.3 Main Results

We find pervasive evidence of excess volatility in each of the term structures we study. Our main findings are summarized in Figure **??**. Clockwise from the upper left, the panels present results for claims to S&P 500 volatility (via options), Spanish sovereign default risk (via CDS), USD/GBP implied volatility and US treasury bonds.

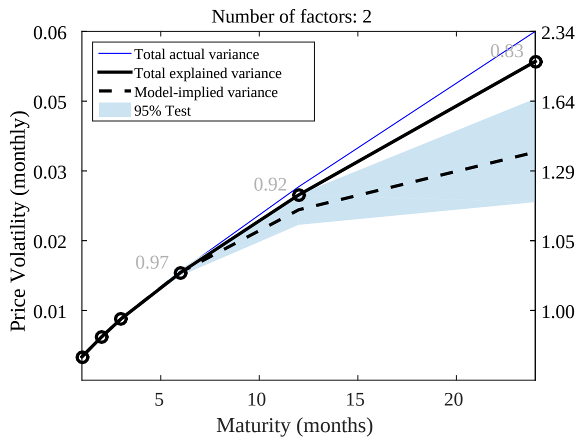
Each figure reports variances of prices across a different term structure term structure. The x-axis shows the maturity of claims. The y-axis on the left side shows the volatility (in standard deviation terms) of prices, while the right axis reports the variance ratio.

The solid thin line in the figure reports the standard deviation of log prices at each maturity. Note that the total volatility at each maturity is a concave function of maturity. This is a first indication that the dynamics under the pricing measure $\mathcal{Q}$ cannot be explosive (in that case, the volatility would be an increasing and convex function of maturity).

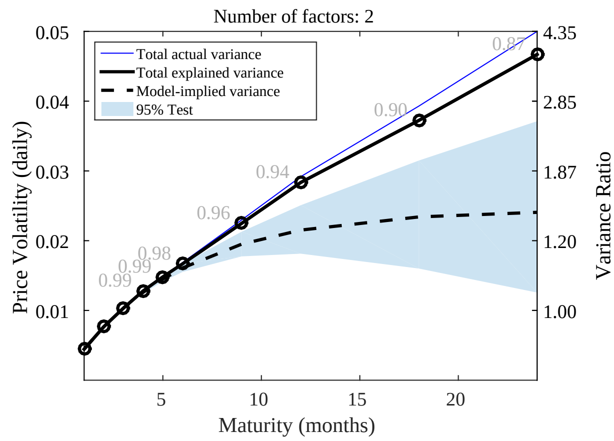
The solid thick line reports the standard deviation of the component of log prices explained by the covariance at the short end, $\sqrt{V(p_t^n)^{unrestricted}}$. The difference between the two solid lines is the part of movements of prices that cannot be explained by comovement with the factors extracted from the short end of the curve: measurement error u_t^n. Note that the factors extracted from the short end of the curve have an extremely high explanation power for *every* maturity along the curve, with R^2 close to 100% even at the very long end of the curve: the unrestricted factor model *fits* the whole term structure extremely well. The dashed line reports the price volatility that the $\mathcal{Q}$ dynamics extracted from the short end of the curve imply at all maturities: $\sqrt{V(p_t^n)^{restricted}}$. Note how in all cases the model-implied volatilities increase with maturity at a much slower rate than the actual

Figure 1: Test of Excess Volatility: 1. Index variances

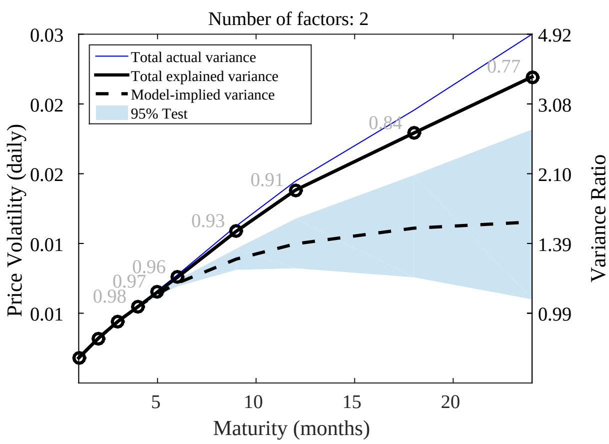
(a) Variance Swaps



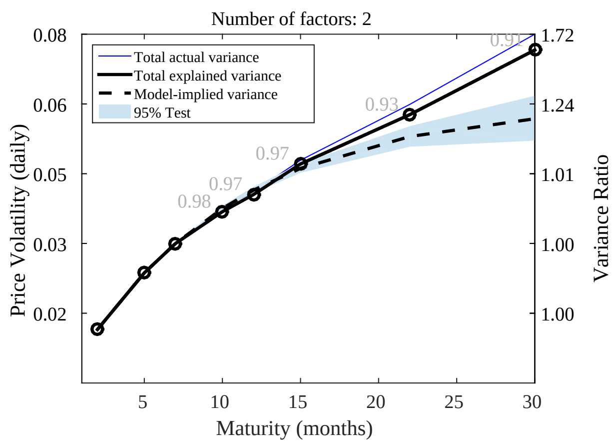
(b) S&P 500 IV



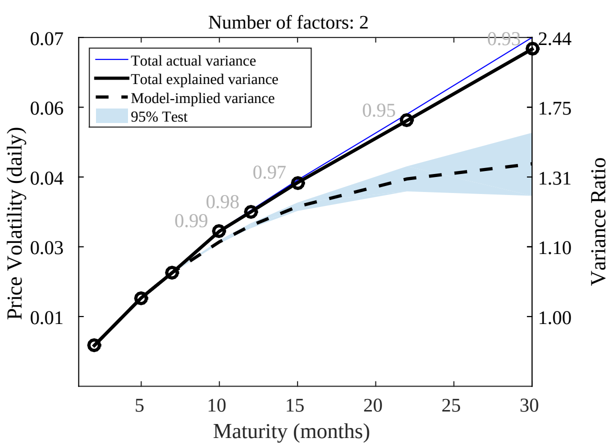
(c) Nasdaq IV



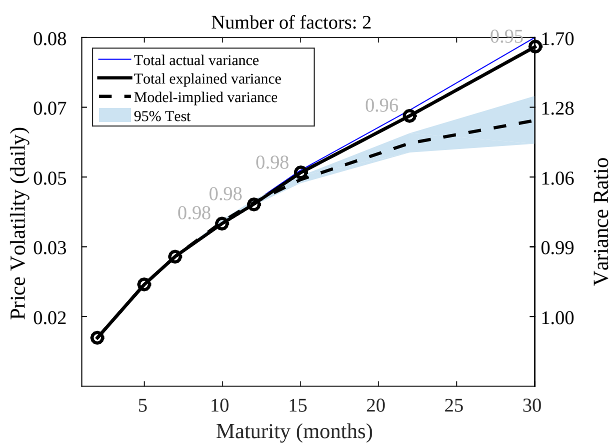
(d) Stoxx 50 IV



(e) FTSE 100 IV



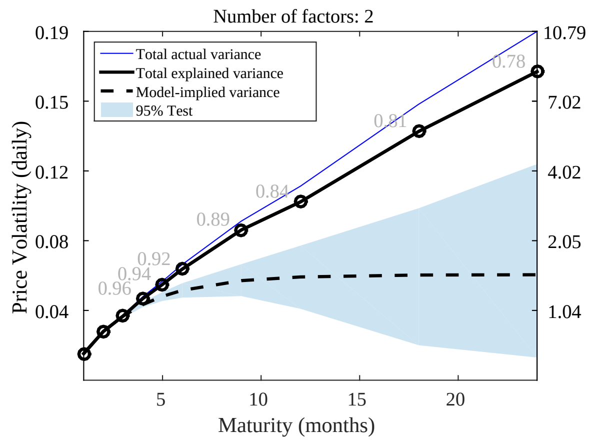
(f) DAX IV



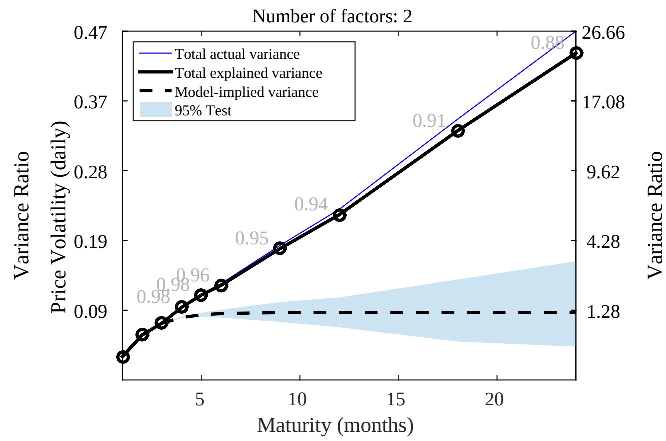
Notes. Each figures plots, for each asset class: 1. the standard deviation of log prices at each maturity (thin solid line); 2. the standard deviation of the component log prices explained by comovement with the short-end factors $\overline{p}_t$, $\sqrt{V(p_t^n)^{unrestricted}}$ (thick solid line); 3. the standard deviation of prices implied by the model estimated from the short end (dotted line), $\sqrt{V(p_t^n)^{restricted}}$. The circles on the unrestricted line represent the maturities we observe in the data. The shaded area encloses the 95th and 5th percentiles of the model-implied variance in bootstrap simulations. The left axis reports the volatility of prices; the right axis reports the ratio $V(p_t^n)^{unrestricted}/V(p_t^n)^{restricted}$.

Figure 2: Test of Excess Volatility: 2. Other variances

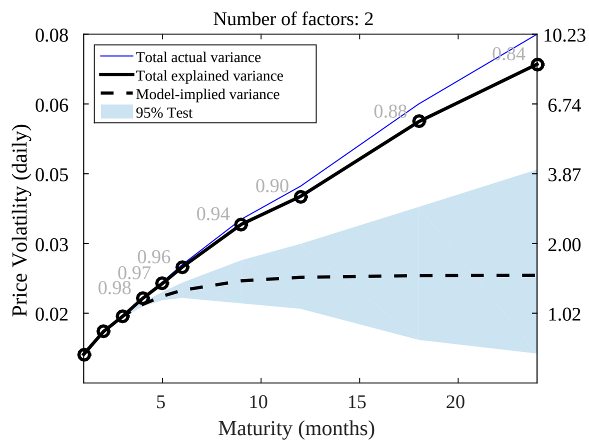
(a) Apple IV



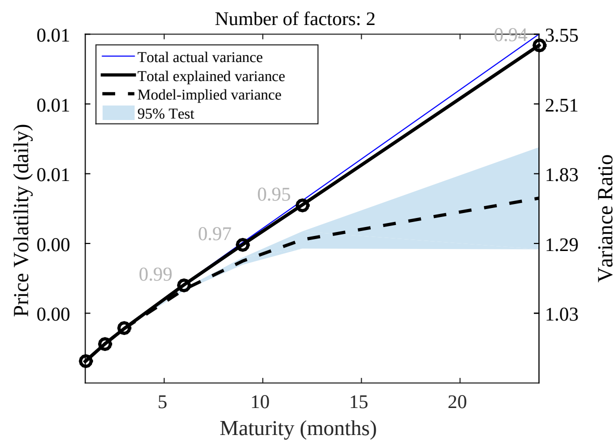
(b) Citigroup IV



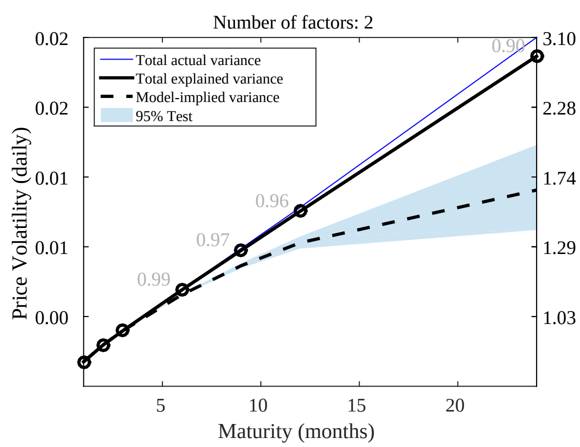
(c) IBM IV



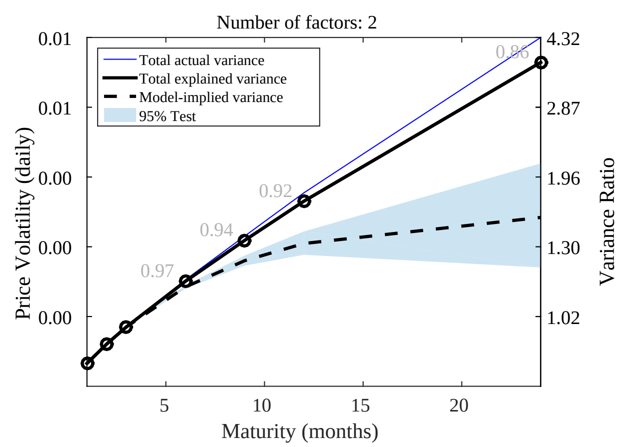
(d) GBP/USD IV



(e) GBP/JPY IV



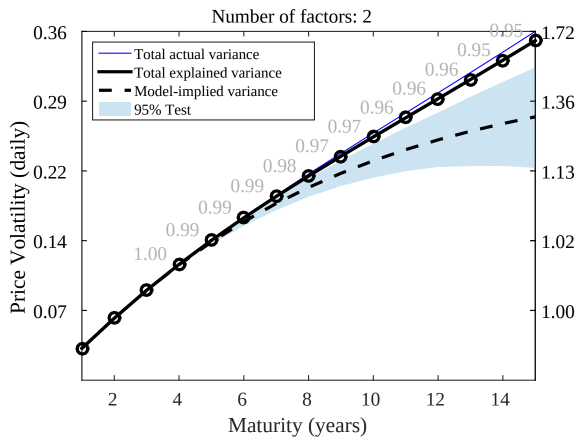
(f) USD/CHF IV



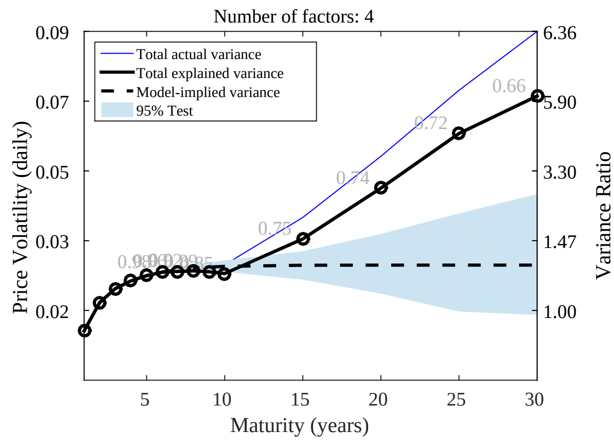
Notes. Each figures plots, for each asset class: 1. the standard deviation of log prices at each maturity (thin solid line); 2. the standard deviation of the component log prices explained by comovement with the short-end factors $\overline{p}_t$, $\sqrt{V(p_t^n)^{unrestricted}}$ (thick solid line); 3. the standard deviation of prices implied by the model estimated from the short end (dotted line), $\sqrt{V(p_t^n)^{restricted}}$. The circles on the unrestricted line represent the maturities we observe in the data. The shaded area encloses the 95th and 5th percentiles of the model-implied variance in bootstrap simulations. The left axis reports the volatility of prices; the right axis reports the ratio $V(p_t^n)^{unrestricted}/V(p_t^n)^{restricted}$.

