Duration Effects in Macro-Finance Models of the Term Structure

Thomas B. King*

Federal Reserve Bank of Chicago

October 15, 2019

Abstract

This paper studies a class of optimizing, no-arbitrage models in which the term structure of interest rates depends on the maturity structure of assets held by investors. The key assumption is that the stochastic discount factor is a function of the return on wealth. Portfolio choice matters for asset prices because it affects the distribution of this return. Such models are inherently nonlinear, and I propose a numerical algorithm for solving them. As an illustration, I solve and estimate a model in which investors price inflation and consumption risk in addition to wealth risk, with short-term rates are determined by a version of a Taylor rule. The equilibrium duration of investors' portfolio is treated as an unobserved factor. This factor is largely responsible for the nominal term premium and is correlated with the quantity of Treasury debt held by the public. Shocks to the factor that are roughly equivalent to the Federal Reserve's large-scale asset purchases reduce the ten-year nominal term premium by about 70 basis points on impact and lead to moderate increases in consumption.

---

*230 S. LaSalle Street, Chicago, Illinois, 60604. Phone: 312-322-5957. Email: thomas.king@chi.frb.org. This paper developed jointly with its companions, King (2015) and King (forthcoming). For helpful comments on this work, I thank Lena Mareen Boneva, Stefania D'Amico, Robin Greenwood, Sam Hanson, Don Kim, Philip Mueller, Dimirtri Vayanos, Andrea Vedolin, and seminar participants at FRB Chicago, the 2014 ECB/Bank of England workshop on Understanding the Yield Curve, the 2015 Federal Reserve "Day Ahead" Conference, the 2018 AFA meetings, the 2018 Banque de France Workshop on Monetary Policy and Asset Prices, the 2018 ECB Conference on Monetary Policy, and the 2019 CAMF Asset Pricing Workshop. Roger Fan and Zachry Wang provided excellent research assistance. The views expressed here do not reflect official positions of the Federal Reserve.

1 Introduction

Over the last decade, economists have made considerable progress in reconciling the behavior of the yield curve with standard consumption- and production-based asset pricing.[1] In these models, the term structure of interest rates is explained by investors' attitudes toward inflation and consumption risk, and by the rule that monetary policymakers use to determine the short-term interest rate.

At the same time that these models have been developed, several central banks have tried to shift the yield curve through purchases of long-term debt. Recent empirical work has been nearly universal in concluding that those purchases and other fluctuations in the structure of government liabilities have significant effects on the term structure of interest rates and, most likely, on other asset prices.[2] Perhaps the most commonly cited explanation for these results is that a reduction in the quantity of longer-term bonds that investors must hold leads them to require less compensation for bearing the remaining interest-rate risk in their portfolios; consequently, expected returns, term premiums, and yields on bonds fall. This phenomenon is sometimes known as the "duration channel" of government debt.

This type of mechanism is completely absent from the structural, consumption-based term-structure literature. In that literature, a shift in the quantity and distribution of government debt is either neutral or undefined. Indeed, partly on these grounds, several authors argue that the empirical evidence on the effects of asset purchases may reflect some other mechanism. A common refrain is that, in frictionless markets, asset quantities should be irrelevant for asset prices, a critique exemplified by Eggertsson and Woodford's (2003) proof that, in a particular class of general-equilibrium models, the structure of government debt available to the public makes no difference for either asset prices or macroeconomic outcomes.

Meanwhile, advocates of the duration channel frequently point to models such as Vayanos and Vila (2009) and Greenwood and Vayanos (2014) (collectively, "GVV") for theoretical support. In those models, shifts in the supply of long-term assets available to investors change their equilibrium exposures to interest-rate risk, causing fluctuations

[1] Among others, see Wachter (2006), Piazzesi and Schneider (2007), Van Binsbergen et al. (2012), and Rudebusch and Swanson (2012).

[2] See Bernanke et al. (2004), Kuttner (2006), Gagnon et al. (2010), Greenwood and Vayanos (2010, 2014), Krishnamurthy and Vissing-Jorgensen (2011, 2013), Meaning and Zhu (2011), Swanson (2011), D'Amico et al., (2012), Hamilton and Wu (2012), Joyce et al. (2011), Li and Wei (2012), Cahill et al. (2013), D'Amico and King (2013), Bauer and Rudebusch (2014), and Rogers et al. (2014), among others.

in term premia.[3] Although these models clearly contain elements that are appealing for those seeking to formalize and explore the link between bond supply and bond pricing, a number of difficulties have prevented their use for policy analysis or their incorporation into broader asset-pricing and macroeconomic models. In the GVV model, investors only hold Treasury bonds, inflation and consumption risk are not priced, and policy actions by the government are not explicitly modeled. And, despite these simplifications, the models are only analytically tractable in certain special cases. These limitations make it difficult to assess the economic importance of their central mechanism and to reconcile it with other macro-finance literature, including Eggertsson and Woodford's neutrality proposition.

This paper attempts to make some progress in incorporing the duration channel into standard macro-finance approaches to the term structure. First, I point out that the essential effect captured by the GVV model can be present in *any* no-arbitrage model in which the stochastic discount factor depends on the return on wealth. In such models, changes in the relative quantities of assets that are held by investors affect the distribution of the wealth return and therefore affect all asset prices. Preferences that result in pricing kernels that depend on the return on wealth have long been common in the finance literature, and, since Epstein-Zin-Weil utility has this property, are increasingly used in macroeconomic models as well. (The Eggertsson-Woodford model does not have it, which is why there are no such effects there.) The equilibrium duration channel that is possible under this specification is descended from the "portfolio balance" effects developed in papers such as Tobin (1968) and Frankel (1985), but, unlike those papers, the models considered here are arbitrage free, obey rational expectations, and do not require anything special about money or short-term debt.

Second, I provide a numerical method for solving such models in a wide variety of cases, iterating on a version of Tauchen and Hussey (1991). The solution is a nontrivial computational task because, even under simplifying assumptions about funcitonal forms, equilibrium asset prices involve a nonlinear recursion in multidimensional function space. (This is why Vayanos-Vila can only be solved analytically in limiting cases.) One advantage of the approach I propose is that it allows for arbitrary nonlinearities in the state vector and in the pricing kernel. A key nonlinearity is introduced by allowing investors to have relative, rather than absolute, risk aversion. Another is introduced

[3]These models have been extended and applied in various ways by Hamilton and Wu (2012), Altavilla et al. (2015), Greenwood et al. (2015), King (2015), Haddad and Sraer (2015), Malkhozov et al. (2016), and King (forthcoming), among others.

by imposing an effective lower bound (ELB) on the nominal short rate.

Finally, I apply the approach to a model with two observable macro factors and two latent factors. The short-term interest rate follows a version of the Taylor rule with an ELB constraint that is implemented using a "shadow rate" specification.[4] One of the latent factors in the model corresponds to low-frequency variation in the Taylor rule; the other governs the maturity structure of investors' assets. The nominal pricing kernel depends on the return on wealth, inflation, and consumption growth, with a functional form that nests Epstein-Zin. Because the SDF explicitly accounts for inflation, the model allows for the pricing of real as well as nominal bonds.

I estimate the model on data on nominal Treasury yields and macroeconomic data since 1971, as well as inflation-protected (TIPS) yields that become available in 2003, using nonlinear Bayesian filtering methods. The model fits the yield data well—far better than a comparable model that ignores the return on wealth. It generates a decomposition of the yield curve that is broadly in line with other models, including a downward drift since the early 1980s in expected inflation, the expected real short rate, and the nominal term premium. Although the price of inflation risk and the volatility of inflation are constant, the model exhibits a time-varying inflation risk premium through the nonlinear interaction of inflation with wealth. This premium broadly has the properties of other estimates of the inflation-risk premium in the literature, declining from a value of around 1.5% during the 1970s to around 0.5% by the end of the sample.

I estimate significant time variation in the effective duration of investors' exposures. In particular, the estimated duration factor rises sharply around 1980 and then fluctuates roughly with the business cycle over the next three decades. Although the level of this factor governs fluctuations in both real and nominal term premia, it does so in a nonlinear way that depends on the levels of all of the other variables in the model and, importantly, on the proximity of the ELB. In particular, term premia drift down over the sample, even though duration displays no secular trend since the early 1980s. In addition, the level of the duration factor is modestly correlated with measures of Treasury supply and duration. The estimates suggest that a four-month increase in duration (about a one-standard-deviation shock, on an annual basis, and about 1/5 of the combined effects of the Federal Reserve's asset purchase programs) results in

[4] See Kim and Singleton (2012), Krippner (2012), and Wu and Xia (2013). Bauer Rudebusch (2014) argue that shadow-rate models do a good job of capturing yield-curve dynamics near the ELB, greatly outperforming traditional affine models.

a contemporaneous increase of about 18 basis points in long-term nominal yields and about 22 basis points in long-term real yields, although these effects are smaller near the ELB. In addition, such shocks lead to modest decreases in consumption growth over the subsequent five years. Finally, I find that conventional monetary-policy shocks (unexpected increases in the short rate) cause significant increases in the duration factor and thus lead to increases in term premia. This is consistent, for example, with the presence of yield-oriented investors, as in Hanson and Stein (2015).

In addition to the studies mentioned above, this paper is related to several strands of the recent literature that have examined the effects of asset supply—especially as it relates to QE—through various lenses. One strand employs dynamic, stochastic equilibrium (DSGE) models to study the effects of QE on the macroeconomy. The stochastic discount factors that emerge in most DSGE models do not depend on the return on wealth and so cannot capture duration effects. (Indeed, most implementations of DSGE models involve linearizations that remove term premia altogether.) Nonetheless, some papers have introduced QE and other fluctuations in Treasury supply into DSGE models by introducing segmentation and limited participation between the short- and long-term bond markets (e.g., Andres et al., 2004; Chung et al.). This device is dificult to reconcile with institutional reality, and it does not capture the risk-based concept that underlies the duration channel in the finance literature.[5]

A few other papers have used Treasury-market data to discipline the parameters in versions of the GVV model, linking the exposures of investors in that model directly to observable measures of Treasury supply (Greenwood et al., 2014; Kaminska and Zina, forthcoming). Relatedly, though they do not impose the cross-equation restrictions implied by the GVV structure, Li and Wei (2014) also use observable measures of Treasury and MBS duration in an affine term-structure model. Proceeding in this way implicitly assumes an extreme form of market segmentationthe arbitrageurs are effectively assumed to hold all of the Treasury bonds outstanding and to face none of the tax liabilities associated with paying off those bonds. Indeed, they are assumed to have no other assets, liabilities, or cash flows at all. Instead, I treat asset exposures as a latent factor, which can produce both positive and negative asset exposures at different maturities, allowing the data to determine the relevant measures of wealth.

[5]Another strand of the macro literature focuses on the potential for QE to replace private intermediation by taking credit risk onto the central banks balance sheet (Gertler and Karadi, 2011; Woodford, 2012). While this channel may well be relevant for some types of asset-purchase programs, it is distinct from the duration channel. It also does not apply to most QE programs in the U.S., which involved only government-backed bonds.

In addition, as noted, the previous term-structure models in this literature abstract from macroeconomic factors and the nonlinearity associated with the ELB, elements that I explicitly incorporate.

Section 2 of the paper sets up the basic class of models considered here and discusses how they relate to those used in the previous literature. Section 3 describes the solution algorithm. Section 4 illustrates with some simple comparative-statics examples, with one- and two-factor models and the duration of investor portfolios held constant. Section 5 discusses the development and estimation of the four-factor macro-finance model. Section 6 presents the results of that model. Section 7 concludes the paper.

2 Asset Portfolios and Returns under No Arbitrage

I consider investors who, at each time t, have claims to a series of certainty-equivalent nominal payments over each of the following N periods. (I use the term "exposure" synonymously with "claim.") Each claim pays one dollar at maturity. I collect the quantities of the claims at each maturity in the vector $\mathbf{X}_t = (X_t^{(1)}, ..., X_t^{(N)})$. Claims at each maturity may take any value on the real line, with negative values denoting short positions. Without loss of generality, I allow the quantity of each claim held by investors to be determined by a vector of "duration factors" $mathbfz_t$.

