A Portfolio-Balance Approach to the Nominal Term Structure

Thomas B. King*

November 14, 2014

Abstract

I examine how the maturity structure of outstanding government liabilities affects the nominal yield curve under a variety of assumptions about investor objectives. In the class of models I consider, equilibria are arbitrage free, expectations are rational, and assets are valued only for their pecuniary returns. Nonetheless, a portfolio-balance mechanism arises through the dependence of the pricing kernel on the return on wealth. This mechanism results in a positive relationship between the duration of bondholders' portfolios and the price of interest-rate risk. Quantitatively, the models suggest that the effects of shifting Treasury supply on yields can be substantial—for example, the increase in the average maturity of U.S. government debt that occurred between 1976 and 1988 may have raised the ten-year yield by 50 basis points. On the other hand, partly reflecting an attenuation of portfolio-balance effects when interest rates are near zero, the Federal Reserve's asset purchase programs likely had a fairly small impact on the yield curve by removing duration from the market.

*Federal Reserve Bank of Chicago. Contact: thomas.king@chi.frb.org. I thank Stefania D'Amico, Jonathan Goldberg, Francisco Vazquez-Grande, Andrea Vedolin, and seminar participants at the Federal Reserve Bank of Chicago, the Federal Reserve Board, and the ECB/Bank of England workshop on Understanding the Yield Curve for helpful comments and discussions. Roger Fan provided excellent research assistance. The views expressed here do not reflect official positions of the Federal Reserve.

1 Introduction

Recent empirical work has been nearly universal in concluding that fluctuations in the structure of government debt have significant effects on the term structure of interest rates and, most likely, on other asset prices.[1] This issue has received particular attention following the recent efforts of several central banks to reduce long-term interest rates by purchasing large quantities of government debt and other securities. Yet, although much of the recent empirical evidence comes from studies of these asset purchases themselves, enough of it derives from other episodes to demonstrate that the mechanism transmitting debt fluctuations to financial conditions is not unique to recent experience or, for that matter, to monetary-policy interventions. 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. Since it supposedly works by shifting investors' exposure to long-term interest rates, this outcome is sometimes referred to as the "duration channel."[2]

Despite its intuitive appeal, the duration channel is all but absent from the extant theoretical literature, and several authors argue that the empirical results noted above probably reflect some other mechanism.[3] 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. For their part, advocates of the duration channel frequently point to the model of Vayanos and Vila (2009) for theoretical support. In that model, shifts in the demand for Treasury bonds on the part of preferred-habitat investors change the equilibrium exposures of arbitrageurs to interest-rate risk, causing fluctuations in term premia. However, the Vayanos-Vila model is both analytically intractable and quite stylized—for example, investors only hold Treasury bonds, inflation and consumption risk are not priced, and policy actions by the government are not explicitly modeled. These features make it difficult to assess the economic importance of its central mechanism and to reconcile it with other macro-finance literature, including Eggertsson and Woodford's neutrality proposition.

This paper notes that the potential for asset quantities to affect asset prices exists in *any* model in which the pricing kernel depends on the return on wealth.

[1] 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), Ihrig et al. (2012), Joyce et al. (2011), Li and Wei (2012), Rosa (2012), Cahill et al. (2013), D'Amico and King (2013), Bauer and Rudebusch (2014), Chabot and Herman (2014), and Rogers et al. (2014).

[2] For example, Gagnon et al. (2010) appeal strongly to this idea. Bernanke (2010) and Yellen (2011) refer to a similar mechanism in discussing how LSAPs work.

[3] See Woodford (2012), Krishnamurthy and Vissing-Jorgensen (2013), and Bauer and Rudebusch (2014).

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. In the spirit of an older literature that studied how asset supply mattered for prices, I refer to these as "portfolio balance effects." However, unlike classical appications of portfolio-balance concept, the equilibria considered here are arbitrage free and obey rational expectations. They do not require that there be anything "special" about money or short-term debt. In the case of Treasury bonds, no-arbitrage portfolio balance is essentially equivalent to the duration channel. 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 Vayanos-Vila model is a member of this class, while the Eggertsson-Woodford model is not.

I study how debt structure affects interest rates in several calibrated models of this type, spanning a variety of assumptions about investor preferences and portfolios. In particular, the models allow variously for inflation and consumption, diversification of investor portfolios into non-Treasury assets, and a zero lower bound on nominal interest rates. Under parameterizations that match the historical data, portfolio-balance effects on the yield curve can be significant. For example, in some cases, an increase in the outstanding duration of Treasury securities of the magnitude that occurred between the mid-1970s and mid-1980s leads to an increase in the 10-year term premium of more than 50 basis points. This finding could explain a portion of the empirical evidence (for example, estimates based on the Kim and Wright (2005) model) that term premiums did increase around this time. However, a critical assumption here is the extent to which the holders of Treasury bonds are exposed *only* to Treasury bonds. For given preferences, portfolio-balance effects of shifting Treasury maturity structure are largest when bondholders do not hold any other assets, implying a high degree of market segmentation.