Figure 3: Test of Excess Volatility: 3. Other term structures

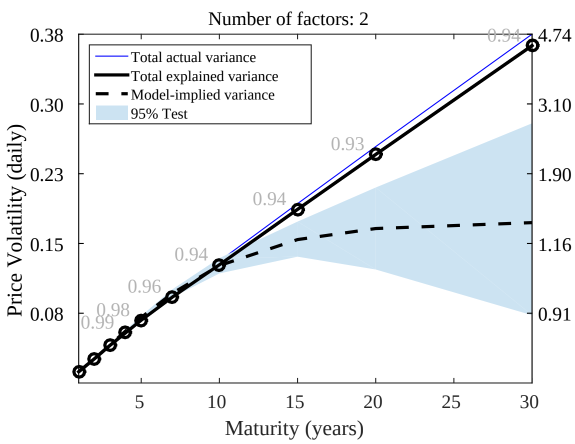
(a) US Treasuries



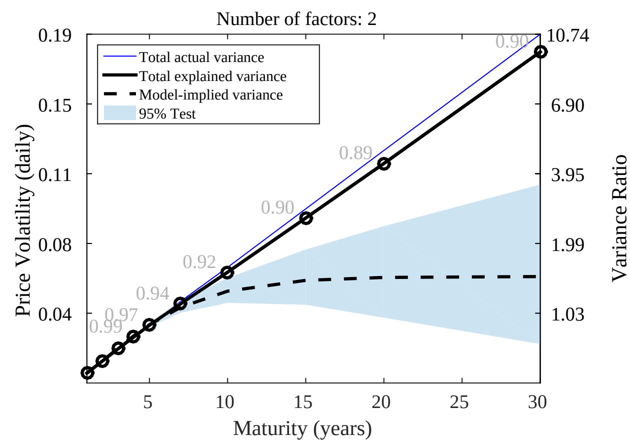
(b) Inflation swaps



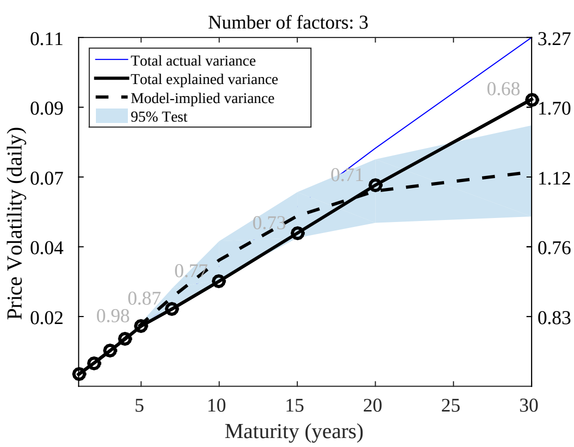
(c) Italy CDS



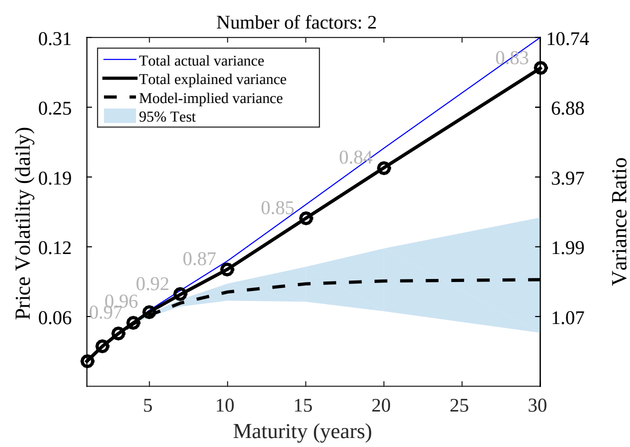
(d) Brazil CDS



(e) JP Morgan CDS



(f) Morgan Stanley CDS



Notes. Each figures plots, for each asset class: 1. the standard deviation of log prices at each maturity (thin solid line); 2. the standard deviation of the component log prices explained by comovement with the short-end factors $\overline{p}_t$, $\sqrt{V(p_t^n)^{unrestricted}}$ (thick solid line); 3. the standard deviation of prices implied by the model estimated from the short end (dotted line), $\sqrt{V(p_t^n)^{restricted}}$. The circles on the unrestricted line represent the maturities we observe in the data. The shaded area encloses the 95th and 5th percentiles of the model-implied variance in bootstrap simulations. The left axis reports the volatility of prices; the right axis reports the ratio $V(p_t^n)^{unrestricted}/V(p_t^n)^{restricted}$.

volatilities of prices. This is precisely what we mean by "excess volatility". The volatility of prices (and the comovement between the long and the short end of the curve) is entirely driven by $\rho^{\mathcal{Q}}$. However, the prices at the long end of the curve react to shocks to the factors driving the short end much more strongly than those dynamics would imply: long-term prices *overreact* to movements in the factors.

The shaded area encloses the 95th and 5th percentiles of the distribution of the volatility under the null of the model (obtained from a bootstrap procedure described in the Appendix). Under a one-sided test of the variance ratio, we reject the null of no overreaction whenever the think solid line ("total explained variance") lies above the shaded area. In all the cases reported here, we find strongly significant evidence of excess volatility.

Finally, two things are important to note. First, the excess volatility we find cannot be explained by movements in discount rates. Under the $\mathcal{Q}$ measure, all prices should just be equal to the expectation of future cash flows. Given the factor model we estimate, this expectation is fully pinned down by $\rho^{\mathcal{Q}}$ (and the current level of the factors), and no other variation in expected cash flows or discount rates can affect prices.

Second, the results of our paper are not about the *fit* of the factor model. The fact that the R^2 of prices on the factors are close to 100% means that even under Q a factor model captures the comovement well. Excess volatility is a test on the *coefficients* that link the long-term prices to movements in the short-term prices (contrast, for example, on other exercises link Pan et al. (XXX) in which the objective of the analysis is the fit of the model itself).

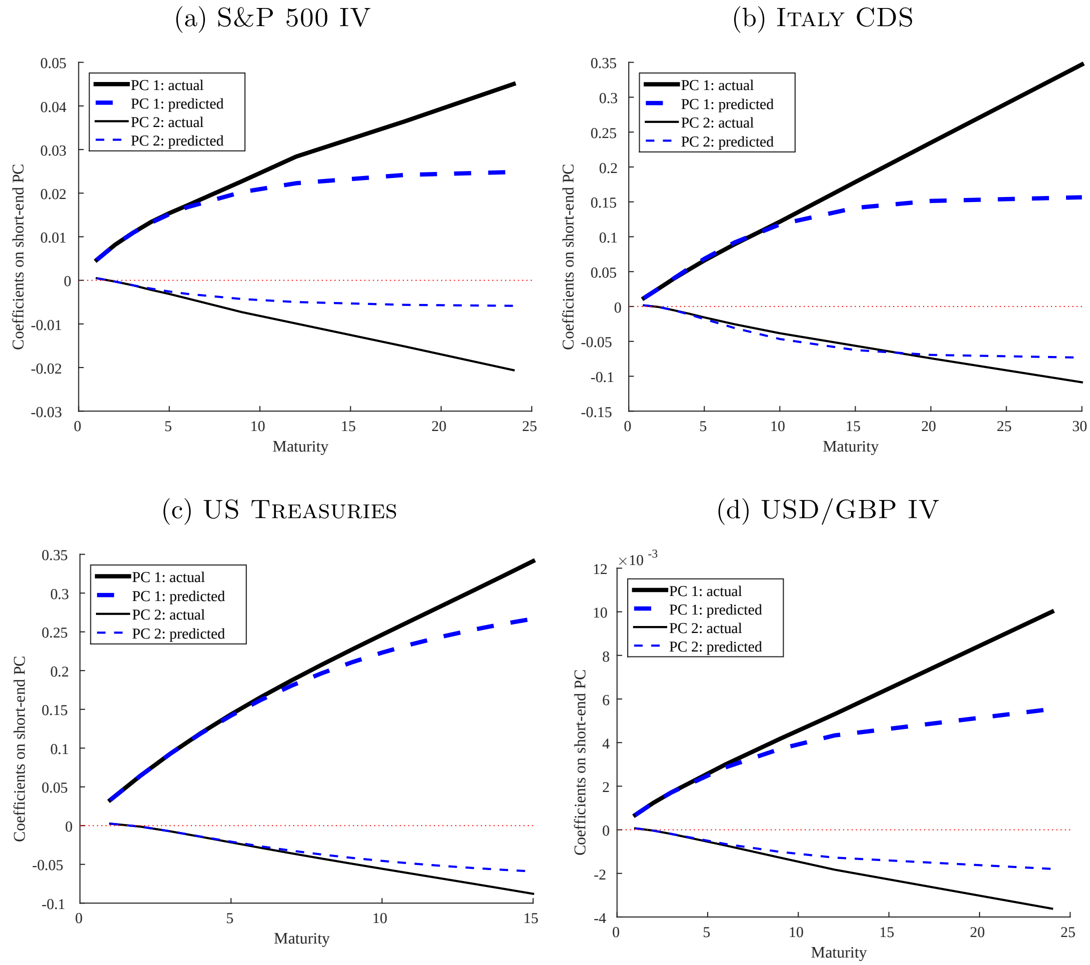
4.3.1 Inspecting the mechanism: regression coefficients

The variance of prices at each maturity (both empirically and under the restrictions of the model) is a function of the *regression loadings* of prices p_t^n onto the factors $\overline{p}_t$ and the variance-covariance matrix of the factors themselves, $V(\overline{p}_t)$. For example, we have

$$V^{restricted}(p_t^L) = b_L' R' V(\overline{p}_t) R b_L$$

where $b_L R$ is the vector of coefficients of the model-implied regression of p_t^L on $\overline{p}_t$. Figure plots the coefficients that link each price in the term structure, p_t^n to each factor of the short end $(\overline{p}_t)$, normalize by the standard deviation of each factor. The figure reports the actual regression coefficients (solid line) as well as the model-predicted ones (dashed line) for a subset of our asset classes. The empirical regression coefficients in all but one case are larger (in absolute value) than the model-predicted ones, which indicates that the

Figure 4: Predicted and Actual Factor Loadings



Notes. Each figures plots the predicted (dashed) and actual (solid) loadings on the principal components of prices of each maturity.

long-term price moves reacts more strongly than predicted by the model to changes in eahc factor. The excess volatility discussed above results from a combination of these overreactions (weighted by the variance-covariance matrix of the factors).

The figure therefore highlights how the true source of "excess volatility" showed in Figure 1 originates from overreaction of the price movements at the long end of the curve to movements in the factors extracted from the short end of the curve.

4.3.2 Dividend strips

For the case of dividend claims we perform a slightly different analysis, since we compare the term structure of dividend strips with the price of the entire stock market. We therefore report all the results in this section.

The term structure of log price-dividend ratios of dividend strips (pd_t^n, $n = 2...7$) has a strong factor structure. The first principal component explains 97.5% of the total variation, and the first two component explain 99.3%. We therefore employ a two-factor model. Given the limited number of maturities available, we extract the two factors using all available maturities.

A first, reduced-form analysis of the relative variances of the log-price-dividend ratio of the dividend strips and of the entire stock market shows an interesting result: for all $n > 1$, we have

$$V(pd_t^n) > V(pd_t)$$

and is essentially the same for $n = 1$. In other words, the claim to the infinite stream of dividends looks *less* volatile than any of the claims to dividends between years 3 and 7, and is as volatile as the claim to a two-year dividend strip. In addition, we can compare the variances of the component of pd_t and pd_t^n explained by the two factors we extract from the dividend curve. In this case, we find that the the variance of pd_t is strictly lower than that of pd_t^n at *all* maturities.

To better understand this (surprising) pattern, it is useful to think about the "duration" of the exposure to each of the factor in our model. In particular, consider any of the factors F_k. Given the diagonal structure of $\rho^{\mathcal{Q}}$, we can always represent the exposure of pd_t^n to that factor by looking of the k_{th} element of $\rho^{\mathcal{Q}}$, $\rho_k^{\mathcal{Q}}$. The loading of pd_t^n to factor k will be:

$$b_{1,k}^n = \sum_{j=1}^{n} (\rho_k^{\mathcal{Q}})^j = \frac{\rho_k^{\mathcal{Q}}(1 - (\rho_k^{\mathcal{Q}})^n)}{1 - \rho_k^{\mathcal{Q}}}$$

whereas the exposure of the infinite-maturity claim will be:

$$b_{1,k}^{\infty} = \frac{\rho_k^{\mathcal{Q}}}{1 - \theta\rho_k^{\mathcal{Q}}}$$

For each factor, we can think of the "duration" of the stock market relative to shocks to that factor as the maturity n of that dividend strip for which the stock market has the same exposure as that dividend strip: the n such that $b_{1,k}^n = b_{1,k}^{\infty}$. It is easy to show that n is equal to:

$$n = \frac{log(\rho) + log(1 - \theta) - log(1 - \theta\rho)}{log(\rho)}$$

For the two $\rho_k^{\mathcal{Q}}$ we estimate (both of them close to 0.3), we find $n \simeq 4$: the stock market should have a similar exposure to the two factors as the 4-year dividend strip. Instead, we find that empirically pd_t has a lower exposure to the first factor (lower than any of the dividend strips), and a higher exposure to the second factor. Given that it is the first factor that drives most of the variance, it is not surprising to find that in this market we find the variance of the infinite-horizon claim pd_t^n is much lower than what would have been predicted by the $\mathcal{Q}$-dynamics implied by the term structure of dividend strips (we confirm by bootstrap that this result is highly significant, with p-value below .0001).