The time-t (real) prices of the claims are denoted by $\mathbf{p}_t = (p_t^{(1)}, ..., p_t^{(N)})$. It is assumed that the prices and quantities of these claims are determined in equilibrium in each period to clear all asset markets. In cases in which the optimization problem faced by agents is specified, the demand and supply functions that give rise to this equilibrium can be solved explicitly. Here, I simply assume that the equilibrium quantities $\mathbf{X}_t$ follow a known reduced-form process, which may be a function of other variables in the economy.

The absence of equilibrium arbitrage opportunities is equivalent to the existence of a stochastic discount factor (SDF) $M_{t,t+n}$ that prices all assets in the economy. In particular, the real price of an arbitrary asset at time t is given by

$$p_t = \mathrm{E}_t \left[M_{t,t+n} q_{t+n} \right] \tag{1}$$

where q_{t+n} is the asset's payoff n periods hence, and E_t indicates the expectation conditioned on information at time t. This condition must hold for all horizons $n > 0$.

The following standard relationships follow immediately:

$$p_t^{(n)} = \mathrm{E}_t \left[M_{t,t+n} \right] \tag{2}$$

$$M_{t,t+n} = \prod_{i=1}^{n} M_{t,t+i} \tag{3}$$

Define the n-period zero-coupon real bond yields in the usual way, as

$$y_t^{(n)} = -\frac{1}{n} \log p_t^{(n)} \tag{4}$$

and define the real short rate $r_t \equiv y_t^{(0)}$. Equations (2) and (4) imply

$$r_t = -\log \mathrm{E}_t \left[M_{t,t+1} \right] \tag{5}$$

Given a real SDF, $M_{t,t+n}$, the nominal SDF is defined as

$$M_{t,t+n}^{\$} \equiv \Pi_{t+n} M_{t+n} \tag{6}$$

where Π_{t+n} is the gross rate of inflation between periods t and $t+n$. Nominal bond prices and yields, denoted $p_t^{\$(n)}$ and $y_t^{\$(n)}$, are given analogously to equations (2) and (4). The nominal short-term interest rate is denoted $i_t = y_t^{\$(0)}$.

I consider models in which the one-period stochastic discount factor takes the form

$$M_{t,t+1} = M(\xi_{t+1}, \xi_t, R_{t+1}) \tag{7}$$

where R_{t+1} is the one-period gross return on investors' wealth, $M(.)$ is a known function, and the vector ξ_t summarizes the time-t state of the economy. I restrict attention to cases in which investors do not care about the quantities of the particular securities that they hold *per se.* This rules out, for example, models with convenience yields, monetary services, or other special benefits that might attach to certain assets beyond their pecuniary returns. Formally, I assume $\mathbf{z}_t \notin \xi_t$. I collect the duration factors and economic factors into the state vector $\mathbf{s}_t = (\xi_t \quad \mathbf{z}_t)$. I assume that $\mathbf{s}_t$ follows a first-order Markov process on the support $\mathbf{S}$ with transition density $\tau(\mathbf{s}_{t+1}|\mathbf{s}_t)$.

The return on wealth is defined as

$$R_{t+1} \equiv \frac{\mathbf{X}_t'\mathbf{q}_{t+1}}{\mathbf{X}_t'\mathbf{p}_t} \tag{8}$$

where the payoff vector $\mathbf{q}_{t+s} = \begin{pmatrix} 1 & p_{t+1}^{(1)} & \dots & p_{t+1}^{(N-1)} \end{pmatrix}$. It is through (7) and (8) that asset quantities are related to asset prices. Fluctuations in the state of the economy that change the value of $\mathbf{X}_t$ will change $\mathbf{p}_t$ because expected returns—and therefore current prices—must adjust to make investors willing to hold the outstanding net positions at each point in time. For convenience, define $\mathbf{x}_t = \begin{pmatrix} x_t^{(1)} & \dots & x_t^{(N)} \end{pmatrix}$ as the vector of par value asset shares, i.e., $x_t^{(n)} \equiv X_t^{(n)} / \sum_{m=1}^{N} X_t^{(m)}$. Since the same vector $\mathbf{X}_t$ appears in both the numerator and denominator of (8), the dollar values of assets outstanding will not themselves be relevant for pricing in the class of models considered here, only their relative quantities will be. In particular, note that $R_{t+1} = \mathbf{x}_t'\mathbf{q}_{t+1}/\mathbf{x}_t'\mathbf{p}_t$.

Define the log real return on an n-maturity asset as

$$\rho_{t+1}^{(n)} = \log \frac{q_{t+1}^{(n-1)}}{p_t^{(n)}} \tag{9}$$

If one were willing to assume that $M(.)$ was exponentially affine and that $\mathbf{s}_t$ and R_t were jointly Gaussian, then expected returns could be written as

$$\mathrm{E}_t \left[\rho_{t+1}^{(n)} \right] = r_t + \mathrm{cov}[\boldsymbol{\lambda}'\xi_{t+1}, \rho_{t+1}^{(n)}] + \lambda^R \mathrm{cov}[R_{t+1}, \rho_{t+1}^{(n)}] + J^{(n)} \tag{10}$$

where $J^{(n)}$ is a term reflecting Jensen's inequality. However, while equation (10) is linear in $\log R_t$, it is not linear in the quantities $\mathbf{X}_t$. This means that it will generally not be possible to provide closed-form solutions for expected returns (or prices) as functions of portfolio quantities.

The GVV models mentioned in the introduction achieve an analytical solution by instead assuming constant *absolute* risk aversion. In particular, the expected log return on an n-maturity bond in those models (using the notation of this paper) is

$$\mathrm{E}_t \left[\rho_{t+1}^{(n)} \right] = r_t + \lambda^R \mathrm{cov}_t \left[\mathbf{X}_t'\mathbf{q}_{t+1}, \rho_{t+1}^{(n)} \right] + J_t^{(n)} \tag{11}$$

In continuous time, this equation gives each asset's expected excess return as proportional to the covariance of that asset's return with the *dollar value* change in wealth. Under additional functional-form assumptions about the process driving $\mathbf{z}_t$, GVV are able to obtain analytical solutions for bond prices using (11). However, it is clear that the CARA case in (11) can only be equivalent to the CRRA case in (10) if the market value of wealth $\mathbf{X}_t'\mathbf{p}_t$ is constant in all states of the world—a condition that will generally be violated if prices adjust to exogenous changes in exposures. Resolving this difficulty necessarily introduces nonlinearity into the CRRA model.

As the only model to incorporate supply effects into an affine representation of the term structure, GVV has been highly influential in the way that economists have designed and interpreted recent empirical studies.[6] However, the unusual assumption of constant absolute risk aversion that is needed to solve it has uncomfortable asset-pricing implications. For example, the model implies that term premia should generally trend upward with wealth, which runs counter to historical evidence. Moreover, it is not obvious how to incorporate additional features, such as inflation or the ELB, into the model while retaining tractability. The method proposed here, by solving the models numerically, overcomes these problems.

3 Solution Method

The central difficulty in solving models like the above—in which $M(.)$ is a funciton of R_{t+1} and R_{t+1} is determined endogenously—is that the solution for asset prices involves the moments of future prices, and, under rational expectations, the future prices themselves depend on the same fundamental process. While it is common in asset-pricing models for today's asset prices to depend on the distribution of tomorrow's asset prices, the particular difficulty here is that the SDF itself depends upon both of these objects.

I propose to solve these models numerically for the time-t vector of asset prices $\mathbf{p}_t$ using an iterative, discrete-state projection method. This approach has the added advantage that it places very few constraints on either the functional form of the pricing kernel or the dynamics of the state vector. Consequently, it is straightforward to consider models with potentially important nonlinearities, such as the ELB.

I first make explicit that prices and quantities depend on the state of the economy. Namely, let $X^{(n)}(\mathbf{z}_t)$ be the function that maps the duration factors into the quantity of asset n, and let $\Pi(\xi_t)$ be the function that maps the macroeconomic state vector into gross inflation. It is assumed that the form of the SDF in equation (7), the laws of motion for the states, and the dependence of quantities on the states

[6] See, for example, Hamilton and Wu (2011), and Li and Wei (2012).

are known—that is, we (and investors) have knowledge of the functions $\tau(\mathbf{s}_{t+1}|\mathbf{s}_t)$, $M(\xi_t, \xi_{t+1}, R_{t,t+1})$, $M^{\$}(\xi_t, \xi_{t+1}, R_{t+1})$, and $X^{(n)}(\mathbf{z}_t)$. We seek vector-valued functions $\mathbf{p}(\mathbf{s}_t) = \begin{pmatrix} p^1(\mathbf{s}_t) & \ldots & p^N(\mathbf{s}_t) \end{pmatrix}$ and $\mathbf{p}^{\$}(\mathbf{s}_t) = \begin{pmatrix} p^{\$(1)}(\mathbf{s}_t) & \ldots & p^{\$(N)}(\mathbf{s}_t) \end{pmatrix}$ that describe how all asset prices depend on $\mathbf{s}_t$ and $\mathbf{X}_t$.

With this notation, the nominal price of asset n is given by

$$p^{\$(n)}(\mathbf{s}_t) = \int_{\mathbf{S}} \tau(\mathbf{s}'|\mathbf{s}_t) M^{\$}(\xi_t, \xi', R_{t+1}) q^{\$(n)}(\mathbf{s}') d\mathbf{s}' \tag{12}$$

where the integral is taken over all dimensions of the state and $\mathbf{q}^{\$}(\mathbf{s}_t) = \begin{pmatrix} 1 & \mathrm{p}^{\$(1)}(\mathbf{s}_t) & \ldots & p^{\$(N-1)} \end{pmatrix}$ is the vector of functions determining nominal asset payoffs in state $\mathbf{s}_t$. An analagous relationship holds for real prices $\mathbf{p}(\mathbf{s}_t)$ with respect to the real SDF $M(\xi_t, \xi_{t+1}, R_{t+1})$ and real payoffs $\mathbf{q}(\mathbf{s}_t)$.

Given a distribution for the market return, R_{t+1}, (12) is a system of linear Fredholm equations of the first kind, which in principle can be discretized and solved by quadrature in one step (see Tauchen and Hussey, 1991). However, the fact that R_{t+1} is defined as in (8) requires us to iterate by, first, solving (12) using a given distribution of R_{t+1}, and, second, given the resulting pricing fuctions finding the updated distribution of R_{t+1}. These steps can be repeated to convergence.

Specifically, let $\mathbf{p}_d^{\$,k}(\mathbf{s}_t)$ be a proposal for the nominal pricing function on a discretization of the state space $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_G) \in \mathbf{S}^G$, where G is the number of nodes and $k = 0, \ldots, K$ indexes iterations, and let $\mathbf{q}_d^{\$,k}$ be the corresponding discretization of $\mathbf{q}^{\$,k}(\mathbf{s}_t)$. Suppose that the nodes are uniformly distributed over the state space, so that the conditional transition probability from node j to node h can be approximated by[7]

$$\widehat{\tau}(\mathbf{d}_h|\mathbf{d}_j) \equiv \tau(\mathbf{d}_h|\mathbf{d}_j) \left[\sum_{g=1}^{G} \tau(\mathbf{d}_g|\mathbf{d}_j) \right]^{-1} \tag{13}$$

The solution algorithm proceeds as follows.
Set the iterator $k = 0$.

1. Guess a function $\mathbf{p}^{\$,k}(.)$ such that $\mathbf{p}_t^{\$} = \mathbf{p}^{\$,k}(\mathbf{s}_t)$ on a discretization of $\mathbf{S}$. Find the corresponding values of $\mathbf{p}_d^k(\mathbf{d}_t)$ at each node in $\mathbf{D}$.

2. Based on this function, compute the real return on wealth between each pair of

[7]The uniform discretization is only for expositional ease and is not essential. Indeed, standard quadrature methods are likely to be more efficient.

nodes (j,g) as

$$R^k(d_j,d_g) = \frac{\mathbf{x}(\mathbf{d}_j)'\,\mathbf{q}_d^{\$,k}(\mathbf{d}_g)}{\mathbf{x}(\mathbf{d}_j)'\,\mathbf{p}_d^{\$,k}(\mathbf{d}_j)} - \Pi(\mathbf{d}_g) \tag{14}$$

3. Compute the updated nominal pricing function for the vector $\mathbf{p}_d^{\$,k+1}(\mathbf{d}_j)$ at each node $j = 1,...,G$ by setting

$$p_{\mathbf{d}_j}^{\$(0),k+1}(\mathbf{d}_j) = \exp[i(-\mathbf{d}_j)] \tag{15}$$

and

$$p_d^{\$(n),k+1}(\mathbf{d}_j) = \sum_{g=1}^{G} \widehat{\tau}(\mathbf{d}_g|\mathbf{d}_j) M^{\$}(\mathbf{d}_j,\mathbf{d}_g,R(d_j,d_g))\, p_d^{\$(n-1),k+1}(\mathbf{d}_g) \tag{16}$$

for $n = 1,...,N$. Set $\mathbf{q}_d^{\$,k+1}(\mathbf{d}_g) = \begin{pmatrix} 1 & p_d^{\$(1),k+1}(\mathbf{d}_g) & ... & p_d^{\$(N-1),k+1}(\mathbf{d}_g) \end{pmatrix}$.

4. Set $k = k+1$ and return to step 2.

This procedure constitutes a contraction mapping on $\mathbf{D}$ so long as the moments of the pricing kernel are well behaved. The Banach Theorem then guarantees for any given discretization $\mathbf{D}$, $\mathbf{p}_d^{\$,k}(\mathbf{d}_j) \longrightarrow \mathbf{p}_d^{\$}(\mathbf{d}_j)\ \forall\ \mathbf{d}_j \in \mathbf{D}$, where $\mathbf{p}_d^{\$}$ is the (unique) nominal pricing function that obtains if $\widehat{\tau}$ is the data-generating process. But continuity of τ ensures that, for any node j,

$$\lim_{G\longrightarrow\infty} p_d^{\$(n)}(\mathbf{d}_j) = \mathrm{E}_t\left[p^{\$(n-1)}(\mathbf{s}_{t+1})\, M^{\$}\left(\mathbf{d}_j,\xi_{t+1},\frac{\mathbf{x}(\mathbf{d}_j)'\,\mathbf{q}_d^{\$,k}(\mathbf{s}_{t+1})}{\mathbf{x}(\mathbf{d}_j)'\,\mathbf{p}_d^{\$,k}(\mathbf{d}_j)} - \Pi_{t+1}\right)\right] \tag{17}$$

i.e., in the limit, the pricing function solves the no-arbitrage condition (2). Since Π_t is a known function of the state, this argument also guarantees that the algorithm finds the unique real SDF.

Finally, by construction, if the algorithm converges, any point of convergence is a rational-expectations equilibrium. This follows immediately, since convergence is *defined as* the fixed point at which the joint distribution of $\mathbf{p}_{t+1}^{\$}$ and $M_{t+1}^{\$}$ is consistent with the vector $\mathbf{p}_t^{\$}$, for each point in the state space.

It is important to note that, although the algorithm only solves for the vector of prices at G points in the state space, once these solutions are in hand it is straightforward to calculate equilibrium prices at *any* point through the Nystrom extension. In particular, take an arbitrary state value $\mathbf{s}_t$. For G large enough, we have

$$p^{\$(n)}\left(\mathbf{s}_t\right) \approx \left[\sum_{g=1}^{G} \tau(\mathbf{d}_g|\mathbf{s}_t) M^{\$}\left(\xi_t, \mathbf{d}_g, \frac{\mathbf{x}\left(\mathbf{s}_t\right)' \mathbf{q}_d^{\$,k}\left(\mathbf{d}_g\right)}{\mathbf{x}\left(\mathbf{s}_t\right)' \mathbf{p}_d^{\$,k}\left(\mathbf{s}_t\right)} - \Pi\left(\mathbf{d}_g\right)\right) p_{n-1}^{\$}(\mathbf{d}_g)\right]\left[\sum_{g=1}^{G} \tau(\mathbf{d}_g|\mathbf{s}_t)\right]^{-1} \tag{18}$$

and similarly for real prices. Once the algorithm has converged, the quantities on the right-hand side are all known. Thus, real and nominal claims can be priced in at any point in $\mathbf{S}$.

Figure 1 displays some results on the convergence of the solution algorithm for the one-factor model discussed in the next section. The only state variable in that model is the "shadow" short rate i_t^*. The top panel shows the computed 2-, 5-, 10-, and 15-year yields, shown for i_t^* at its average value of 5.2%, across the first 30 iterations ($k = 1, \ldots\ 30$). The algorithm is initialized at a price vector $\mathbf{p}_d^0\left(\mathbf{d}_j\right) = (0.95, \ldots, 0.95)$ for all values of $\mathbf{d}_j$ and uses $G = 8$ nodes distributed uniformly across the range $i^* = (-0.05, 0.15)$. It is evident from this figure that, for each maturity n, the solution converges very quickly once $k > n$.

The middle panel shows convergence in the number of gridpoints by displaying the computed yield curve (after $k = 30$ iterations), again using $i_t^* = 0.052$ for illustration. Yield curves are shown for $G = 2$, 4, 8, and 16, in each case spaced equally across possible values of the state variable. While 4 nodes is clearly too few to achieve convergence, the solutions using 8 or more nodes are indistinguishable from each other.

For brevity, these results were shown for the average value of the shadow short rate. Similar convergence results obtain for other points in the state space, although solutions will not be accurate near the bounds if the underlying state process itself is not actually bounded. For example, in the above case, we would not expect the procedure to generate correct solutions near $i_t = 0.15$. However, so long as the bound on the state space is imposed far enough away from the values of the states that are actually realized in practice, this limitation has a negligible effect on the results. The bottom panel of the figure illustrates this claim by comparing the yield curve computed above with the yield curve computed when the grid for i_t^* is extended over ranges of 30 and 40 percentage points, rather than the 20-point range used above.

4 Some simple examples

To illustrate some of the properties of these models, I briefly consider a series of one- and two factor models in which the asset distribution is fixed. In the one-factor model, the nominal short rate i_t is the only source of stochastic variation; the two-factor model adds inflation. I take periods to be one year in length and suppose that assets have maturities of up to $N = 15$ periods.

I impose that the nominal short rate is bounded below by adopting a "shadow-rate" process. (See Kim and Singleton, 2012, and Wu and Xia, 2016, among others.) In particular, suppose that the shadow short rate i_t^* follows the linear process

$$i_t^* = \phi_0 + \phi_1 i_{t-1}^* + \varepsilon_t \tag{19}$$

where ε_t has variance σ^2. The short rate i_t is given by

$$i_t = \max[i_t^*, b] \tag{20}$$

where the parameter b defines the ELB. I assume zero inflation, so $i_t = r_t$ at all t.

For the moment, assume that inflation is always zero. Then let the nominal SDF be given by

$$M_{t,t+1}^{\$} = \delta_t \exp\left[\lambda^R (R_{t+1} - 1)\right] \tag{21}$$

where λ^R is a risk-aversion parameter. The variable δ_t fluctuates to ensure that the no-arbitrage condition is met for any exogenous variation in the short rate. In particular, δ_t is immediately determined as

$$\delta_t = \frac{\exp[-i_t]}{\mathrm{E}_t\left[\exp\left[\lambda^R (R_{t+1} - 1)\right]\right]} \tag{22}$$

From an economic point of view, fluctuations in δ_t can be interpreted as changes in time or liquidity preference.

For the purposes of illustration, I set $\phi_0 = 0.0052$, $\phi_1 = 0.9$, $\sigma = 0.01$, $b = 0.002$, and $\lambda^R = -8$. These values are calibrated roughly to match the dynamics of the short-rate and the average value of the 10-year yield in the data.

4.1 Effects of the asset distribution

I begin by considering the case in which the maturity structure of assets has a normal distribution that is constant over time. That is

$$x^{(n)} \propto \exp\left[-\frac{(n-n^*)^2}{2\theta^2}\right] \tag{23}$$

where the parameter n^* is the average maturity outstanding, and θ is a scale parameter. Panel A of Figure 2 illustrates the insensitivity of the results to the shape of the asset-maturity distribution, as governed by θ. The left-hand graph shows the distribution with $n^* = 8$ and $\theta = 0.1$ (blue) or $\theta = 2$ (orange). The right-hand graph shows the corresponding yield curves, in both cases taking the short rate to be $i_t = 5.2\%$. The curves are nearly identical. This should not be surprising because the individual asset share $x_t^{(n)}$ does not matter for the individual asset price $p_t^{(n)}$. Only the weighted sums of asset shares shown in equation (8) matters, and they affect all prices in the same way through $M(.)$. Consequently, this model cannot produce local-supply effects from large quantity gluts or shortages in particular sectors of the market.[8]

Panel B shows how a shift in the average duration in investors' portfolio translates into yields. Again taking the short rate to be 5.2%, I consider duration values of $n^* = 5$ years (blue) and $z = 10$ years (orange). In both cases, θ is set to 1, so that, as illustrated on the left, this is just a uniform transposition of the distribution of assets to higher maturities. As shown on the right, the yield curve shifts upward in response—by 72 basis points at the ten-year maturity. Because the expected path of short rates is the same in both cases, the entire difference is attributable to a change in the (real) term premium.

Using the normal distribution in (23) for the asset supply imposes that exposures are always positive at all maturities. In general, one may want to allow for negative exposures (i.e., short positions). A convenient functional form that permits this was proposed by Greenwood et al. (2015):

$$x^{(n)} = 1 + \left(1 - \frac{2n}{N}\right)z \tag{24}$$

where z is a parameter that determines the average maturity of investors' holdings and N is the maximum maturity available. This specification—in which higher values

[8] Malkhozov et al. (2016) make a similar point.

of z tilt the distribution of claims toward lower maturities—is arguably more realistic than the normal distribution and also allows for negative exposures at some maturities. (The four-factor model of the following section adapts this specification allow the maturity distribution to change over time by letting z follow a stochastic process.) One advantage is that the average maturity is linear in z, making interpretation straightforward.

Figure 3 shows how the ten-year yield varies with different values of z in this model, when the short rate is at its average value. The dashed line shows the expectations component of the yield, which does not depend on z. The difference between the dashed and solid lines is the ten-year term premium. As above, the term premium is monotonically increasing in duration (increasing in the value of z). The effects of duration are non-linear, increasing as the amount of duration held by investors increases. For low values of duration, representing negative exposure to long-term assets, the term premium becomes negative.

4.2 Effects of the ELB

Since risk prices are themselves a function of the quantity of risk held by investors, heteroskedasticity can cause risk prices—and therefore term premiums—to differ significantly across states of the world. A particularly important case of this is the lower bound on the nominal short rate. The presence of this bound, all else equal, implies that there is less uncertainty about short-term interest rates in the near future when the current value of those rates is near zero. The reduction in the volatility of short-term interest rates induced by the ELB will dampen duration effects.

Figure 4 illustrates this effect by conducting the same comparative-statics exercise on duration that was depicted in Figure 2.B, but this time with a shadow-rate value of $i_t^* = -3\%$. (This is close to the average value of the shadow rate in the empirical estimates of Krippner (2012).) Yield curves in this region of the space have an "S" shape, due to the expectation for the short rate to remain at zero for some time. Again, the shift from a portfolio duration of 5 years to 10 years causes an increase in longer-term yields, but it is smaller than we obtained away from the ELB. In particular, in this case, the increase in the 10-year yield is only 61 basis points. King (forthcoming) explores the effect of the ELB in similar models in detail.

4.3 Effects of the price of wealth risk

Figure 5 shows the effect of different values of the price of wealth risk, λ^R, for $i_t = 5.2\%$, $z = 8$ years, and $\theta = 1$. A reduction in this parameter, from -8 to -4 in this case, acts much like a reduction in duration, causing the yield curve to fall at longer maturities. Again, this entire decline—77 basis points on the ten-year yield—is due to term premia.