We therefore document a significant violation of the law of iterated values when comparing the dividend strip curve and the infinite-claim. Interestingly, it points to a *lower* volatility of the stock market claim relative to the volatility of the corresponding-duration claim in the dividend strips market.

We believe there are two possible interpretation for this surprising result. First, it may reflect different factors driving the dividend market and the infinite claim, for exmaple frictions or tax effects (see Boguth et al.). If the two markets are not well integrated, and respond to different factors, it can look to the observer as a violation of the law of iterated expectations.

A second, maybe more interesting, possibility, is that given that the dividend strips market is actually a derivative market of the much-larger market for the stock market, what we are documenting here is overreaction and excess volatility in this market relative to the movements of the entire stock market. This may explain why dividend strips have been found to be extremely volatile (for example by Binsbergen et al.), though naturally is not enough to explain the declining term structure of Sharpe ratios across maturities documented by Binsbergen et al.

5 Robustness

Key to the methodology of the paper is extracting $\rho^{\mathcal{Q}}$ from the short end of the term structure and comparing the predicted and actual variance of prices of higher maturities. To do so, in our main analysis we extracted K factors from the short end of the curve, where K was chosen as the number of factors that explain 99% of the variation of the whole term structure, and the "short end" was defined to be the composed of the K shortest maturities available. In this section we consider several robustness tests. All the results are reported in the Appendix Figures.

First, we extract K factors by looking at more than K maturities at the short end of the curve. We follow the methodology presented in Section 2 to extract the K principal components out of the first $H > K$ maturities. In the appendix we consider $H = K + 1$ and $H = K + 2$ maturities. Doing so has two effects. On the one hand, it potentially reduces measurement error in extracting the factors from the short end of the curve. On the other hand, it increases the maturities used as a "short end" of the curve, in some cases getting very close to the "long end" of the curve. Expanding the maturities used H therefore makes our test less sharp, because it often uses much longer maturities. While in most cases doing so results in a lower estimate of excess volatility, for the vast majority of cases the results are still strong and statistically significant.

In a third robustness test, we set again $H = K$ as in the main text, but now require enough factors K to explain 99.9% of the variation in the term structure. This greatly increases the number of factors required, and again, using longer maturities, makes our test less sharp. Except in a couple of cases, all of our results still hold.

Finally, we report a version of our main results where we use asymptotic standard errors (instead of bootstrapping them).

6 Conclusion

We document excess volatility and a violation of the law of iterated values in a large cross-section of asset classes. Our test of excess volatility exploits the overidentification restrictions offered by observing a term structure of claims on the same cash flows.

We use the short end of the term structure to learn investor's implied dynamics of cash flows under the pricing measure Q. This gives us a model of expectations under Q at all maturities, which are linked by the law of iterated expectations and the implied dynamics of the factors driving the cash flows. We find that prices of long-maturity claims are dramatically more variable than justified by the behavior of short maturity claims. This excess volatility cannot be explained by time variation in discount rates, because that is already accounted for by the risk-neutral expectations we extract from the short end of the term structure. Our results therefore show that the excess volatility puzzle first highlighted by Shiller (1982) cannot be fully accounted for via rational variation in discount rates.

A Appendix

A.0.3 The Origins of $\mathcal{Q}$, Heteroskedasticity, and Other Considerations

Before moving to other asset classes, we take advantage of the well-researched interest rate setting to discuss a few considerations in more detail. We begin with a discussion of the pricing measure, $\mathcal{Q}$.

Up to now, we have assumed the absence of arbitrage, which ensures the existence of a measure $\mathcal{Q}$ under which prices are given by equation (**??**). $\mathcal{Q}$-measure event probabilities are distorted versions of the objective probability measure, $\mathcal{P}$, that is observed by the econometrician and that describes the evolution of cash flows. The distortion of $\mathcal{Q}$ has a natural economic interpretation. In particular, the $\mathcal{Q}$-probability of a given state of the world arises from scaling the $\mathcal{P}$-probability by consumers' marginal utility in that state. That is, $\mathcal{Q}$ over-emphasizes the probability of high marginal utility states and under-emphasize low marginal utility states relative to $\mathcal{P}$. By adjusting for marginal utility, the $\mathcal{Q}$-probabilities directly embed the valuation effects of investor risk preferences state-by-state.

We briefly consider how the statistical $\mathcal{P}$-dynamics of cash flows map to the $\mathcal{Q}$-measure dynamics that we infer from price variances. To do so, we specify investor preferences in the form of a log stochastic discount factor (SDF):

$$m_{t+1} = -x_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t' u_{t+1}$$

The log SDF describes how investors assess and discount future payoffs under $\mathcal{P}$, and allows us to describe bond prices as a $\mathcal{P}$-expectation that is equivalent to (**??**):

$$P_{t,n} = E^{\mathcal{P}}\left[\exp\left(\sum_{j=1}^{n} m_{t+j}\right)\right]. \tag{9}$$

The vector λ_t summarizes investor risk attitudes regarding the uncertain future shocks u_{t+j}. It directly maps into risk premia, and is thus referred to as the price of risk. If λ_t is zero, investors are risk-neutral, if it is constant then risk premia are constant, and if it varies then discount rates also vary. We parameterize risk prices as a function of the same factors governing cash flow dynamics

$$\lambda_t = \lambda + \Lambda F_t.$$

Suppose that the $\mathcal{P}$-dynamics for factors are also linear, though generally with different parameters:

$$F_t = c + \rho F_{t-1} + u_t \tag{10}$$

Then the fact that prices are the same under $\mathcal{P}$ and $\mathcal{Q}$ recovers the mapping between the two measures. Most importantly for tests of excess variance, we have

$$\rho_{\mathcal{Q}} = \rho - \Sigma\Lambda.$$

What is the point? Knowing the process for cash flows does not allow us to make specific predictions about price variance. Specifically, this is because price variation depends both on preferences (discount rates) and on cash flow dynamics. As we argued in the introduction, any degree of price variation is reconcilable with observed physical cash flow dynamics by allowing for the appropriate variation in discount rates. This shows exactly why it is important to do all modeling under $\mathcal{Q}$. It pins down variation arising from either cash flows *or* discount rates.

The thrust of our approach is to combine prices, which are representable as expectations under $\mathcal{Q}$, with a flexible model of $\mathcal{Q}$-dynamics. This allows us to describe price volatility throughout the term structure. s the form

NEXT: HETEROSKEDASTICITY. CAN BE HANDLED IN GAUSSIAN FRAMEWORK WITH $\Sigma_t\Sigma_t'$ and $\Sigma_t\lambda_t$ LINEAR IN FACTORS. CAN ALSO CONSIDER AFFINEQ STUFF

A.1 A note on heteroskedasticity

In this section we discuss the effect of heteroskedasticity on our results. To do that, we will distinguish between linear and log-linear models. In all of the models, we will keep assuming that

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

$$F_t = c^{\mathcal{Q}} + \rho^{\mathcal{Q}} F_{t-1} + u_t$$

and we will discuss which conditions we need to impose on the conditional distribution of u_t.

The easiest case is that of a linear model, in which we can write the price directly as a risk-neutral expectation of future values of x:

$$P_t^n = E_t^Q[\sum_{j=1}^{n} x_{t+j}]$$

In this class of models, it is easy to show that we will have

$$P_t^n = a + B_{n1}' F_t$$

with

$$B_{n1}^{'} = \sum_{j=0}^{n-1} (\rho^{\mathcal{Q}})^j$$

For linear models in levels, not only the conditional variance of u_t does not matter, but the entire conditional distribution of u_t is irrelevant in determining B_{n1}. Therefore, as long as we look at an asset class whose price can be expressed as (XXX), we will not have to worry about specifying the conditional distribution of the shocks u_t, as long as they ensure that the process is well behaved. For example when we talk about variance swaps, we require that the unspecified process for $V_t(u_{t+1})$ converges to 0 as the underlying state variable approaches zero, so that – at least in the continuous time limit – a negative variance never occurs.

The second case we consider in this paper is the case in which prices P_t^n are exponentially affine in x. Since among the asset classes we consider the ones that fall in this category are bonds and CDS spreads (which are modeled in a way very similar to bonds), for the rest of this section we will look at bond models alone. The variable x_t in these cases should be interpreted as $-r_t$ (for bonds), and $-R_t$ (for CDS). In the case of bonds,

$$P_t^n = E_t^Q[e^{-\sum_{j=0}^{n-1} r_{t+j}}]$$

$$r_t = \delta_0 + \delta_1' F_t$$

While the results that follow can be generalized to much more sophisticated models (see for example Creal and Wu (2014)), for ease of exposition we will work with a variant of a CIR model, in which the volatility of all shocks u_{t+1} scales with a volatility parameter proportional to the interest rate. In particular, suppose that the variance-covariance matrix of the Gaussian shocks u_{t+1} is:

$$V_t u_{t+1} = \overline{\Sigma} + \Sigma \sigma_t^2$$

where $\overline{\Sigma}$ and Σ are constant symmetric positive definite matrices (this ensures that the variance will be positive semidefinite no matter what value σ_t^2 takes), and $\sigma_t^2 = ar_t = \delta_1' F_t$, so that the conditional variance of all shocks is proportional to the interest rate r_t as in CIR models. Note that this also ensures that – in the continuous time limit – volatility cannot become negative.

In this setup, log prices can be expressed as:

$$p_t^n = -A_n - B_n F_t$$

where the coefficients B_n follow the recursion

$$B_n = \delta_1' + B_{n-1}\rho^Q - \frac{1}{2} B_{n-1}' \Sigma B_{n-1} a \delta_1'$$

When volatility is not time varying ($\Sigma = 0$), the recursion collapses to the usual case:

$$B_n = \delta_1' + B_{n-1}\rho^Q$$

Finally, note that since the sequence B_n is increasing with n (at least until the point where the last term dominates) the effect of volatility on prices is larger at longer maturities than at short maturities.

Two remarks are important to understand the effect of stochastic volatility on our results. First, the volatility adjustment term $-\frac{1}{2} B_{n-1}' \Sigma B_{n-1} a \delta_1'$ *lowers* the term B_n relative to the case in which there is no stochastic volatility, and, given that this adjustment is larger for longer maturities, it pushes B_n down more and more along the maturity structure. The reason why the adjustment for volatility decreases with maturity is that the term $B_{n-1}' \Sigma B_{n-1} a \delta_1'$ is positive: Σ is positive semidefinite (so that $B_{n-1}' \Sigma B_{n-1} \geq 0$), and $a > 0$ as empirically volatility of interest rates is higher when the interest rate is high. The consequence of the volatility adjustment is that the series B_n (which ultimately determines the predicted volatility of the long-end of the curve) grows more slowly than in the homoskedastic case, making the predicted volatility of long-term bond prices even lower than the homoskedastic model would predict. Therefore, ignoring the variance correction makes our results more conservative.

A second remark is that a simple calibration of this variance adjustment shows that this variance adjustment is actually quite small. To see it empirically, note that in this model we havewe have:

$$V_t(p_{t+1}^{n-1}) = B_{n-1}'\overline{\Sigma}B_{n-1} + B_{n-1}'\Sigma B_{n-1}(a\delta_1' F_t) = B_{n-1}'\overline{\Sigma}B_{n-1} + B_{n-1}'\Sigma B_{n-1} a r_t$$

In other words, the variance adjustment for B_n is precisely one half of the loading of the conditional variance of the log price of the bond with maturity $n-1$ onto the short-term rate, times δ_1'. We can then estimate this adjustment for each maturity by running the regression

$$(p_{t+1}^{n-1} - E[p_{t+1}^{n-1}])^2 = a + br_t + \epsilon_t$$

and estimating the variance adjustment then as:

$$\frac{1}{2}B_{n-1}'\Sigma B_{n-1} a\delta_1' =$$

and in particular:

$$V_t(p_{t+1}^2) = \delta_1'\overline{\Sigma}\delta_1 + \delta_1'\Sigma\delta_1 a r_t$$

The magnitude of the adjustment to B_n will precisely be one half of the coefficient of the regression of the variance of the n-th log price onto the real rate.

The following figure shows the case of the two-factor model for bonds estimated in the text. The figure shows clearly that the volatility effect is negligible at least at short maturities in our calibration. Most importantly, when it does have a nontrivial effect, it reduces the growth rate of B_n, which, in turn, makes the predicted long yields react *even less* to movements in the factors extracted from the short end of the curve. Therefore, the assumption of homoskedasticity is, for our purposes, highly conservative.

Finally, the precise same logic applies to CDS, in which all the formulas hold with R_t instead of r_t, and where we also have that empirically volatility is higher when the implied default spreads are higher ($a > 0$).

A.2 Interest Rates

We begin with a model of the US government bond market. This is a convenient starting point because it builds directly on the large interest rate term structure literature familiar to many researchers. Our development follows Hamilton and Wu (2012), who study the class of Gaussian affine term structure models developed by Vasicek (1977), Duffie and Kan (1996), Dai and Singleton (2002), and Duffee (2002), and studied by many others.

In the Guassian 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 shown in (**??**) with factor dynamics under the pricing measure described by (**??**). The price of a risk-free bond that pays \$1 after n periods is

$$P_{t,n} = E^{\mathcal{Q}}\left[\exp\left(-\sum_{j=1}^{n} x_{t+j}\right)\right]. \tag{11}$$

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

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

The factor loading depends only on the persistence of the factors:[6]

$$b_n = (I + \rho'_{\mathcal{Q}} + ... + (\rho'_{\mathcal{Q}})^{n-1})\delta_1. \tag{12}$$

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

This setting illustrates how our empirical analysis will proceed. If the risk-neutral factor persistence is known, then the price variance at any point in the term structure can be perfectly predicted. In Section XXX, we show that the dynamics of F_t $(K \times 1)$ can be inferred from the prices of claims with $K+1$ different maturities. If $N > K+1$ maturities are available, we can test for consistency between the price variance for a subset of $K+1$ maturities and the variance of the remaining $N-K-1$ maturities.