4.4 Effects of inflation

To this point, the model has contained only one factor, the nominal short rate. This specification makes no distinction between real and nominal yields or, equivalently, assumes that inflation is always zero. Given its historical importance as a risk factor in bond pricing and monetary policy, this is a significant omission. To incorporate inflation, I make two adjustments to the above model. First, consistent with equation (6), I distinguish between the real and nominal SDFs as

$$M_{t,t+1} = \delta_t \text{exp}\left[\lambda^R(R_{t+1} - 1 - \pi_t)\right] \tag{25}$$

$$M^{\$}_{t,t+1} = \delta_t \text{exp}\left[\pi_t + \lambda^R(R_{t+1} - 1 - \pi_t)\right] \tag{26}$$

Second, I modify the short-rate process to depend on inflation, similarly to a simple interest-rate rule for monetary policy:

$$i^*_t = \kappa_\pi \pi_t + s_t \tag{27}$$

where

$$s_t = \phi^s_0 + \phi^s_1 s_{t-1} + \varepsilon^s_t \tag{28}$$

The random variable s_t is a factor that drives exogenous variation in the short rate. I calibrate $\kappa_\pi = 1.5$, $\phi^s_0 = 0.0052$, and $\phi^s_1 = 0.9$.

There is now a second factor in the model: inflation. I assume that it's dynamics are

$$\pi_t = \phi^\pi_0 + \phi^\pi_1 \pi_{t-1} + \varepsilon^\pi_t \tag{29}$$

and I calibrate $\phi^\pi_0 = 0$, $\phi^\pi_1 = 0.9$, and the standard deviation of $\varepsilon^\pi_t = 1.6\%$. Note that, for ease of exposition, the unconditional mean of inflation is set to zero.

Panel A of Figure 6 shows how nominal risk premia in this model can be decomposed into compensation for inflation risk and compensation for real risk. It plots the real and nominal yield curves, relative to the expectations component, evaluated at the sample means of the two state variables. Since expected inflation is zero, the difference between the nominal and real yield curves is the inflation-risk premium, and the difference between the real yield curve and the expectations component is the real term premium. Nominal yields are uniformly higher in this model, compared to the one-factor model above—even though the risk aversion parameter is the same—because there is now an additional source of risk that commands a positive premium.

Panel B shows how the real and inflation-risk components of the 10-year term premium vary with the duration of investors' exposures. Both premia are increasing in duration, though they behave somewhat differently. The differences reflect how the covariance of wealth with inflation and (real) short-rate risk changes as investors' holding of assets at different maturities changes.

5 A Macro-Finance Model with Portfolio Effects

5.1 Model setup

I now turn to a fully specified four-factor macro-finance model in which the duration channel is operative, and I take that model to the data. As noted earlier, some papers have proxied investors' duration exposures using observable measures of government bond supply. Recognizing that, in reality, investors have duration exposures through many types of payment claims and obligations, of which the Treasury universe is likely only a small part, I instead treat the duration as an unobserved factor. This factor is identified only from fluctuations in bond yields themselves, together with the cross-equation restrictions implied by the model.

In reduced form, the model is similar to Ang and Piazzesi (2003) and other term-structure models that incorporate observable macroeconomic factors. However, the model here features strong cross-equation restrictions that potentially limit its ability to fit the data, relative to an unrestricted four-factor model. First, one of the latent factors is restricted to only affect the level of the short-term interest rate. Second, the other latent factor, z, only enters the SDF as a combination of asset returns, as in equation (7). Third, the estimation takes into account the fit of the macroeconomic data in addition to the fit of the yield curve. Thus, the factor dynamics are not completely free to fit yields alone.

Specifically, I consider a four-dimensional state vector, $\mathbf{s}_t = (\pi_t, g_t, s_t, z_t)$, where π_t

is inflation, g_t is the growth rate of real consumption, s_t is a factor affecting the level of the nominal short-term risk-free rate, and z_t is a factor governing investor exposures at different maturities. I assume that the state follows a VAR(1) process:

$$\mathbf{s}_t = \phi + \mathbf{\Phi s}_{t-1} + \mathbf{e}_t \tag{30}$$

where the reduced-form error vector $\mathbf{e}_t$ has covariance matrix $\mathbf{\Sigma}$. The short-term nominal rate is given by equation (20), where I calibrate the value of b to 20 basis points and where the shadow rate i_t^* is now specified as a function of the macroeconomic data:

$$i_t^* = \kappa_\pi \pi_t + \kappa_g g_t + s_t \tag{31}$$

Thus, the factor s_t has an interpretation as a deviation from a Taylor-type rule, or as low-frequency variation in such a rule (such as changes in the inflation target or natural rate of interest).

The unobserved factor z_t determines equilibrium asset quantities as in equation (24), taking $N = 10$:

$$x_t^{(n)} = \frac{1}{10} + (1 - \frac{2n}{10}) z_t \tag{32}$$

As noted in the previous section, this formulation follows Greenwood et al. (2015), except that it applies to the future values of claims, rather than to the market value, and that it is normalized to sum to 1, such that the value of $x_t^{(n)}$ at each maturity can be interpreted as a share of total claims. Again, as discussed above, the precise specification of the supply factor loadings generally has only second-order effects. The magnitude of the z_t is linear in the weighted-average maturity (WAM) of investors' cash flows.[9]

I suppose that investor preferences are such that the nominal SDF is

$$M_{t,t+1}^{\$} = \delta_t \exp\left[\pi_{t+1} + \lambda_g g_t + \lambda_R (R_{t+1} - 1 - \pi_t)\right] \tag{33}$$

where, as above, R_{t+1} is the gross return on wealth, given by equation (8). As in the previous section, the random variable δ_t can be calculated exactly, given the other time-t model objects, as

$$\delta_t = \exp\left[-i_t\right] \mathrm{E}_t \left[\exp\left[\pi_{t+1} + \lambda_g g_t + \lambda_R (R_{t+1} - 1 - \pi_t)\right]\right]^{-1} \tag{34}$$

[9]In particular, it is straightforward to show that $WAM_t = \frac{N}{2} + \frac{1}{6}(2 + 3N + N^2))z_t$.

I note that, unlike most modern term-structure models, the model here has both constant coefficients in the SDF and homoscedastic factors. In standard affine models, these two restrictions would result in term premia that are constant. Instead, here, term premia fluctuate with the state of the economy for two reasons. First, the nonlinearities in the model endogenously affect the way that factor volatility passes through to asset-price and SDF volatility. For example, as illustrated above, near the ELB, volatilities are damped, putting downward pressure on term premia. Second, more importantly, changes in the factor z_t shift the composition of wealth between assets of different volatility. As a consequence, the SDF is heteroskedastic, even though the underlying factors are not. Again, this is the same mechanism that drives term premia in GVV. A productive extension of the model would be to allow for coefficients and factor volatilities to also be state-dependent, but the results presented below suggest that a good deal of yield-curve variation can be explained without the need for these extensions.

5.2 Estimation

I estimate the model on annual data from 1971 through 2017. There are two reasons for using annual data. First, from a practical standpoint, having both fewer observations and fewer points on the yield curve greatly increases computational efficiency. Second, it is important to capture lower-frequency properties of the macro data to fit bond yields. (See, e.g., Piazzesi and Schneider, 2007.) Using annual data allows for this, while still maintaining a relatively parsimonious first-order dynamic process.

I fit the model to PCE inflation rates, growth of nondurables and services from the NIPA data, and 1-, 5-, and 10-year nominal Treasury yields, and 5- and 10-year TIPS yields. All yields are Gurkaynak et al. (2007) zero-coupon yields and are averaged to produce annual values. Let the 5×1 vector $\mathbf{y}_t$ collect the yield data, and denote by Ω the 5×5 covariance matrix of the error terms on the long-term yields produced by the model. For ease of notation, I collect all of the other model parameters in the vector $\boldsymbol{\Theta} = (\boldsymbol{\phi} \quad \boldsymbol{\Phi} \quad \boldsymbol{\Sigma} \quad \kappa_g \quad \kappa_\pi \quad \lambda_g \quad \lambda_R)$.

Because Treasury yields depend on the duration factor and the shadow rate in a nonlinear way without a closed-form solution, I estimate the series $\{z_t\}$ by means of a particle filter (see Doucet et al., 2001). Specifically, for a given set of parameters $\boldsymbol{\Theta}j$ and $\boldsymbol{\Omega}_j$, estimation proceeds as follows.

1. Calculate the real and nominal SDF M_j and $M_j^{\$}$ based on $\mathbf{\Theta}j$, using the procedure described in Section 3.

2. Using $\mathbf{\Theta}j$ for the state dynamics, run the particle filter. Set $t = 1971$ (corresponding to the first annual observation). Then,

   (a) Draw values of $z_{t,k} \sim p\left(z_t|\mathbf{\Theta}j, \mathbf{s}_{t-1}\right)$, for $k = 1$ to $100,000$.

   (b) For each draw $z_{t,k}$, use the state dynamics to construct model implied yields as conditional expectations of the SDF, given the time-t state.

   (c) Evaluate the weights $w_{k,t} \propto \Pr(\mathbf{y}_t|z_t, \pi_t, g_t, i_t, \mathbf{\Theta}j, \Omega_j)$ based on the yield data. For the early part of the sample, where TIPS yields do not exist, the weights are only proportional to the PDF of the nominal yield errors.

   (d) Resample 10,000 draws from the distribution of $\{z_t\}_{1971}^{t}$, using the weights $w_{k,t}$.

   (e) $t = t + 1$.

   Note that, in running the particle filter, I keep the whole history of each resampled particle. The resulting distribution of paths is a random sample from the posterior distribution $p\left(\{z_t\}_{1971}^{2017}|\mathbf{\Theta}j, \Omega_j, \{\pi_t, g_t, i_t, \mathbf{y}_t\}_{1971}^{2017}\right)$. That is, it is a smoothed, rather than a filtered, estimate.

The fixed parameters of the system are estimated by maximum likelihood, with standard errors computed via the delta method.

6 Results

6.1 Yield curve fit and decomposition

Table 1 reports the parameter estimates. The short-rate coefficients κ_g and κ_π are roughly in line with other estimates of simple interest-rate rules. The coefficients λ_g and λ_R, reflecting the market prices of consumption risk and wealth risk, respectively, are 0.9 and -4.2. However, they are estimated with substantial uncertainty.

Figure 7 shows the yields used in the estimation (in red), together with the posterior medians and 10% - 90% credibility intervals of the model-implied yields. (The 15-year nominal yield was not used in the estimation but is shown for comparison.) As shown in Table 1, the estimated standard deviations of the error terms, evaluated at the posterior median, range from 25 to 31 basis points.

Figure 8 shows the model-implied decomposition of the 10-year yield into its four components: the expected real short rate, the expected inflation rate, the real term premium, and the inflation-risk premium. The real term premium and the inflation risk premium sum to the nominal term premium by definition. The average expected nominal rate and average expected inflation over the next ten years are computed, for each period's state vector, based on the state dynamics. The average expected real rate is the difference between these series. The corresponding real and nominal term premia are simply the differences between the model-implied (real or nominal) 10-year yield and its expectation component. The series shown are the means of the model-implied distributions of each series.

The model has several features that are common to less-restricted term-structure models (i.e., those with multiple latent factors and no economic interpretation of the SDF), such as Kim and Wright (2007), Adrian et al. (2017), and D'Amico et al. (2017). In particular, it implies a fairly rapid increase in the nominal term premium in the late 1970s and early 1980s, followed by a gradual downward drift. The real term premium ranges from about 0.5% to about 2%, and the inflation risk premium ranges from about 0.5% to about 1.5%. They are highly correlated, and both contribute to the decline in the overall term premium since 1980. Expected inflation also moved lower in the 1980s, largely following realized inflation. The ten-year expected real short rate has drifted lower by about 200 basis points between about 2000 and 2017, consistent with some other estimates of "r-star."

It is interesting to note that the model exhibits significant time variation in the inflation-risk premium, because (apart from the effects of the ELB) the variance of inflation is constant, and it enters into the SDF with a constant exponent of 1. Thus, the time variation arises solely from the covariance of inflation with the return on wealth, which is a function of the duration factor z_t.

6.2 The latent factors

Figure 9 shows the estimates of the two latent factors in the model. Panel A presents the short-rate factor. It moves somewhat lower since the early 1980s, consistent with downward drift in both the inflation target and the neutral rate of interest. Of course, this variable is pinned down fairly precisely by short-term yields. However, it becomes less precisely estimated during the ELB period, when those yields are no longer informative about its value.

Panel B shows the estimated path of the duration factor over time (median and 80% credibility interval). For ease of interpretation, the factor value is converted to the WAM of claims in investors' portfolio. It starts the sample period low but rises sharply in the early 1980s. It displays little systematic trend over the remainder of the sample, although it spikes around the trough of every recession. Although the level of the estimated z_t is correlated with the other model objects shown in Figure 8, it does not bear a one-to-one correspondence with any of them due to the nonlinearities of the model.

The factor has an interpretation as the average maturity of claims to future payments held by investors. While these claims may take a number of forms—including privately issued securities, real assets, and future income streams—a case of particular interest is that of Treasury debt. To see whether the size and structure of actual Treasury debt bears any relation to the duration factor, I regress the series in Figure 9 on the WAM of all Treasury debt outstanding (WAM) and on the maturty-weighted debt-to-GDP ratio (MWD). Both measures are calculated using all outstanding Treasury securities available in CRSP, and I control for 1-year yields in the regressions.

Table 2 reports the results. Both measures are strongly positively correlated with the estimated factor. (In fact, coincidentally, they take almost identical coefficient values.) Of the two, WAM is more consistent with the interpretation given to the factor in the model, since it is in the same units. According to the regression, a one-year increase in Treasury WAM is associated with a 1.36-year increase in model WAM. We cannot reject that the coefficient is equal to 1.

Figure 10 shows the estimated relationship between the duration factor and the level of the 10-year yield. In the top, black line, the other three state variables are all set to their sample means, giving a short rate of 5%. In the bottom, blue line, inflation and growth are set to their sample means, but the shadow rate i_t^* is set to -10%, so that the ELB is binding. In both cases, there is a positive association between duration and yields. As in the one-factor model of the previous section, this association is attenuated at the ELB.

Figure 11 shows that a significant portion of the level of and variation in the ten-year yield is driven by z_t. Specifically, I set $\lambda^R = 0$, so that the return on wealth is no longer an element of the SDF, holding all other parameters and state variables the same. The resulting counterfactual 10-year yield is shown by the blue line, while the black line is the yield implied by the baseline model. The shaded region, which averages about 1.5 percentage points, represents the contribution of the return on wealth to long-term rates. Looked at in this way, the duration channel accounts for the bulk of the nominal term premium (which averages 1.9%). This result is also shown in the second column of Table 3, which reports the model errors in the counterfactual model. (Column 1 reports the errors in the baseline model, for comparison.)

A related question is whether a model that includes only consumption and inflation in the SDF can do equally well in fitting the data. To answer this question, I re-estimated the model without including R_t in $M(.)$. Note that, even in this case, the unobserved factor z_t may continue to help, to the extent that it might absorb predictable variation in nominal yields that is not accounted for by the observable variables, although it would no longer have an interpretation as reflecting portfolio duration. The last column of Table 3 reports the model fit in this case. Re-optimizing the parameters of the model improves the fit substantially relative to the second column, mostly because the model is able to raise the level of yields relative to that column and eliminate the downward bias that was present in Figure 7. However, the errors in matching yields are still about twice as large as in the baseline model (even though the number of observable and unobservable factors is the same). Evidently, adding the return on wealth to the SDF and treating z_t as reflecting portfolio duration is advantageous for fitting the yield-curve data.

6.3 Impulse-response functions

I now consider the dynamic effects of shocks to the short rate and the duration factor. For this purpose, I adopt a structural decomposition of $\mathbf{\Sigma}$ using short-run ordering restrictions. Given the use of annual data, it is not realistic to impose the usual assumption that certain variables cannot respond to others for a least one period. Therefore, I estimate the shocks using higher-frequency data, as follows. First, using the Φ parameters and the estimated annual values of z_t, I interpolate the z_t series to a quarterly frequency. Then, I re-run the model using quarterly data for all four variables, including three lags of the states. I apply the Cholesky decomposition to the covariance matrix of the error terms of this higher-frequency model to obtain the short-run ordering. I order the duration factor last and the short rate second-to-last.

The responses of the real and nominal yield curves to the two shocks is shown in

Figure 12, using the maximum-likelihood parameter values. Monetary-policy shocks decay nearly monotonically, and the expectation of this behavior is reflected in the initial response of the yield curve, which is greatest at the short end. (This shape is somewhat different at the ELB, not shown.). These shocks have modest effects on inflation, so most of this behavior is passed through to real yields.

The duration shock results in a much different shape of the initial response. In response to a one-standard-deviation shock (about 4 months of duration, at an annual frequency) short-term yields do not move at all, while nominal yields beyond about 7 years rise by about 18 basis points. The real yield curve rises even a bit more than the nominal curve, although that response exhibits a slight hump shape across maturities, with the peak around the 5-year sector. The effect of duration shocks on both real and nominal yields is hump-shaped over time, and decays very slowly after the first few years.

Taking a closer look at these responses, Figure 13 shows the response of the expectations and term-premium components of nominal yields across maturities, in the period when the shocks occur. Most of the effect of the monetary-policy shock is on the expectations component, but there is a sizeable effect on term premia as well. I will return to this result momentarily. The duration shock has a small effect on short rate expectations that arises through the dynamic response of the economy to changes in z_t and the associated response of the interest-rate rule. Perhaps surprisingly, the effect of the duration shock on the term premium is fairly modest.

The shocks also have implications for the dynamic paths of the economic variables. These are shown in Figure 14. Both shocks have no significant effect on inflation but a modestly negative effect on consumption. Interestingly, the duration factor rises following a shock to the nominal short rate. This reaction is responsible for the modest increase in term premia associated with this shock, noted in the previous figure. It is consistent with investors moving into longer-term assets in response to increases in interest rate, a behavior that Hanson and Stein (2015).

Using these results, we can do a back-of-the-envelope calculation to consider the effects of the Federal Reserve's asset purchase programa. In King (forthcoming) I calculate that the Fed's asset purchases collectively reduced the dollar duration in public hands by about 20%, relative to what it would have otherwise been at the end of the program. MWD at the end of 2014 was 5.0. Thus, QE may have been responsible for a reduction of about 1.25 in this variable. From the second column of Table 2, this would map into a value of z_t that is about 1.7 years lower. The IRFs above suggest that this would translate into about a 80 basis point decrease in long-term nominal yields and about a 95 basis point decrease in long-term real yields, with about two-thirds of this effect arising through term premia. In addition, such a shock would have raised consumption growth by about 0.8% over each of the following three to four years and perhaps have resulted in a some boost to inflation (though the latter response is not statistically significant.)

There are several caveats to this calculation. First, QE programs were implemented at the ELB, and the attenuating effects of that environment have already been noted. Second, the programs were spread over several years, rather than occurring as a single large shock. That timing difference could matter in the presence of nonlinearities. Third, the dynamics of central bank asset purchases are likely to differ from those of Treasury debt or other elements of investors' equilibrium portfolios. Since investors perceptions of these dynamics matters for their response, the effect of QE on yields and economic variables might be different from the effects of other types of shocks to z_t.

7 Conclusion

This paper has presented a method for solving a broad class of models in which the maturity distribution of investors' assets matters for bond yields (and other asset prices) through the dependence of the pricing kernel on the return on wealth. These models are inherently nonlinear and analytically intractable, and I develop an algorithm for solving them. I set up and estimate one such model, which includes both inflation and real activity as observable factors. To my knowledge, this is the first attempt to integrate portfolio-balance / duration effects of the type explored in Vayanos-Vila into a structural macro-finance asset pricing model. The model allows one to examine such issues as the relative effects of short-rate and bond-supply shocks, which might be of interest for calibrating monetary policy.

Generally, the model suggests that the direct effects of duration shocks on term premia are fairly small. However, the *presence* of duration in the model, through the return-on-wealth term in the stochastic discount factor, is very important for explaining the behavior of yields. Put somewhat differently, fluctuations in duration are less important for the term structure than fluctuations in real asset prices, which then feed back through their affect on aggregate portfolio returns. Adding these features to more-realistic and fully specified models of the macroeconomy is an important direction for future research.

References

Adrian, T.; Crump, R. K.; and Moench, E., 2013. "Pricing the Term Structure with Linear Regressions." *Journal of Financial Economics* 110(1): 110-38.

Ang, A., and Piazzesi, M., 2003. A No-Arbitrage Vector Autogression of Term Structure Dynamics with Macroeconomic and Latent Variables. " *Journal of Monetary Economics* 50: 745-87.

Bauer, M. D. and Rudebusch, G. D., 2014. "The Signaling Channel for Federal Reserve Bond Purchases." *International Journal of Central Banking* 10(3): 233-89.

Bernanke, B. S.; Reinhart, V., R.; and Sack, B. P., 2004. "Monetary Policy Alternatives at the Zero Bound: An Empirical Assessment." *Brookings Papers on Economic Activity* (2): 1 – 78.

Cahill, M.; D'Amico, S.; Li, C; and Sears, J., 2014. "Duration Risk versus Local Supply Channel in Treasury Yields: Evidence from the Federal Reserve's Asset Purchase Announcements." FEDS paper 2013-35 (April).

Chung, H.; Laforte J.-P.; Reifschneider, D.; and Williams, J. C., 2012. "Have We Underestimated the Likelihood and Severity of Zero Lower Bound Events?" *Journal of Money, Credit and Banking* 44 (s1): 47 - 82.

D'Amico, S.; English, W.; Lopez-Salido, D.; and Nelson, E., 2012. "The Federal Reserve's Large-Scale Asset Purchase Programmes: Rationale and Effects." *Economic Journal* 122 (564): F415-46.

D'Amico, S. and King, T. B., 2013. "Flow and Stock Effects of Large-Scale Treasury Purchases: Evidence on the Importance of Local Supply." *Journal of Financial Economics* 108(2): 425-48.

Doucet, A.; de Freitas, N.; and Gordon, N., 2001. *Sequential Monte Carlo Methods in Practice.* Springer: New York.

Eggertsson, G. and Woodford, M., 2003. "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity* (1): 139-211.

Epstein, L. G. and Zin, S. E., 1989. “Substitution, Risk Aversion, and the Temporal Behavior of Consumption Growth and Asset Returns I: A Theoretical Framework.” *Econometrica* 57(4): 937-69.

Frankel, J. A., 1985. “Portfolio Crowding-out, Empirically Estimated.” *Quarterly Journal of Economics* 100: 1041-65.

Gagnon, J.; Raskin, M.; Remache, J.; and Sack, B., 2010. “Large-Scale Asset Purchases by the Federal Reserve: Did They Work?” *International Journal of Central Banking.*

Gertler, M.; and Karadi, P., 2011. ”A Model of Unconventional Monetary Policy.” ” *Journal of Monetary Economics* 58: 17-34.

Greenwood, R. and Vayanos, D., 2010. “Price Pressure in the Government Bond Market.” *American Economic Review* 90(2): 585-90.

Greenwood, R. and Vayanos, D., 2014. “Bond Supply and Excess Bond Returns.” *Review of Financial Studies* 27(3): 663-713.

Gurkaynak, R. S.; Sack, B.; and Wright, J. H., 2006. “The U.S. Treasury Yield Curve: 1961 to the Present.” *Journal of Monetary Economics* 54(8): 2291-394.

Hamilton, J. D. and Wu, J. C., 2012. “The Effectiveness of Alternative Monetary Policy Tools in a Zero Lower Bound Environment.” *Journal of Money, Credit and Banking* 44(s1): 3 – 46.

Hanson, S. G. and Stein, J. C., 2015. ”Monetary Policy and Long-Term Real Rates.” *Journal of Financial Economics* 115(3): 429-48.

Joyce, M. A. S.; Lasaosa, A.; Stevens, I.; and Tong, M., 2011. ”The Financial Market Impact of Quantitative Easing in the United Kingdom.” *International Journal of Central Banking* 7(3): 113-61.