A.3 Variance Swaps and Related Derivatives

Derivatives, especially put and call options, are traded on virtually all major classes of assets. Typically, a given asset has options with a range of maturities. Option prices at different maturities crucially depend on the implied volatility at each maturity. By linking the term structure of volatilities to our framework of Section XXX, we can produce our tests of excess price volatility for claims to volatility at different horizon, extracted from

[6] As we discuss in Section XXX, the the payoff's factor loading δ_1 are normalized to equal the vector of ones.

option prices.

The simplest case of volatility derivatives to analyse is the market for variance swaps. Variance swaps are claims to the integrated variance (RV, for realized variance) between the time the contract is signed and maturity. All money is exchanged at maturity, so the pricing will have no risk-free rate adjustment.

The price of a variance swap with maturity n at time t can be written as the expectation under $\mathcal{Q}$ of realized variance $x_t = RV_t$ from t to $t + n$.

$$P_t^n = E_t^{\mathcal{Q}} \left[\sum_{j=1}^{n} RV_{t+j} \right]$$

where realized variance is the sum of squared log daily returns in each month. In contrast to the price model of Section XXX, the payout of a variance swap depends on the *level* of the RV_t process (not on an exponential function of it). In this case it is the price level, as opposed to the log price, that is linear in factors:

$$P_t^n = a_n + b_n' F_t \tag{13}$$

with factor loading b_n given by equation (**??**). An attractive feature of the simple payoff structure of variance swaps is that dependence of prices on factors, $b_n' F_t$, 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 (**??**).

A second and related point is that because variances are non-negative, the linear Gaussian model of equation (**??**) 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[7]

$$F_t = c^{\mathcal{Q}} + \rho^{\mathcal{Q}} F_{t-1} + \Sigma_{t-1} u_t$$

[7] 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 F_{t-1}. Because this is only a scale modification of u_t, the expected *level* payoff in is unaffected, hence equation (**??**) is also unaffected. Different versions of this model are applied by Ait-Sahalia et al. (), Egloff et al. () and Dew-Becker et al. (2015).

Our variance swap pricing framework is directly applicable to the study of excess volatility in options markets. A variance swap contract with maturity n can be synthesized as a portfolio of options with various strike prices and each with maturity n (Carr and Wu (2006)). The CBOE's well known VIX index is precisely this, a portfolio of put and call options on the S&P 500 index constructed to replicate the value of a one month S&P 500 variance swap. In particular, Carr and Wu (2006, 2009) show that

$$VIX^2_{n,t} \cong E^{\mathbb{Q}}\left[\sum_{j=1}^{n} RV_{t+j}\right]$$

and characterize the approximation error with respect the the price of the traded variance swap. This same principle can be applied to a replicate the price of a variance swap on any underlying security when the necessary call and put options are available, and for any maturity at which the options are available.

Constructing the synthetic variance swap price (VIX^2) requires integrating across (infinitely) many strikes of puts and calls. In many cases, it is hard to construct reliable measures of the term structure of the VIX due to the availability of strikes at different maturities. As an approximation, we will work here with at-the-money (ATM) implied volatility instead. The advantage of working with ATM implied variance is that it can be computed reliably for a large number of asset classes. In addition, it gets closer to the original setting of Stein (1989). Finally, we note that at least to a first-order approximation, ATM implied volatility corresponds in fact to the price of a volatility swap (the claim to realized volatility), as shown by Carr and Lee (2009).

A.4 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 XXX.

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 bond can be written as:

$$P_t^n = E_t^Q \left[exp(-\int_t^n R_s ds) \right]$$

with

$$R_s = r_s + \lambda_s L_s$$

where h_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. Note that under the assumption that both r_s and $\lambda_s L_s$ are linear in the factors, we will have:

$$p_t^n = log(P_t^n) = -ny_t^n = (a_r^n + a_{\lambda L}^n) + (b_r^n + b_{\lambda L}^n)F_t$$

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

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

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. While theoretically this relationship holds only for floating-rate bonds, in practice it holds approximately well for fixed-coupon bonds too.

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:

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

so that:

$$nz_t^n = nlog(1 + Z_t^n) \simeq ny_t^n - ny_{F,t}^n = a_{\lambda L}^n + b_{\lambda L}^n F_t$$

This representation allows 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.

A.5 Inflation swaps

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

$$(1+b_t)^{(T-t)}-1 = E_t^Q\left[\frac{I(T)}{I(t)}-1\right]$$

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

$$B_t^n = e^{b_t n} = E_t^Q\left[exp(\int_t^{t+n}\pi_s ds)\right]$$

Note that we have no risk-free rate adjustment since all payoffs occur at maturity $t+n$. Just as in the case of bonds, we will have that log prices $n\cdot b_t$ will be linear in the factors:

$$n\cdot b_t = a^n + b^n F_t$$

A.6 Currency swaps

Currency swaps are claims to the future exchange rate. In particular, if E_t is the exchange rate process, the price of a $n-$maturity currency swap is:

$$P_t^n = E_t^Q[E_{t+n}]$$

We can map these prices to our framework of Section XXX as follows. Consider the process for the log exchange rate innovations, $\Delta e_t = dlnE_t$. Then we can write:

$$P_t^n = E_t^Q[E_{t+n}] = E_t E_t^Q[exp(\int_t^{t+n}\Delta e_s ds)]$$

or:

$$\frac{P_t^n}{E_t} = E_t^Q[exp(\int_t^{t+n}\Delta e_s ds)]$$

The rescaled price therefore fits precisely in our framework. Therefore, we will work with the log prices $p_t^n = log(\frac{P_t^n}{E_t})$.

A.7 Dividend futures

We obtain the log price-dividend ratios for maturities 2 through 7 for dividend futures in the following way (following Binsbergen, Brandt and Koijen (2010)). In each time t, we linearly interpolate the "equity yields" $e_t^n = \frac{1}{n} ln(D_t/P_t^n)$ of fixed-calendar-maturity dividend futures contracts and obtain the yields of constant-maturity contracts, from which we obtain the log price-dividend ratios

$$pd_t^n = -n \cdot e_t^n$$

The log price-dividend ratio falls exactly in the exponential affine specification of Section 2 if $x_t = \Delta d_t - r_t$ is a liner function of the factors F_t.[8]

For the stock market, the formula in the text can be obtained by writing down the pricing equation:

$$P_t = E_t[e^{m_{t+1}}(D_{t+1} + P_{t+1})]$$

with

$$m_{t+1} = -r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t' u_{t+1}$$

and simplifying it making use of the Campbell-Shiller log-linearization of the price-dividend ratio:

$$1 = E_t[e^{m_{t+1}}(1 + \frac{D_{t+1}}{P_{t+1}})\frac{P_{t+1}/D_{t+1}}{P_t/D_t}\frac{D_{t+1}}{D_t}]$$

$$\simeq E_t[exp\{m_{t+1} + k + \theta pd_{t+1} - pd_t + \Delta d_{t+1}\}]$$

Assuming then

$$\Delta d_t = \delta_0 + \delta_1' F_t$$

$$r_t = a_0 + a_1' F_t$$

[8]The attentive reader will notice a mismatch between the timing of dividends and that of the risk-free discount rate: $x_{t+j} = \Delta d_{t+j} - r_{t+j-1}$. We ignore this one-period mismatch for simplicity and assume that x_{t+j} is a function of factors at time $t + j$.

one can show that pd_t is also affine in F_t:

$$pd_t = b_0 + b_1' F_t$$

and

$$b_1' = (-a_1' + \delta_1' \rho^Q)(I - \theta \rho^Q)^{-1}$$

Finally, ignoring the timing discrepancy between the risk-free rate and dividend growth, one can write

$$b_1' = \eta_1' \rho^Q (I - \theta \rho^Q)^{-1}$$

with $\eta_1' = a_1' + \delta_1'$.

B Testing

B.1 Asymptotic Distribution of Test Under Null

Now we will treat ρ as a K-vector describing the factor dynamics:

$$F_{t+1} = \text{diag}(\rho) F_t + u_{t+1}$$

Let x_t be the vector of K yields observed with no error, so that

$$x_t = (y_{n_{x1},t}, ..., y_{n_{xK},t})^\top = P^\top F_t$$

with P a matrix with row k equal to $\rho^\top$ raised element-wise to the n_{Xk} power.
Throughout, any power of ρ is assumed to be taken element-wise.
Estimation model:

$$\begin{aligned} y_{n_s,t} &= b_{n_s}^\top x_t + \varepsilon_{n_s,t} & (14) \\ y_{n_l,t} &= b_{n_l}^\top x_t + \varepsilon_{n_l,t} & (15) \end{aligned}$$

Null model:

$$\begin{aligned} y_{n_s,t} &= \left(P^{-1} \rho^{(n_s - n_x)}\right)^\top x_t + \varepsilon_{n_s,t} & (16) \\ y_{n_l,t} &= \left(P^{-1} \rho^{(n_l - n_x)}\right)^\top x_t + \varepsilon_{n_l,t} & (17) \end{aligned}$$

where I'm assuming that $n_l > n_s > n_x$.

Assumption B.1. $E[\varepsilon_{i,t}|x] = 0$.

$$Cov(\varepsilon_{1,t}, \varepsilon_{2,s}) = \begin{cases} \Sigma & \textit{if } s = t \\ 0_{2\times 2} & \textit{oth} \end{cases}$$

where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

Under the null,

$$\sqrt{T} \begin{pmatrix} \hat{b}_{n_s} - \left(P^{-1}\rho^{(n_s-n_x)}\right) \\ \hat{b}_{n_l} - \left(P^{-1}\rho^{(n_l-n_x)}\right) \end{pmatrix} \sim N(0, \Sigma \otimes V_x^{-1})$$

where $V_x = X^\top X/T$.

Our variance ratio is a non-linear hypothesis test regarding $(\hat{b}_{n_s}, \hat{b}_{n_l})$. Because we know the joint asymptotic distribution, we can perform such tests using the delta method. Suppose the nonlinear function of interest is $g(b_{n_s}, b_{n_l})$, then we have

$$\sqrt{T}\left[g\left(\hat{b}_{n_s}, \hat{b}_{n_l}\right) - g(b_{n_s}, b_{n_l})\right] \to N\left(0, \nabla g^\top \left[\Sigma \otimes V_x^{-1}\right] \nabla g\right). \tag{18}$$

In our case, the function we study depends on the variance of the projection of y_2 on x. We want to compare $V(y_2|x)$ based on the estimate $\hat{b}_{n_l}$ to $V(y_2|x)$ based on $\hat{b}_{n_s}$. Under the null model, the ratio of these variance is

$$g\left(\hat{b}_{n_s}, \hat{b}_{n_l}\right) = VR = \frac{\hat{b}_{n_l}^\top V_x \hat{b}_{n_l}}{f(\hat{b}_{n_s})^\top V_x f(\hat{b}_{n_s})}$$

where $f(b_{n_s})$ is the appropriate transformation of short coefficients to long coefficient under the null, and $f'(b_{n_s})$ is its derivative.

The gradient is

$$\nabla g = \left(-2VR\frac{V_x f f'}{f^\top V_x f}, \frac{2V_x b_{n_l}}{f^\top V_x f}\right)'$$

B.2 More General Error Dependence

Assumption **??** allows for dependence in "noise" in the cross section of yields. But it doesn't handle serial dependence, which there may be a lot of in these data.

To get a feel for where this matters. The general estimation of the variance of b_{n_s}

$$\begin{aligned} V(b_{n_s}) &= E\left[(X'X)^{-1}X'\varepsilon_{n_s}\varepsilon_{n_s}'X(X'X)^{-1}\right] \\ &= E\left[(X'X)^{-1}(\sum_{s,t} x_t x_s' \varepsilon_{n_s,t}\varepsilon_{n_s,s})X(X'X)^{-1}\right] \end{aligned}$$

Under Assumption **??**, we are calculating the variance of b_{n_s}, the sum terms are zero if $s \neq t$, which is why this reduces to $V(\varepsilon_{n_s,t})(X'X)^{-1}$.

Call that middle term $S = E\left[X'\varepsilon_{n_s}\varepsilon_{n_s}'X\right]$. Instead of calculating it as

$$\hat{S} = \Gamma_0 = \sum_t x_t x_t' \varepsilon_{n_s,t}^2$$

allow for lags:

$$\hat{S} = \Gamma_0 + \sum_{j=1}^{J}(\Gamma_j + \Gamma_j^\top)$$

where

$$\Gamma_j = \hat{S}\sum_{t=j+1}^{T} x_t x_{t-j}' \varepsilon_{n_s,t}\varepsilon_{n_s,t-j}$$

Now, note, this only gives you the (1,1) block of the variance matrix in equation (**??**). For the (2,2) block, you need to redo this using ε_{n_l}. For the (1,2) and (2,1) blocks, it comes from using both ε_{n_s} and ε_{n_l}. I.e. it is estimating

$$Cov(b_{n_s}, b_{n_l}) = E\left[(X'X)^{-1}X'\varepsilon_{n_s}\varepsilon_{n_l}'X(X'X)^{-1}\right]$$

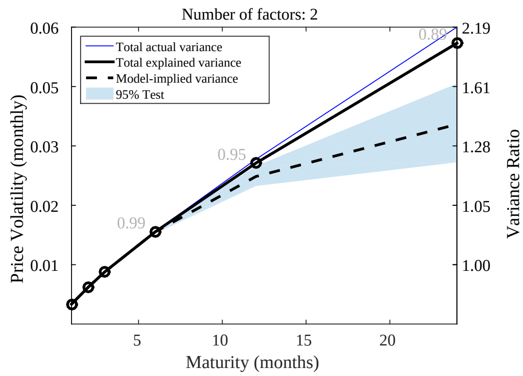
so keep this in mind when running the Newey-West.