Kaminska, I; and Zinna, G., forthcoming. *Journal of Money, Credit, and Banking.*

King, T. B., 2015. ”A Portfolio-Balance Approach to the Nominal Term Structure.” FRB Chicago Working Paper 2013-18.

King, T. B., forthcoming. "Expectation and Duration at the Effective Lower Bound." *Journal of Financial Economics.*

Kim, D. H. and Wright, J. H., 2005. "An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates." FEDS paper 2005-33.

Krishnamurthy, A. and Vissing-Jorgensen, A., 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity* (2): 215-65.

Krishnamurthy, A. and Vissing-Jorgensen, A., 2012. "The Aggregate Demand for Treasury Debt." *Journal of Political Economy* 120(2): 233-67.

Kuttner, K., 2006. "Can Central Banks Target Bond Prices?" NBER Working Paper 12454.

Li, C. and Wei, M., 2012. "Term Structure Modeling with Supply Factors and the Federal Reserve's Large-Scale Asset Purchase Programs." *International Journal of Central Banking* 9(1): 3-39.

Meaning, J. and Zhu, F., 2011. "The Impact of Recent Central Bank Asset Purchase Programmes." BIS *Quarterly Review*, Dec.

Malkhozov, A.; Mueller, P.; Vedolin, A.; and Venter, G., 2016. "Mortgage risk and the yield curve." *Review of Financial Studies* 29(5): 1220-53.

Piazzesi, M. and Schneider, M., 2008. "Bond Positions, Expectations, and the Yield Curve." FRB Atlanta working paper 2008-2.

Rudebusch, G. D. and Swanson, E. T., 2012. "The Bond Premium in a DSGE Model with Long-Run Real and Nominal Risks." *American Economic Journal: Macroeconomics* 4(1): 105-43.

Tauchen, G. and Hussey, R., 1991. "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models." *Econometrica* 59(2) 371-96.

van Binsbergen, J. H.; Fernandez-Villaverde, J.; Koijen, R. S. J.; and Rubio-Ramirez, J., 2012. "The Term Structure of Interest Rates in a DSGE Model with Recursive Preferences." *Journal of Monetary Economics* 59(7): 634-48.

Vayanos, D. and Vila, J.-L.. 2009. "A Preferred-Habitat Model of the Term Structure of Interest Rates." NBER Working paper 15487 (November).

Wachter, J., 2006. "A Consumption-Based Model of the Term Structure of Interest Rates." *Journal of Financial Economics* 79(2): 365-99.

Table 1. Parameter Estimates

κ^G	κ^R	λ^G	λ^R	$\omega^{(1nom)}$	$\omega^{(5nom)}$	$\omega^{(10nom)}$	$\omega^{(5tips)}$	$\omega^{(10tips)}$
1.02 (0.29)	1.16 (0.15)	0.9 (94.3)	-4.2 (6.4)	0.25%	0.27%	0.27%	0.30%	0.31%

Note: Standard errors in parentheses.

Table 2. Regressions of Duration Factor on U.S. Debt Metrics

	Dep. var: Estimated duration factor	
Intercept	-5.37** (2.10)	-2.34 (1.40)
1y yield	30.2*** (10.9)	46.2*** (12.1)
WAM	1.36*** (0.35)	
MWD/GDP		1.36*** (0.35)
Adj. R^2	0.679	0.681

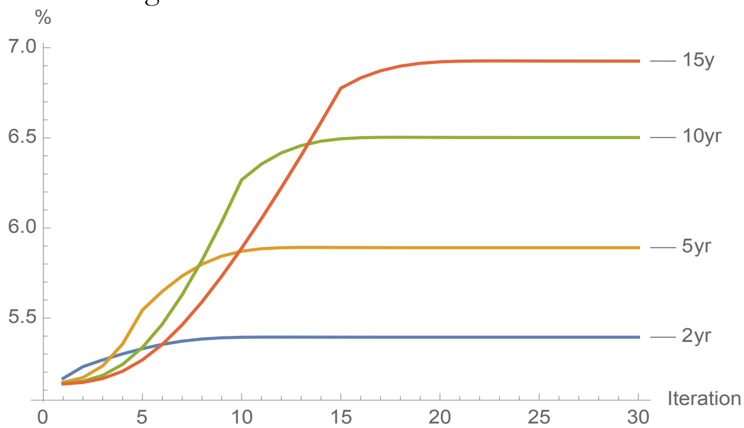
Note: Standard errors in parentheses.

Table 3. Fit Statistics

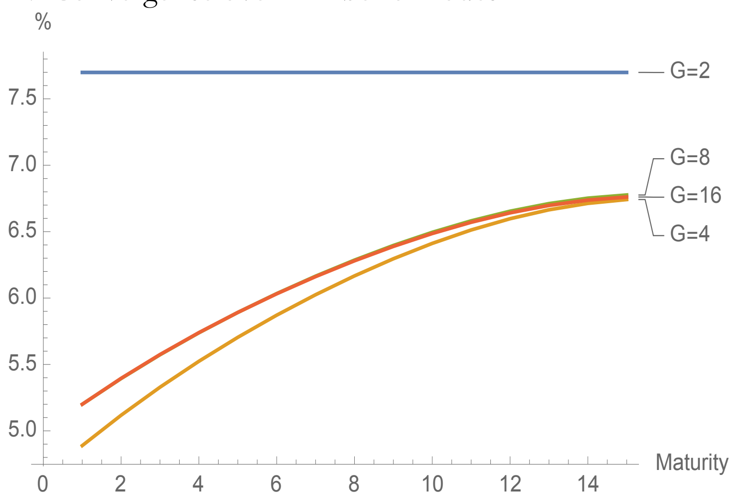
	Baseline	Baseline parameters, but $\lambda^R = 0$	Re-estimated w/o R_{t+1} in SDF
Inflation t+1	1.30%	1.30%	1.38%
Consumption t+1	1.12%	1.12%	1.20%
1y yield t+1	1.30%	1.42%	1.44%
5y nom. yield t	0.35%	2.33%	0.62%
10y nom. yield t	0.35%	1.93%	0.69%

Figure 1. Convergence of the Algorithm in the One-Factor Model

A. Convergence over iterations



B. Convergence over number of nodes



C. Convergence over range of nodes

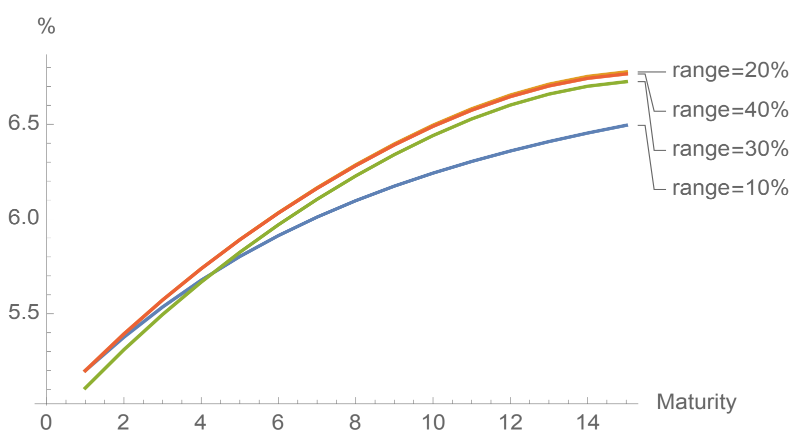
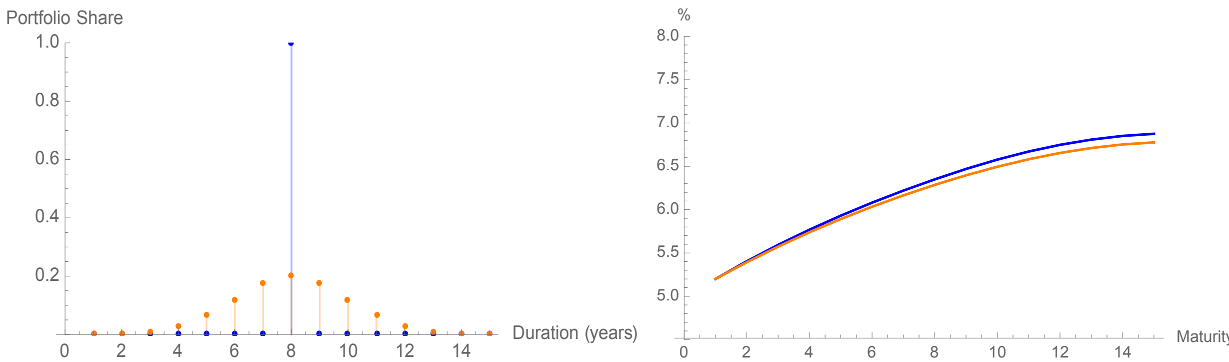


Figure 2. Duration Effects in the One-Factor Model, with Normally Distributed Exposures

A. Effect of the distribution shape



B. Effect of duration

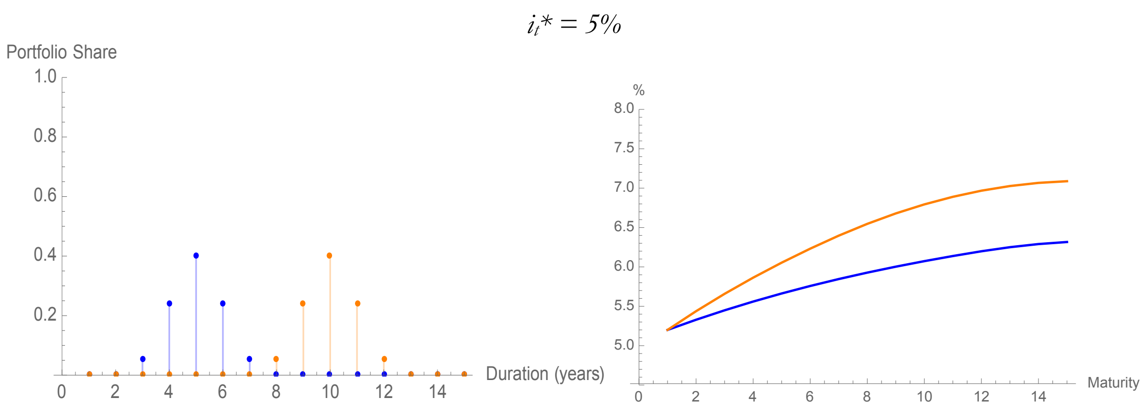


Figure 3. Duration Effects in the One-Factor Model, with Linearly Distributed Exposures