C Appendix Figures

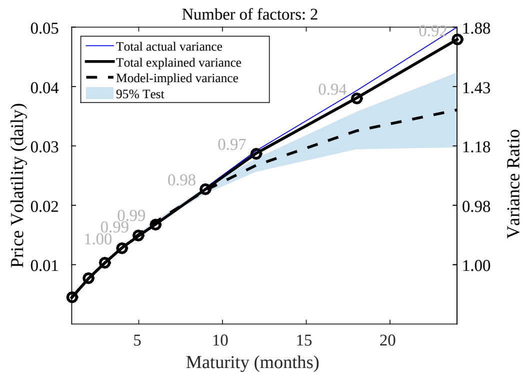
C.1 Robustness Test 1: Using $K+1$ yields to compute $\rho^{\mathcal{Q}}$

Figure 5: Robustness: using $K+1$ yields to compute $\rho^{\mathcal{Q}}$: 1. Index variances

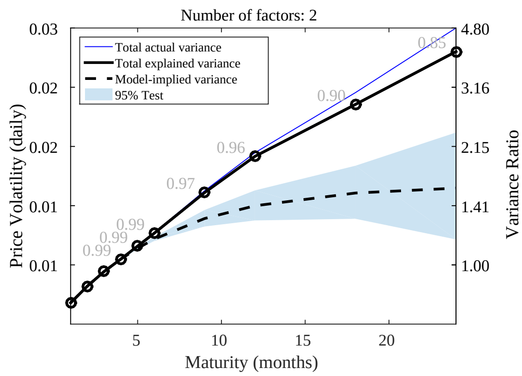
(a) Variance Swaps



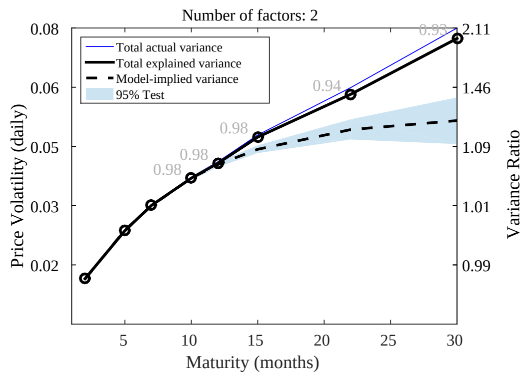
(b) S&P 500 IV



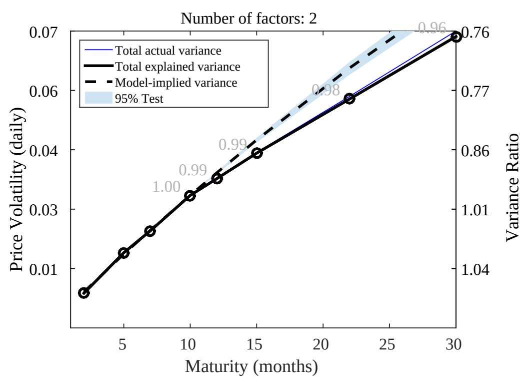
(c) Nasdaq IV



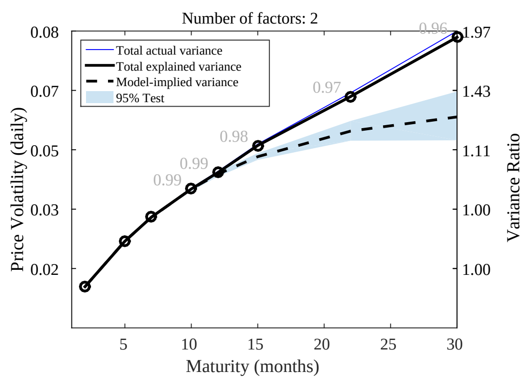
(d) Stoxx 50 IV



(e) FTSE 100 IV



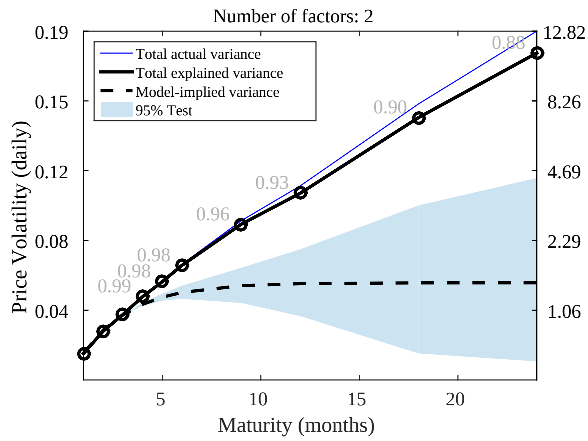
(f) DAX IV



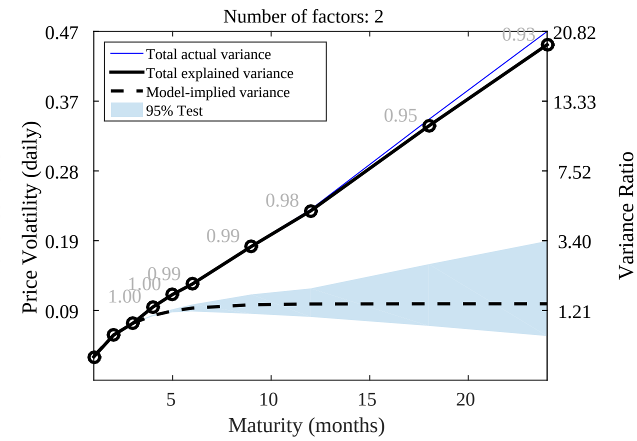
Notes. Same as in Figures 1-3, but the number of yields used to compute $\rho^{\mathcal{Q}}$ is chosen as $K+1$.

Figure 6: Robustness: using $K+1$ yields to compute $\rho^{\mathcal{Q}}$: 2. Other variances

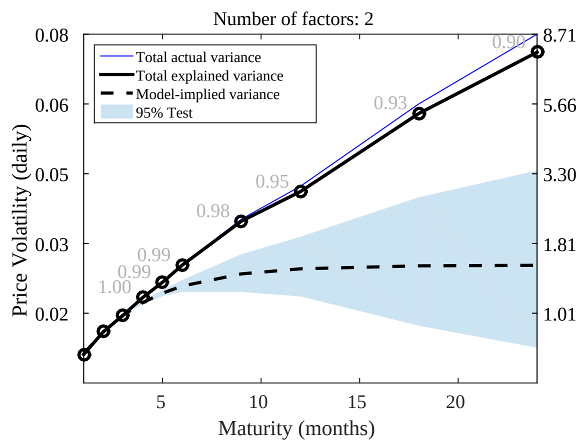
(a) Apple IV



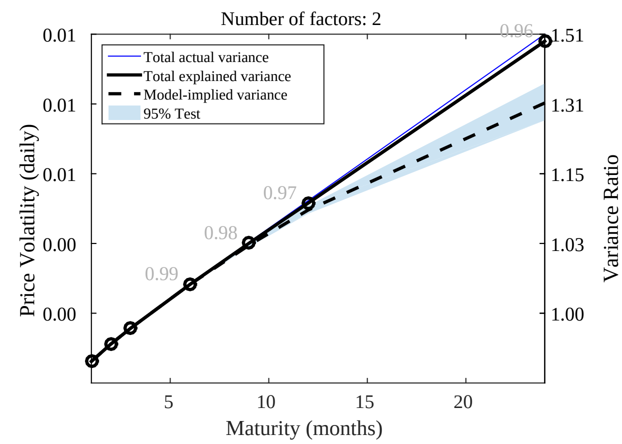
(b) Citigroup IV



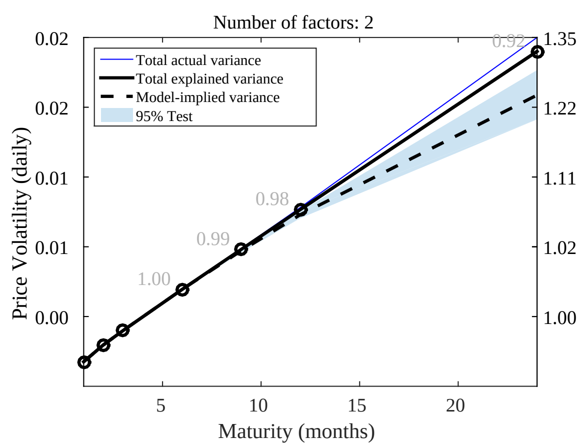
(c) IBM IV



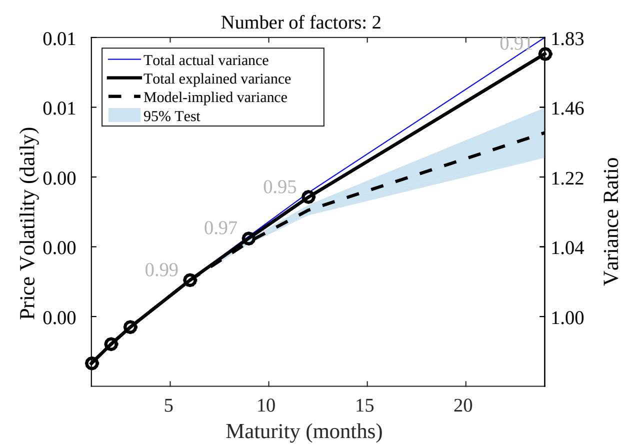
(d) GBP/USD IV



(e) GBP/JPY IV

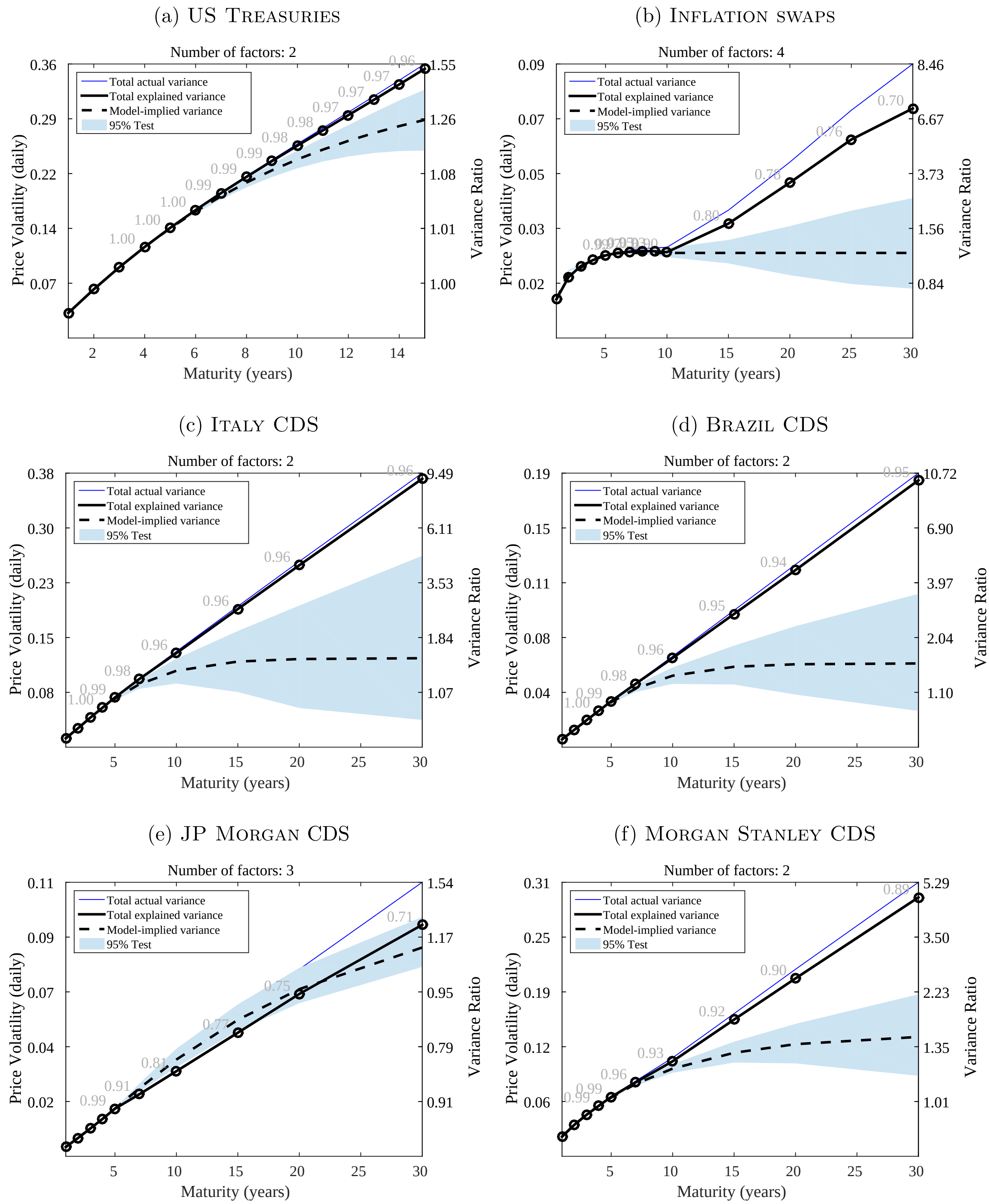


(f) USD/CHF IV



Notes. Same as in Figures 1-3, but the number of yields used to compute $\rho^{\mathcal{Q}}$ is chosen as $K+1$.