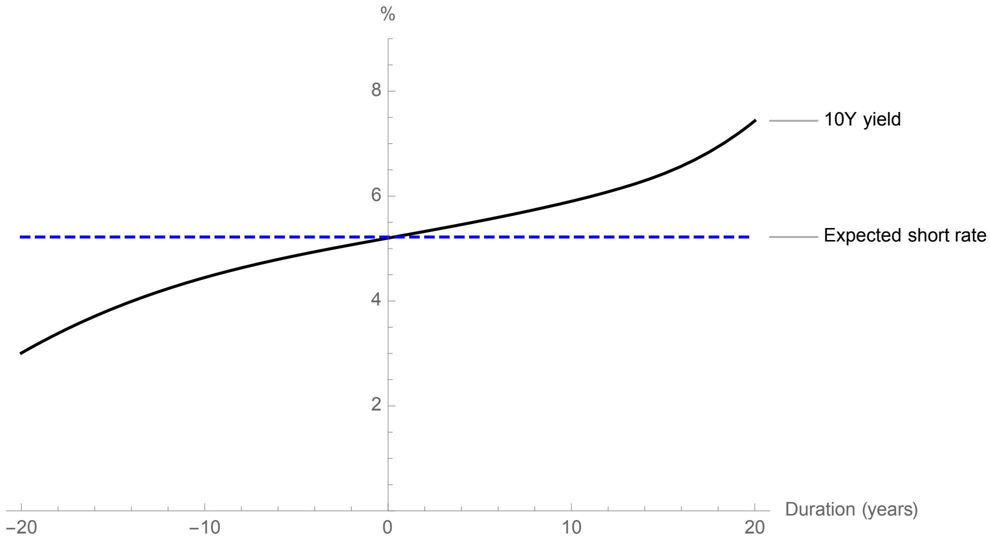


Figure 4. Effect of the ELB in the One-Factor Model

$i_t* = -3\%$

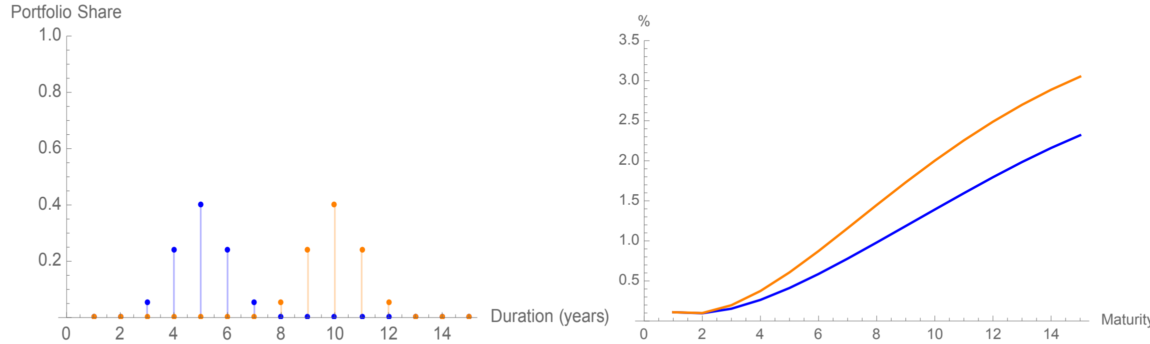


Figure 5. Effect of the Risk Price in the One-Factor Model

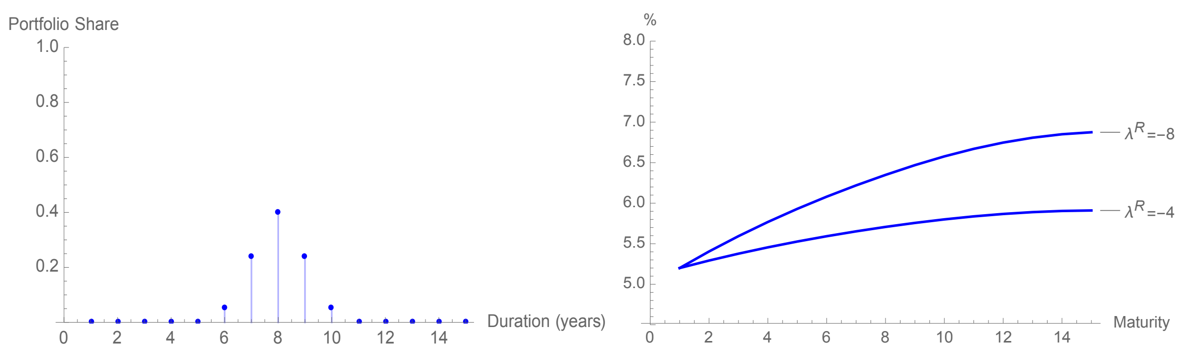
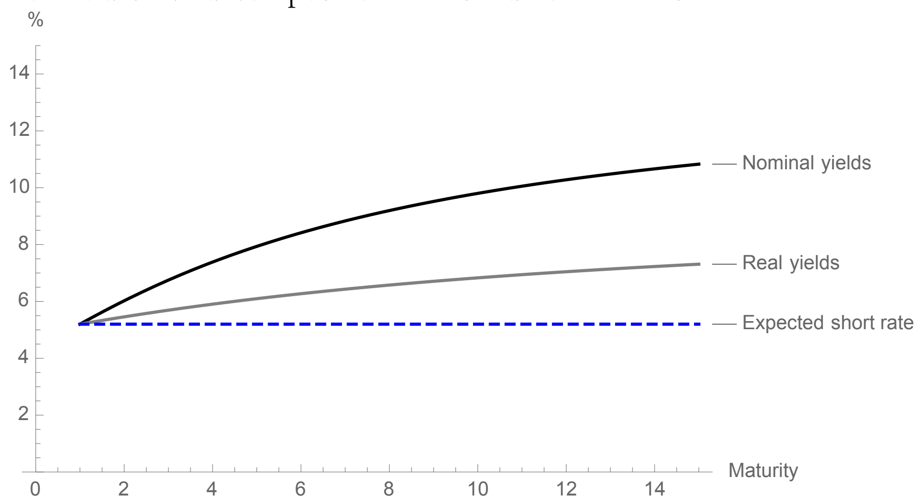


Figure 6. Effect of Inflation in the Two-Factor Model

A. Yield-curve decomposition at unconditional means



B. Duration effects on 10-year risk premia

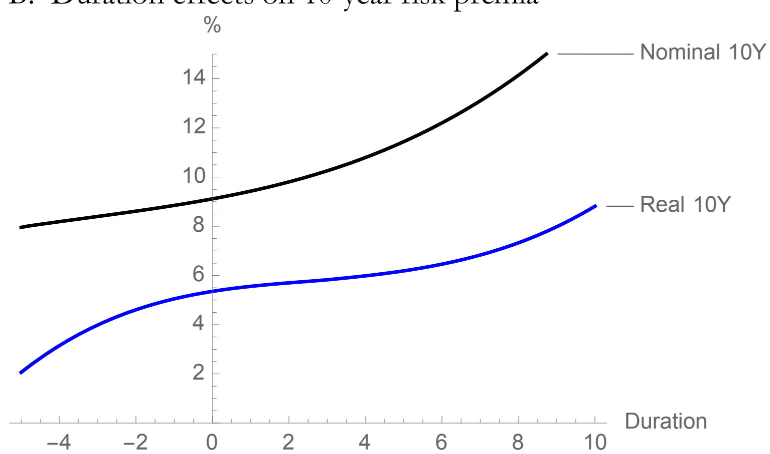


Figure 7. Four-factor model-implied yields vs. data

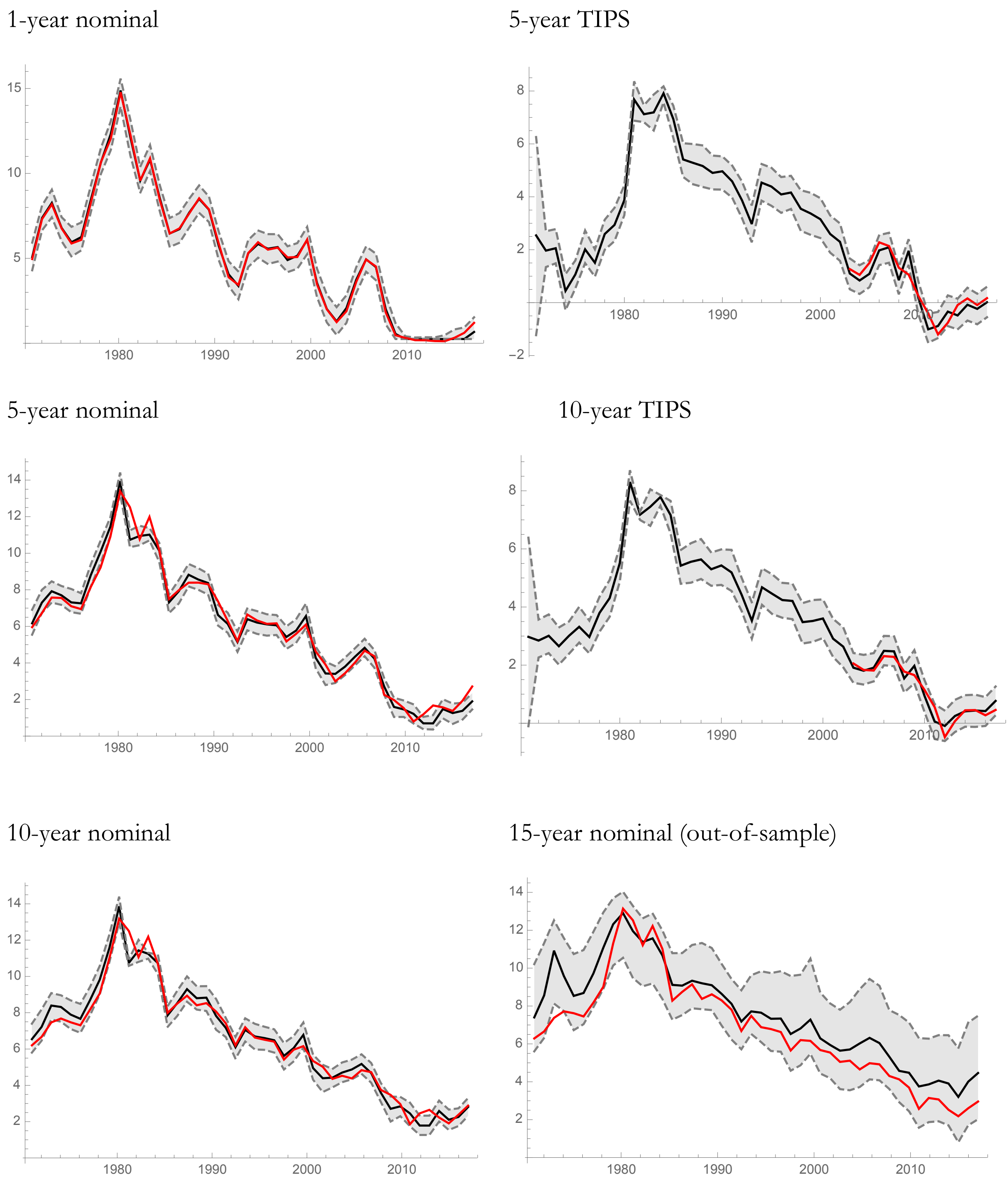
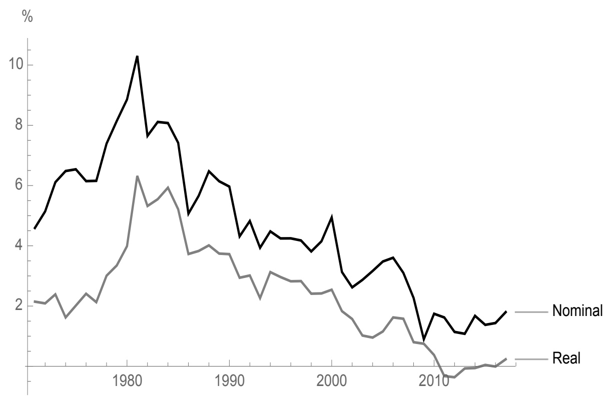