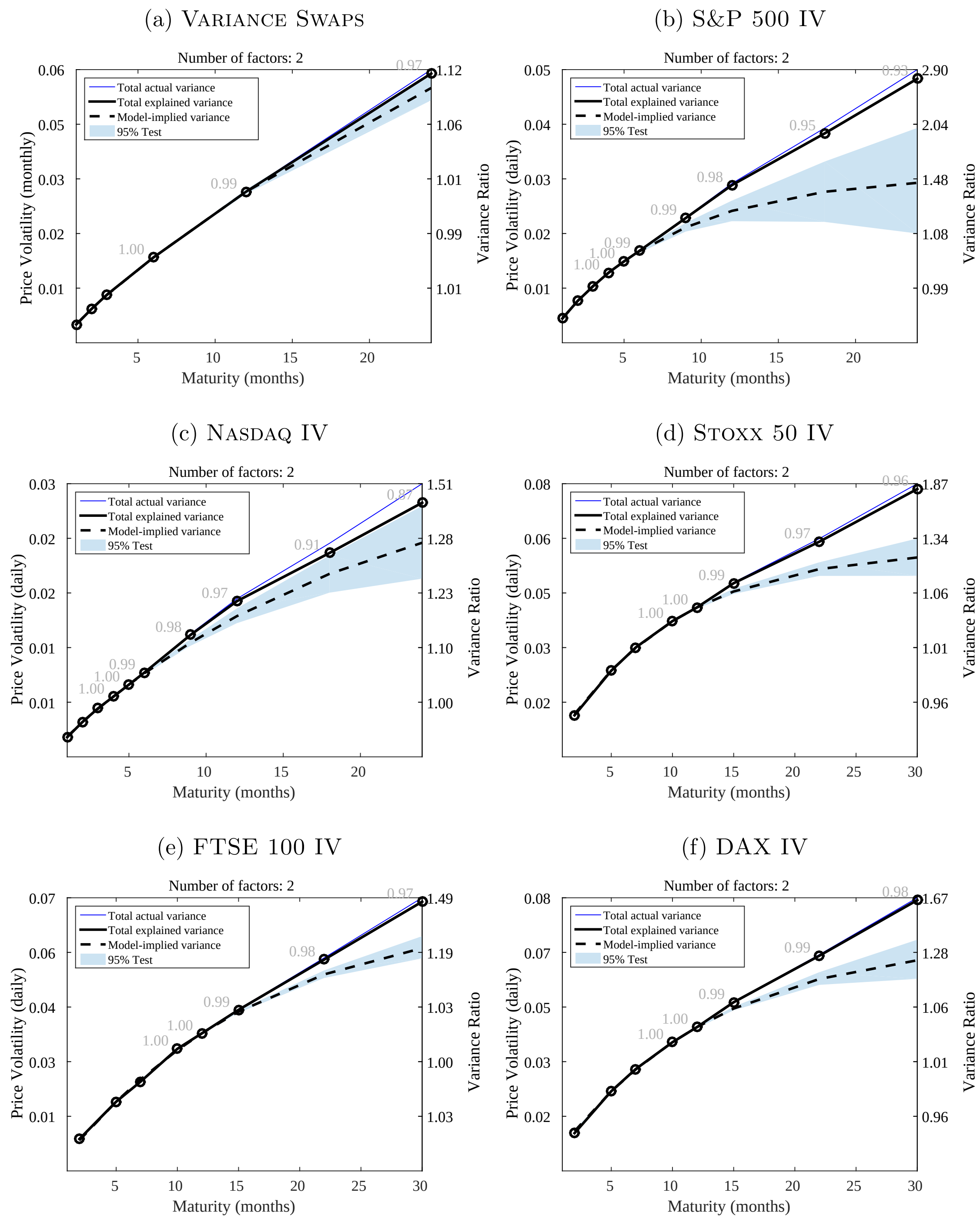
Figure 7: Robustness: using $K+1$ yields to compute $\rho^{\mathcal{Q}}$: 3. Other term structures



Notes. Same as in Figures 1-3, but the number of yields used to compute $\rho^{\mathcal{Q}}$ is chosen as $K+1$.

C.2 Robustness Test 2: Using $K+2$ yields to compute $\rho^{\mathcal{Q}}$

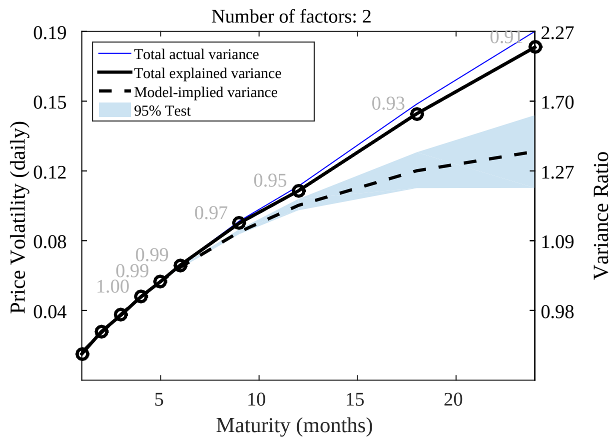
Figure 8: Robustness: using $K+2$ yields to compute $\rho^{\mathcal{Q}}$: 1. Index variances



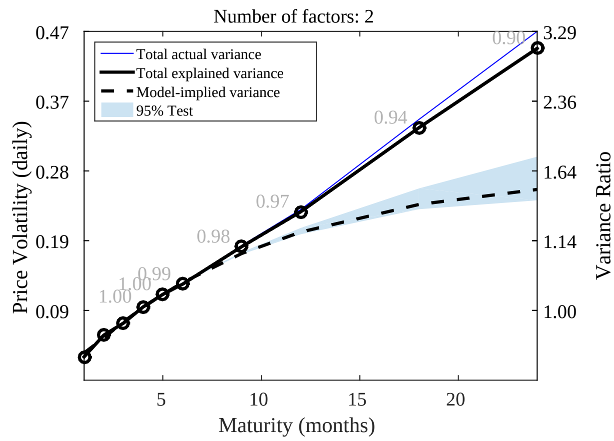
Notes. Same as in Figures 1-3, but the number of yields used to compute $\rho^{\mathcal{Q}}$ is chosen as $K+2$.

Figure 9: Robustness: using $K + 2$ yields to compute $\rho^{\mathcal{Q}}$: 2. Other variances

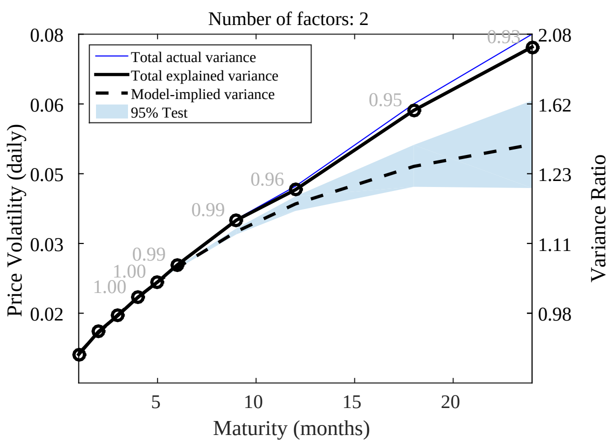
(a) Apple IV



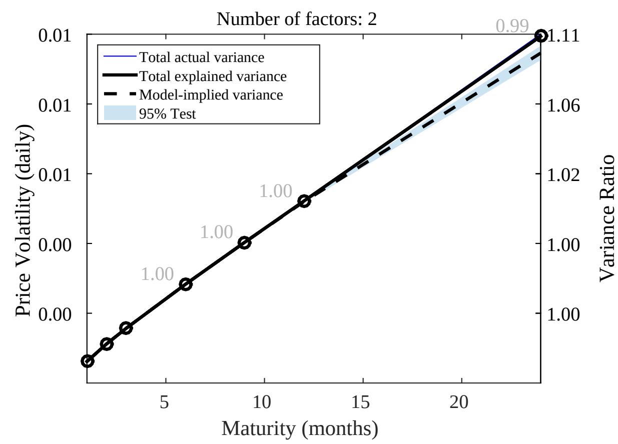
(b) Citigroup IV



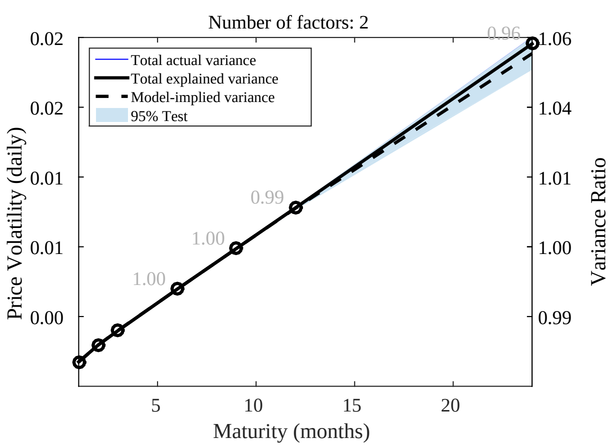
(c) IBM IV



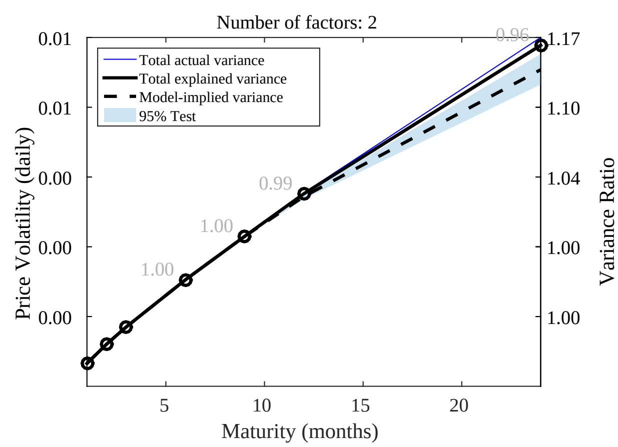
(d) GBP/USD IV



(e) GBP/JPY IV



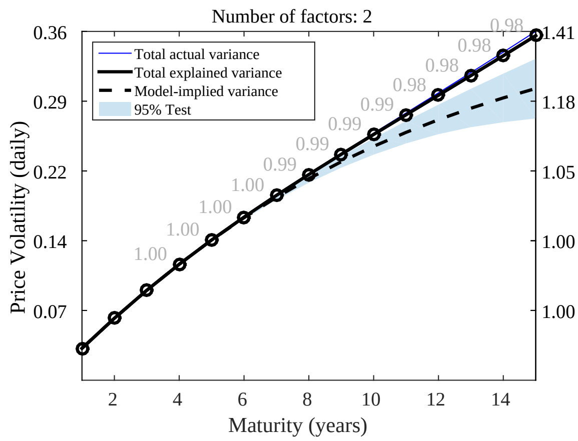
(f) USD/CHF IV



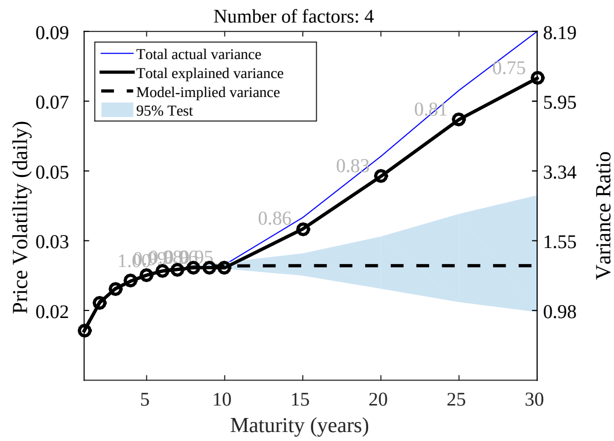
Notes. Same as in Figures 1-3, but the number of yields used to compute $\rho^{\mathcal{Q}}$ is chosen as $K + 2$.

Figure 10: Robustness: using $K+2$ yields to compute $\rho^{\mathcal{Q}}$: 3. Other term structures

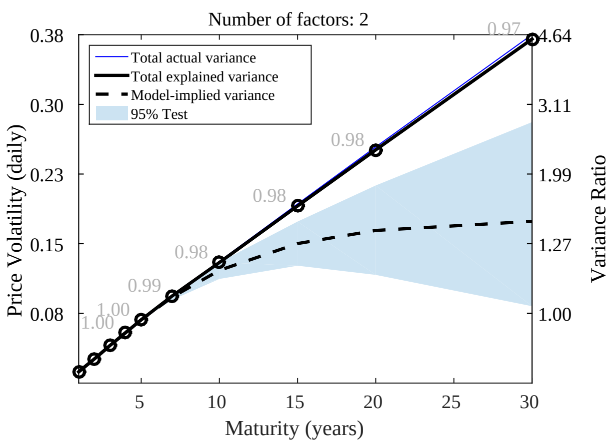
(a) US Treasuries



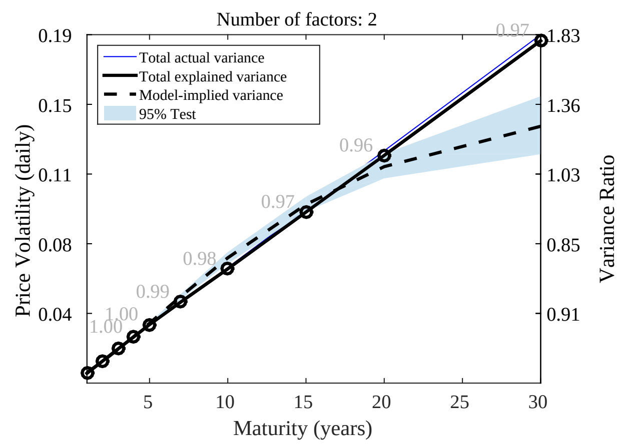
(b) Inflation swaps



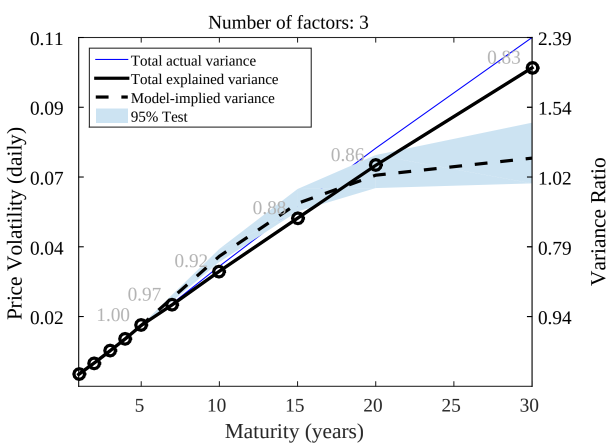
(c) Italy CDS



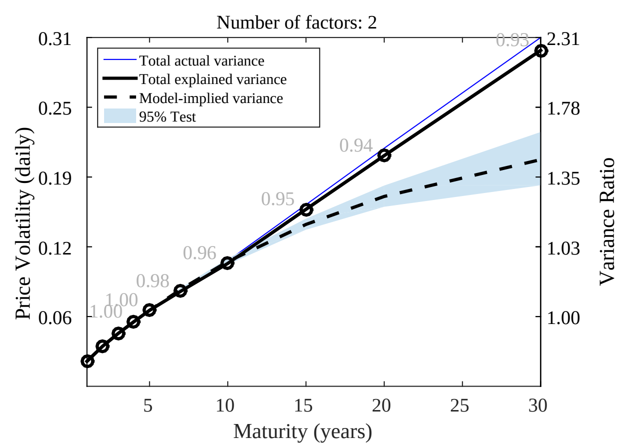
(d) Brazil CDS



(e) JP Morgan CDS



(f) Morgan Stanley CDS

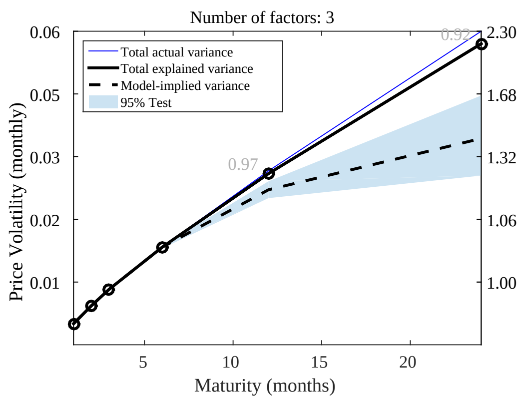


Notes. Same as in Figures 1-3, but the number of yields used to compute $\rho^{\mathcal{Q}}$ is chosen as $K+2$.

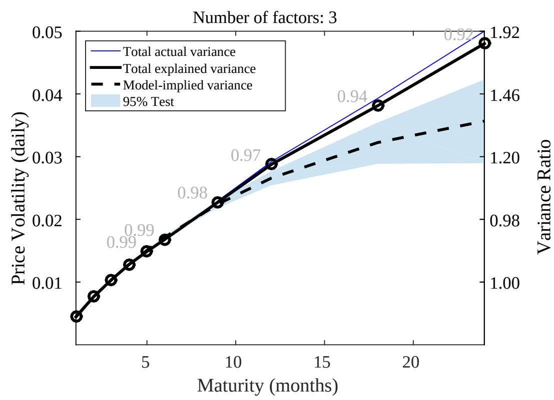
C.3 Robustness Test 3: Choose K to explain 99.9% of the variance

Figure 11: Robustness: explain 99.9% of the variance: 1. Index variances

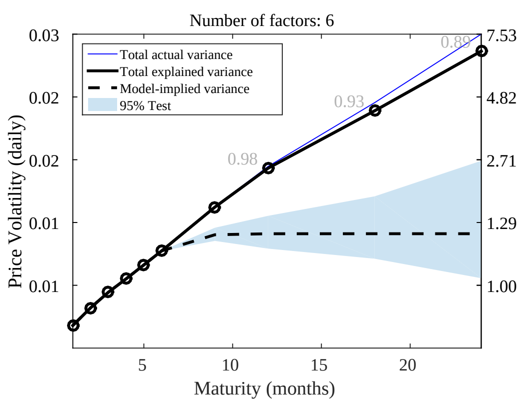
(a) Variance Swaps



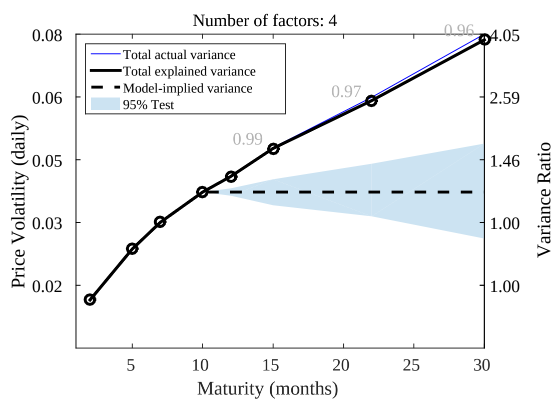
(b) S&P 500 IV



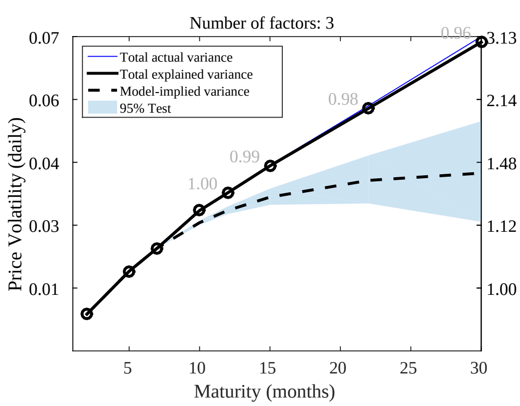
(c) Nasdaq IV



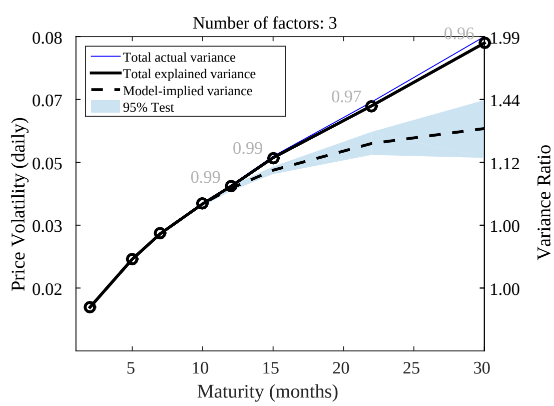
(d) Stoxx 50 IV



(e) FTSE 100 IV



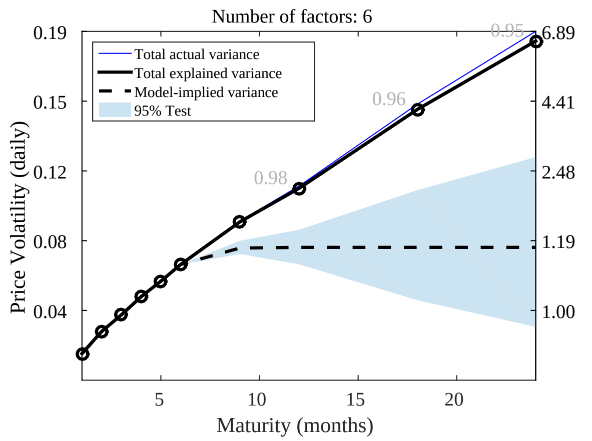
(f) DAX IV



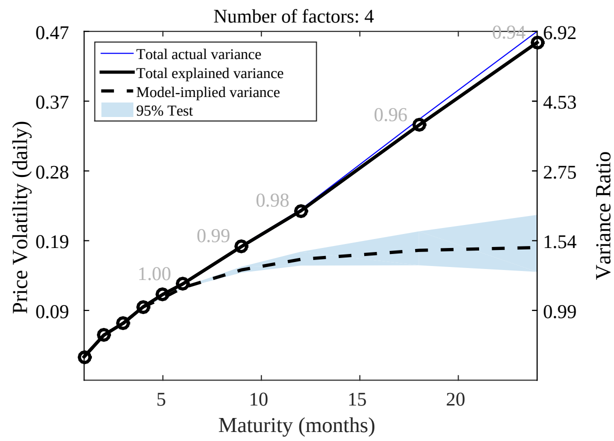
Notes. Same as in Figures 1-3, but the number of factors K is chosen to explain at least 99.9% of the total variation of prices.

Figure 12: Robustness: explain 99.9% of the variance: 2. Other variances

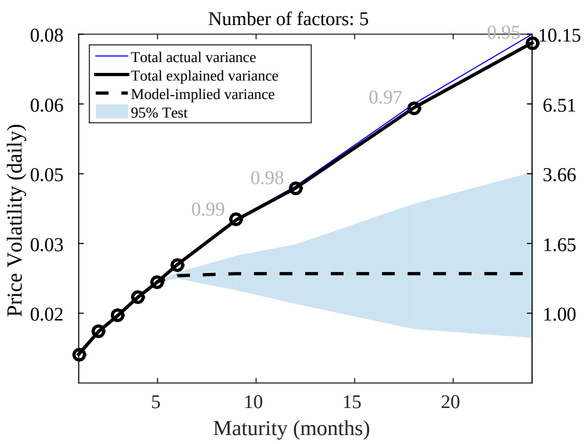
(a) Apple IV



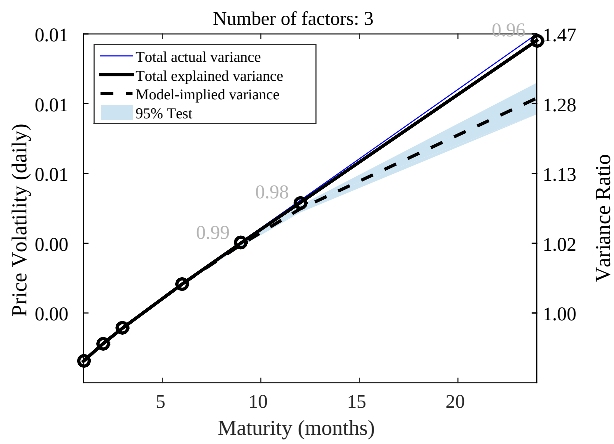
(b) Citigroup IV



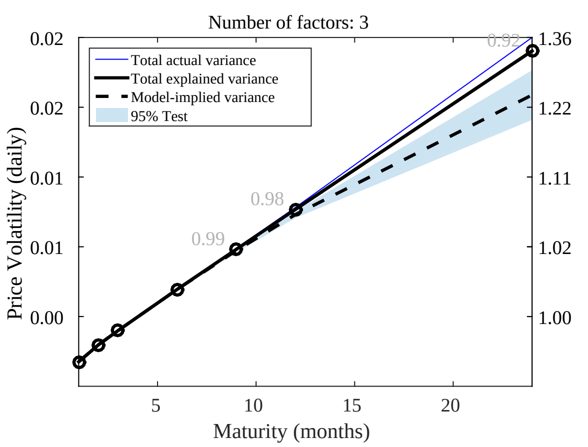
(c) IBM IV



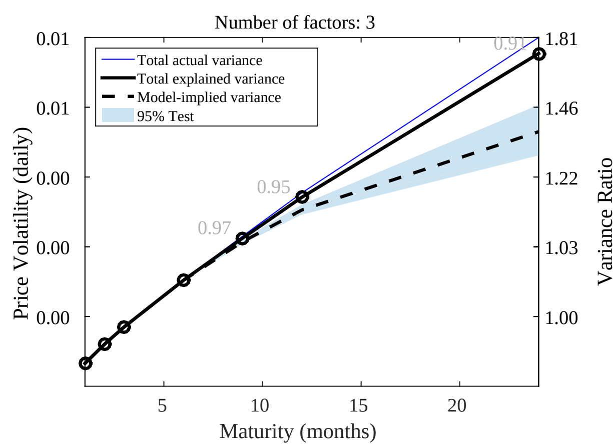
(d) GBP/USD IV



(e) GBP/JPY IV



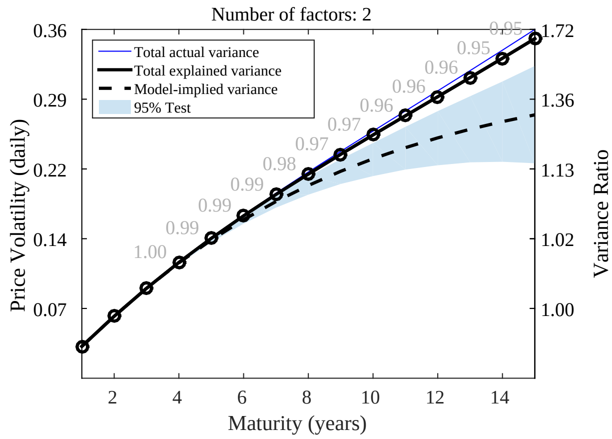
(f) USD/CHF IV



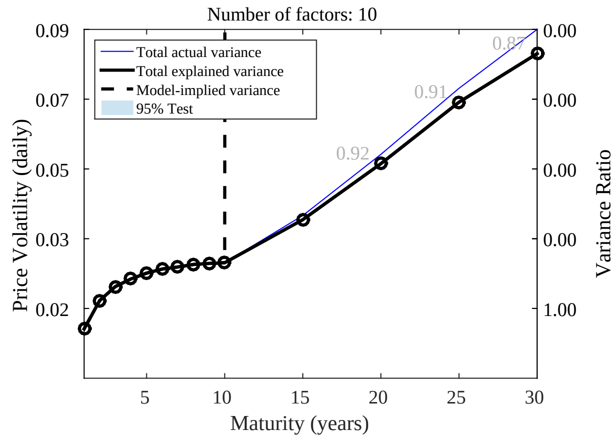
Notes. Same as in Figures 1-3, but the number of factors K is chosen to explain at least 99.9% of the total variation of prices.