Figure 8. Decomposition of 10-year Nominal Yield

Expected short rates



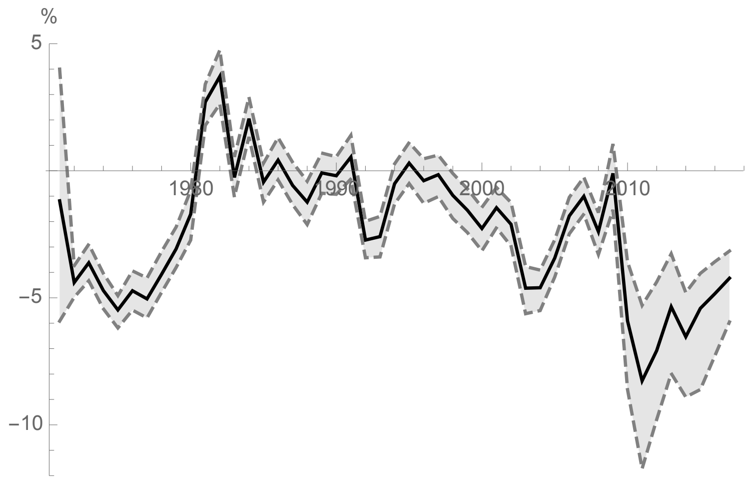
Expected inflation

Term premia

Inflation-risk premium

Figure 9. Latent variable estimates

Short-rate factor



Duration factor

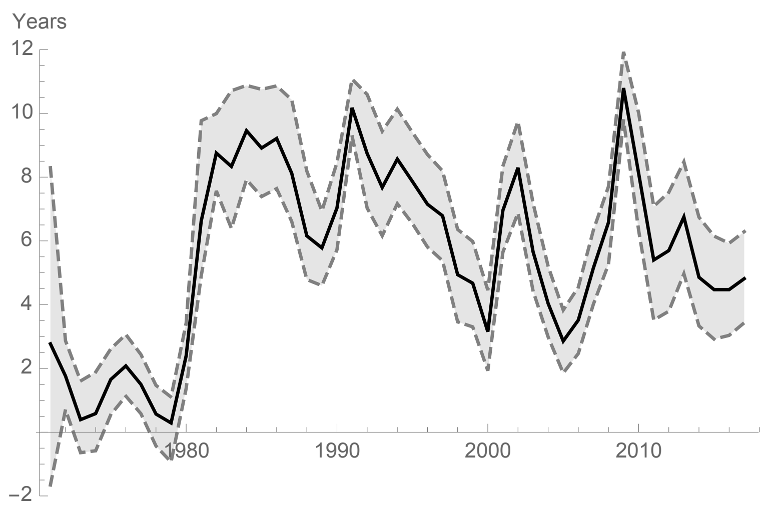


Figure 10. 10-year yield as a function of the duration factor

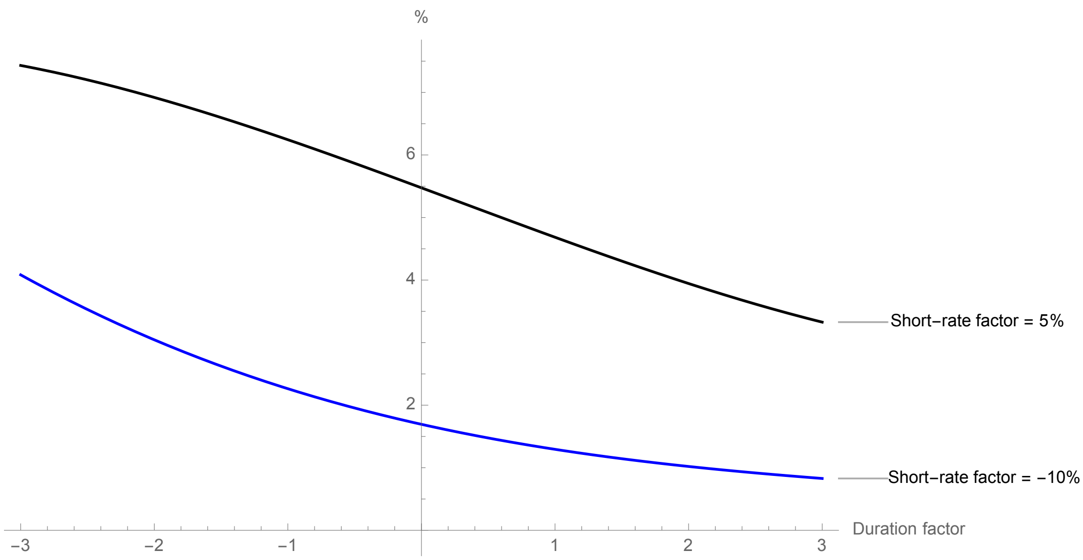


Figure 11. Contribution of duration effect to 10-year yield

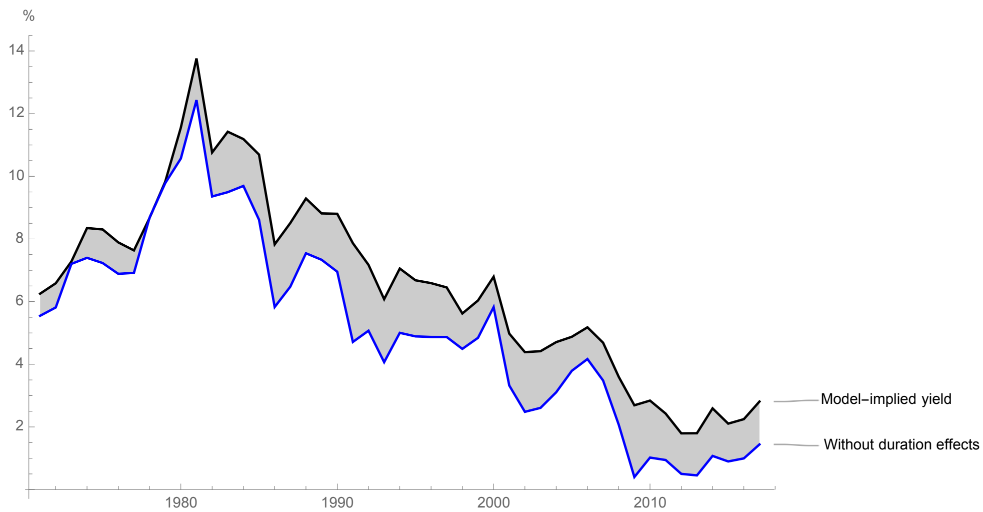
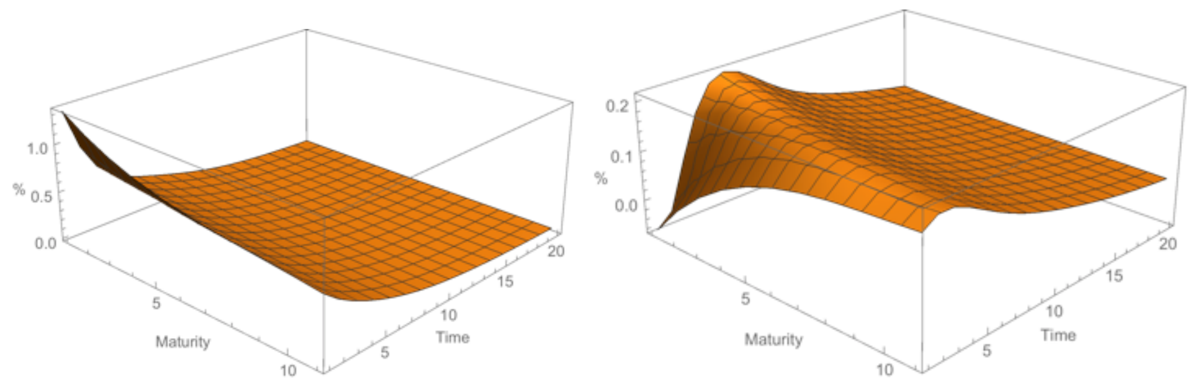


Figure 12. Impulse-response functions – Yield curve

Monetary-policy shock *Duration shock*

Nominal yields



Real yields

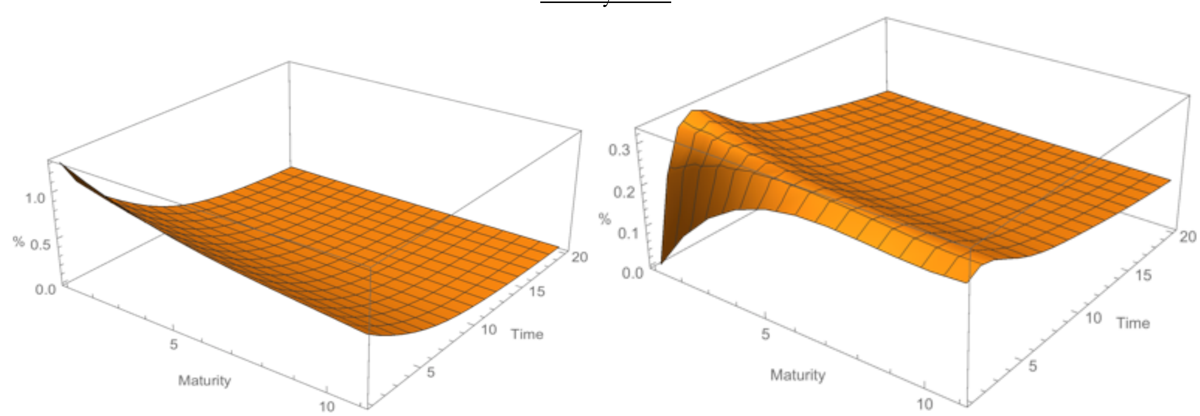


Figure 13. Immediate impact of shocks on yield curve components

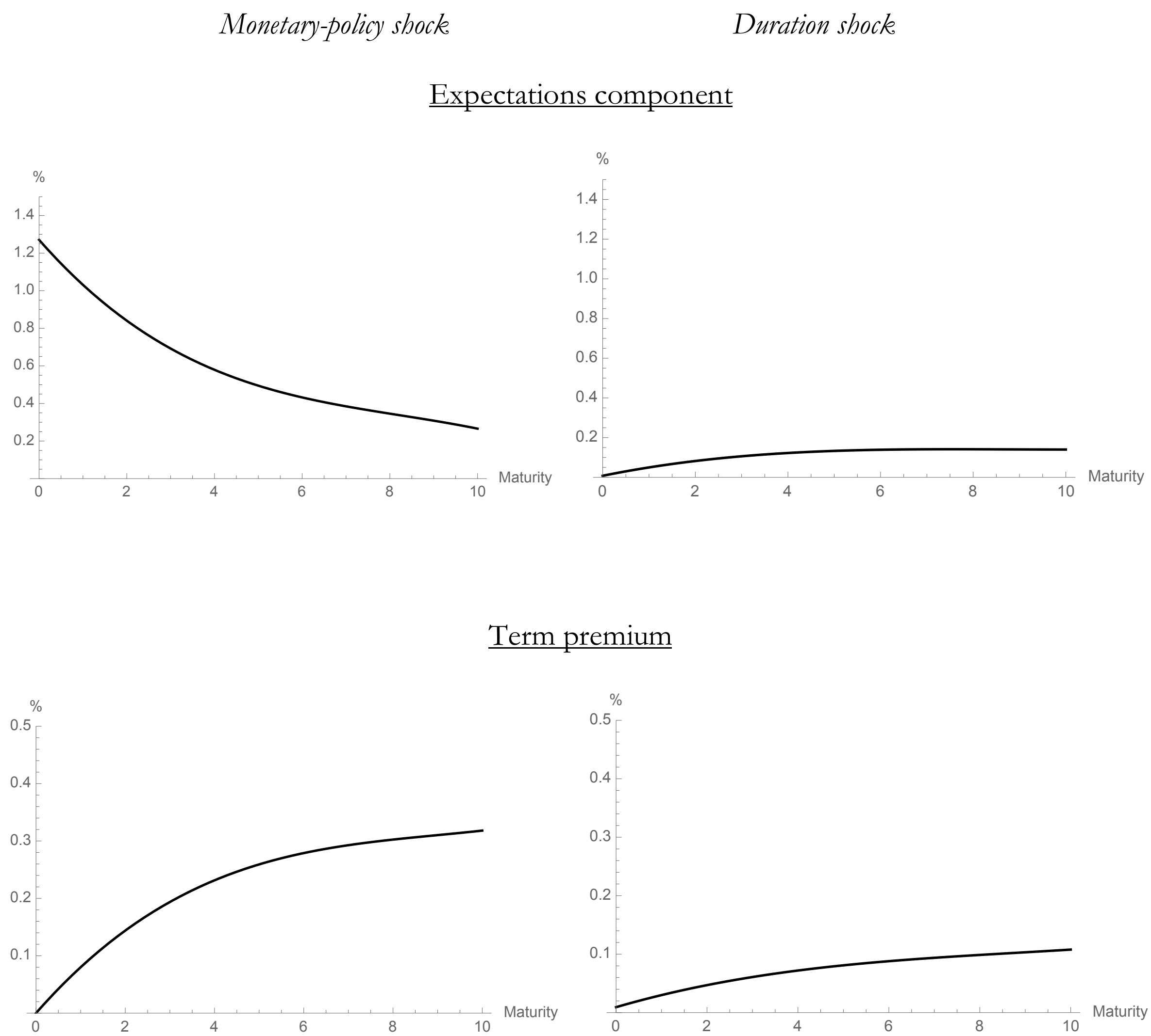


Figure 10. Impulse-response functions – state variables