Figure 13: Robustness: explain 99.9% of the variance: 3. Other term structures

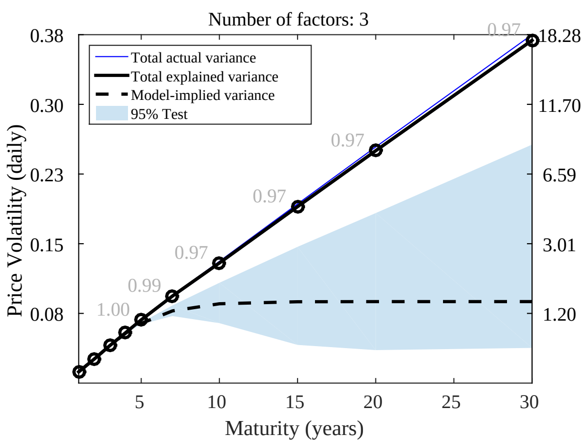
(a) US Treasuries



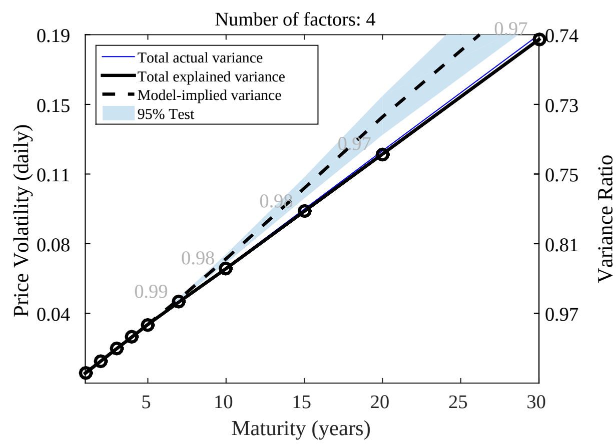
(b) Inflation swaps



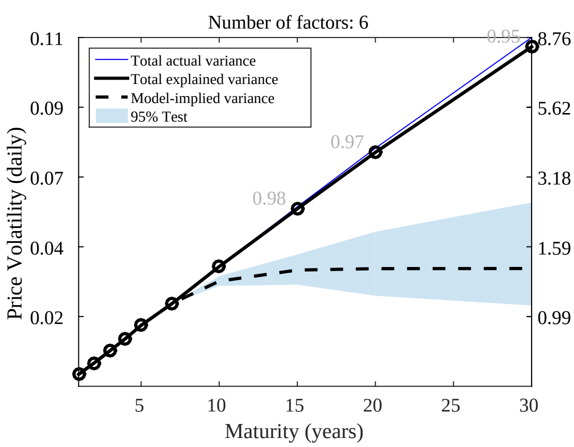
(c) Italy CDS



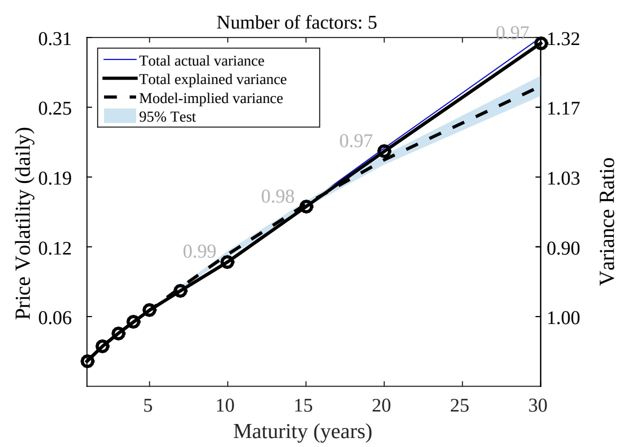
(d) Brazil CDS



(e) JP Morgan CDS



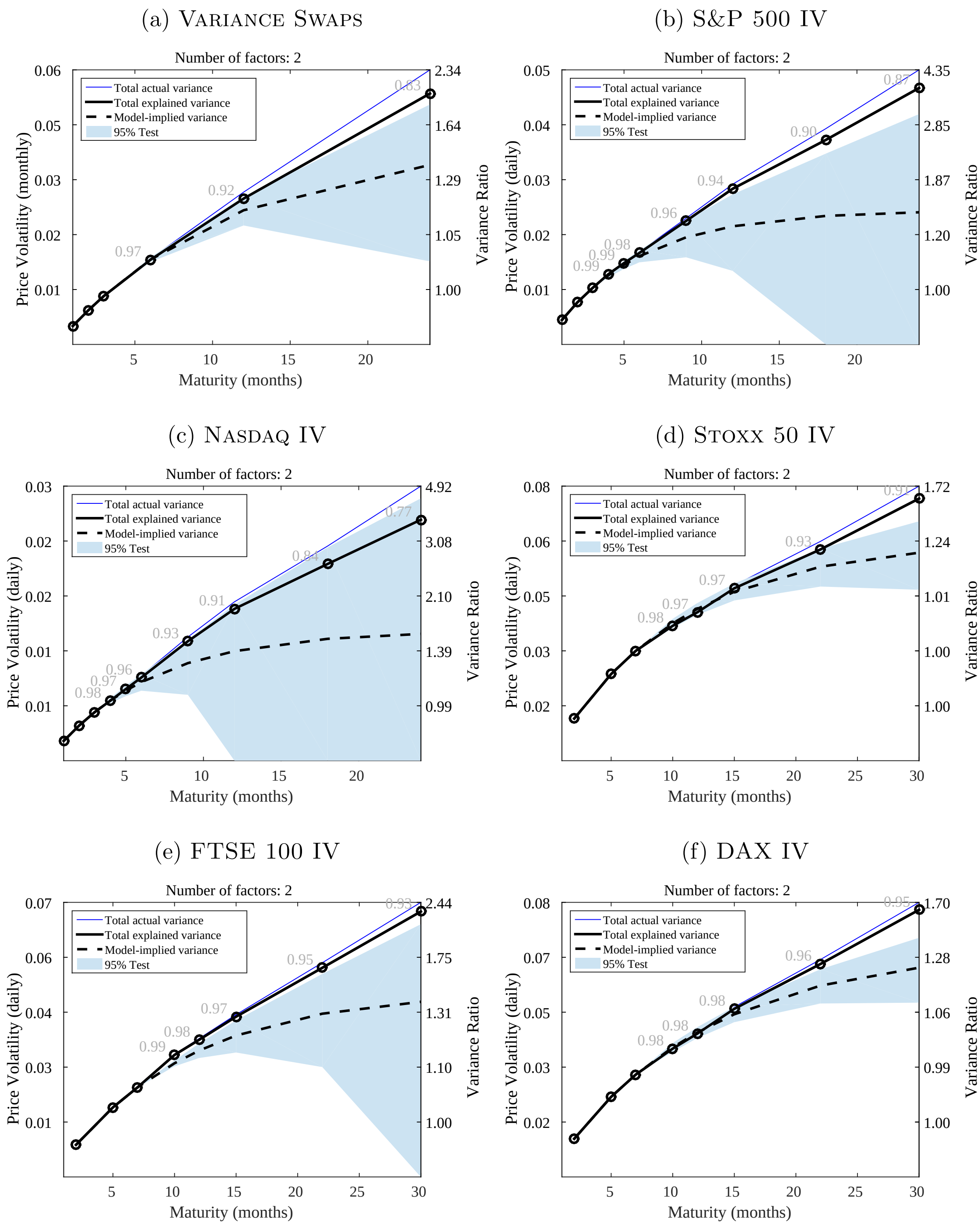
(f) Morgan Stanley CDS



Notes. Same as in Figures 1-3, but the number of factors K is chosen to explain at least 99.9% of the total variation of prices.

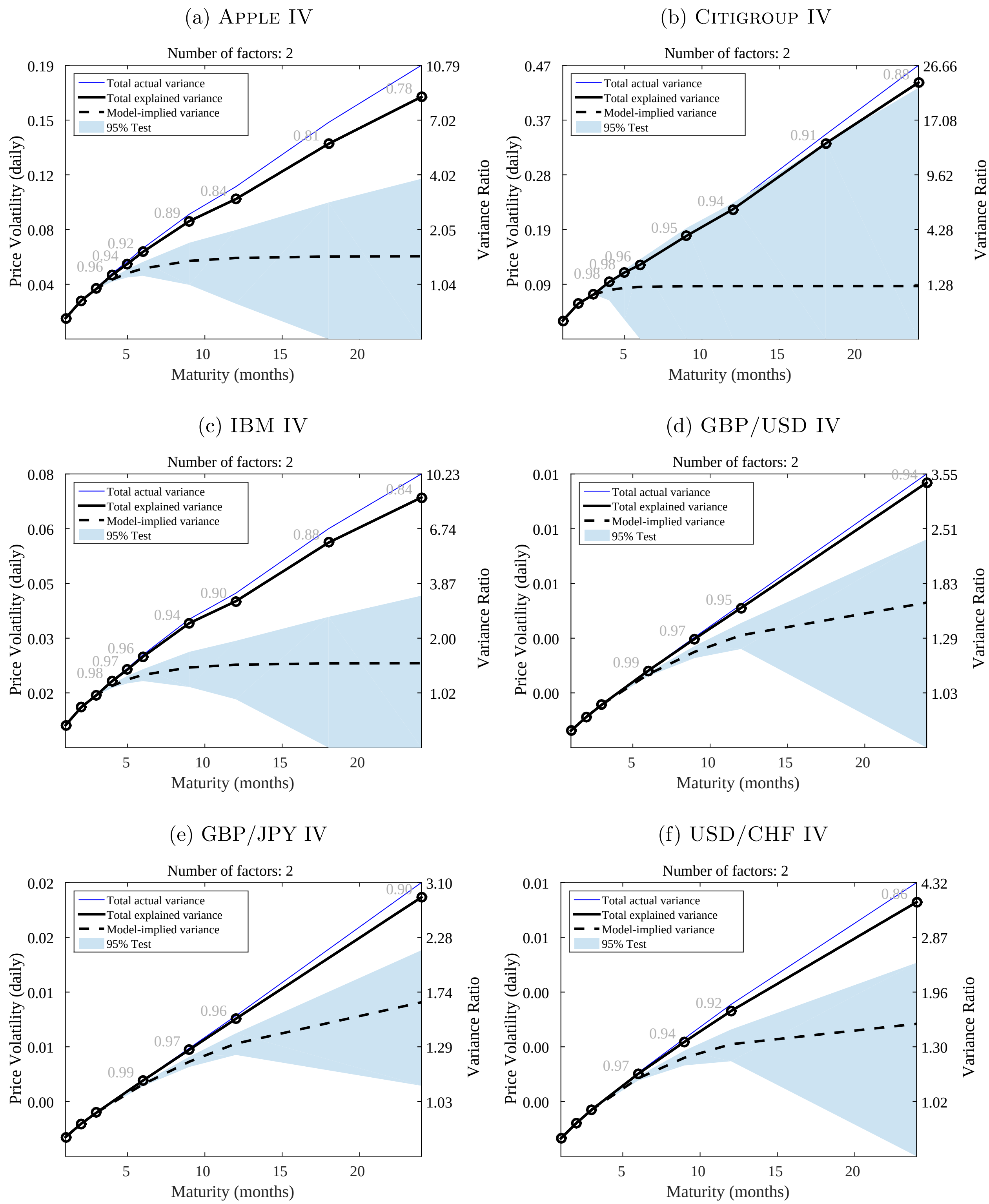
C.4 Robustness Test 4: Asymptotic Confidence Bands

Figure 14: Robustness: Asymptotic confidence bands: 1. Index variances



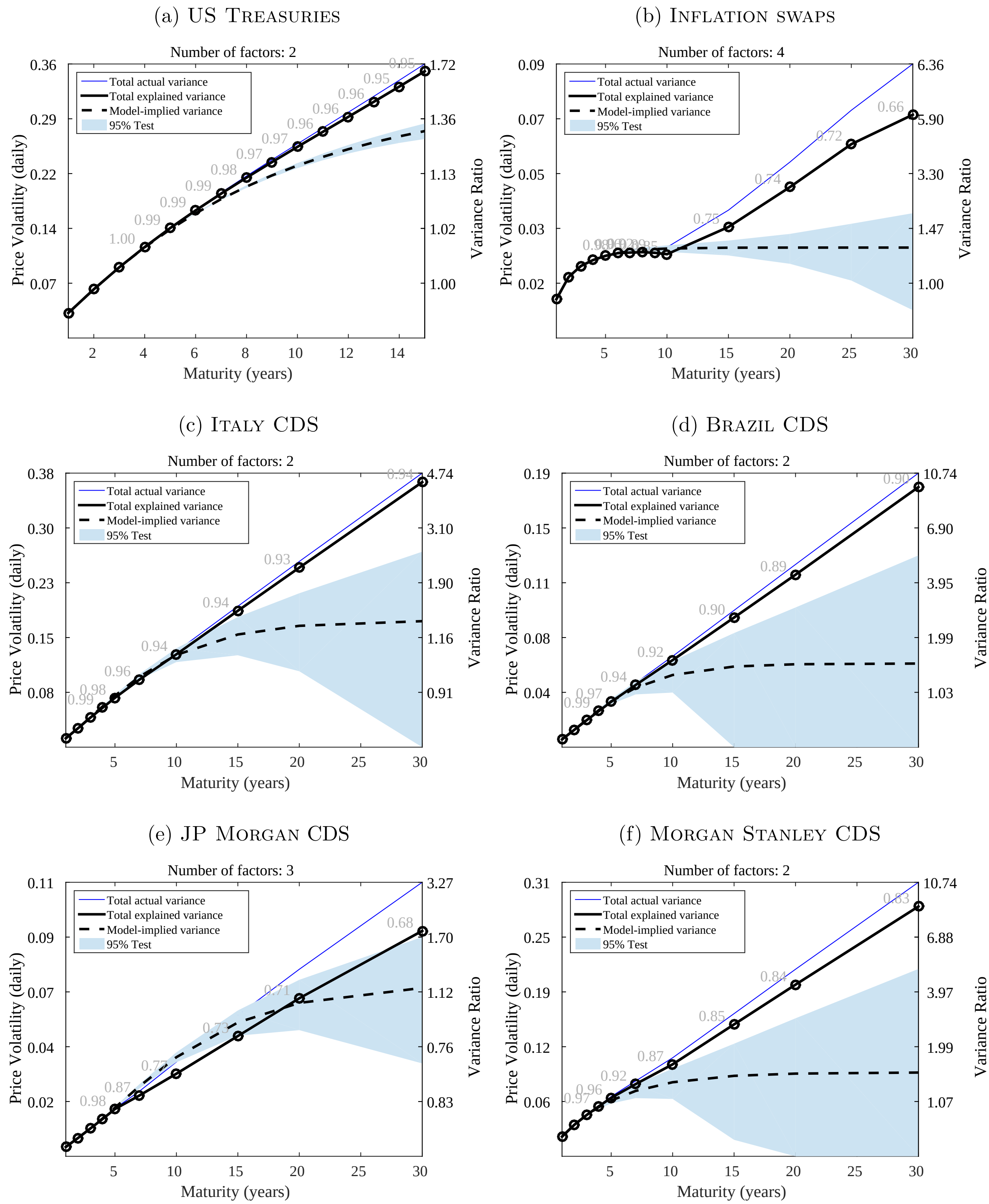
Notes. Same as in Figures 1-3, but the confidence bands are constructed using asymptotic standard errors.

Figure 15: Robustness: Asymptotic confidence bands: 2. Other variances



Notes. Same as in Figures 1-3, but the confidence bands are constructed using asymptotic standard errors.

Figure 16: Robustness: Asymptotic confidence bands: 3. Other term structures



Notes. Same as in Figures 1-3, but the confidence bands are constructed using asymptotic standard errors.