Moreover, even assuming complete isolation of the Treasury market, these models cannot generate duration effects that are anywhere near empirical estimates of the impact of the Federal Reserve's large-scale asset purchases on longer-term interest rates. In part, this impotence has to do with the nonlinearity created by the zero lower bound (ZLB) on nominal interest rates. The truncation of the lower tail of the rate distribution at zero reduces volatility along the entire yield curve, and—since duration effects depend on the product of duration and volatility—render this channel less effective. (Doh, 2010, also makes this point.) The reduction in interest-rate volatility, and therefore in portfolio-balance effects, is even greater when the central bank broadcasts its medium-term intentions for short-term interest rates through so-called "forward guidance."

These findings lead to a number of implications for the interpretation of recent policy actions. The conclusion that duration removal may not be the primary mechanism involved in asset purchases is consistent with the recent event-study analyses of Krishnamurthy and Vissing-Jorgensen (2013) and Cahill et al. (2013). As those authors note, the effects of Treasury purchases might at least partly be explained through a scarcity channel (as in D'Amico and King, 2013) or through the signals that unconventional monetary policy sends about future short-term rates (as in Bauer and Rudebusch, 2014). The result that greater certainty about the course of short rates dampens duration effects implies that forward guidance and asset purchases, two tools of monetary policy that are often viewed as complements at the ZLB, actually offset each other somewhat in terms of their effects on longer-term yields. More broadly, the structure of government debt is likely to matter most for interest rates when interest-rate volatility is high. This observation should inspire caution if the Federal Reserve needs to sell assets at some point, since, to the extent that the path of policy is likely to be less certain in that environment, the duration effect could be asymmetrically large relative to the period of asset purchases. Similar considerations may also be worth policymakers' attention when designing the pattern of Treasury debt issuance.

The relationship of this paper to previous literature is discussed in some detail in Section 2. Briefly, it builds on a long line of models in which the focus is on the adjustment of prices to make investors willing to hold whatever securities are outstanding in each period. The first generation of such models, in the tradition of Tobin (1958), mostly had to do with substitution between money and bonds. These models have been criticized by more-recent authors because they modeled demand for those instruments in an arbitrary, reduced-form way (see for example, Woodford, 2012). A second generation of papers, beginning with Frankel (1985), derived asset-demand functions from investor optimization conditions. While this approach is more consistent with modern finance theory, in the sense that it essentially imposes no-arbitrage conditions, implementations of it have typically had to assume that investors form expectations of future asset prices in an arbitrary, non-rational way in order to obtain analytical solutions. What distinguishes the class of models discussed here from these previous efforts is that investors' demand for assets is determined from their optimization in each period, assuming that they realize that security prices will be determined in the same way in the next period. Unlike Tobin-style models, investors do not care about quantities of particular securities, only about the overall risk of their portfolios. Unlike Frankel-style models, expectations are fully rational. The only paper to date that has successfully characterized the equilibrium in a structural yield-curve model with both of these properties is Vayanos and Vila (2009), which, as noted, is a special case of the class of portfolio-balance models I consider.

Finally, a technical contribution of the paper is to develop a numerical algorithm, iterating on a version of Tauchen and Hussey (1991), to solve for state-contingent prices in this class of models. Solving no-arbitrage portfolio-balance models 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 requires only weak conditions on the functional form of the pricing kernel and the dynamics of the state of the economy. In particular, the approach allows for departures from linearity, normality, and homoscedasticity, and there are essentially no restrictions in modeling the dynamics of the asset-supply distribution itself. An important example of a relevant nonlinearity is the ZLB.

The paper proceeds as follows. Section 2 lays out the general properties of portfolio-balance models and explains the conditions under which changing quantities can affect asset prices under no-arbitrage. It also puts these models in the context of previous literature. Section 2.3 sketches the solution algorithm I use to solve such models, with the details contained in an appendix. Section 3 shows how portfolio-balance applies specifically to the term structure of interest rates and illustrates how it results in duration effects in a calibrated one-factor model, both with and without the ZLB imposed. Section 4 extends this model to several more realistic cases with inflation risk and stochastic bond supply. Section 5 applies the model to study the Federal Reserve's asset purchases, and Section 6 concludes.

2 No-Arbitrage Portfolio Balance

2.1 General Asset-Pricing Relations

I use the term "portfolio balance" to encompass any model in which the equilibrium prices of financial assets are related to the quantities of those assets that investors must hold and in which the quantities can be considered to have an exogenous or policy-dependent component. In particular, suppose there exists a finite number of assets N, with time-t prices $\mathbf{p}_t = (p_{1t}, ..., p_{Nt})$, and par values $\mathbf{X}_t = (X_{1t}, ..., X_{Nt})$. In equilibrium, investor wealth must be equal to the value of all assets: $W_t = \mathbf{X}_t' \mathbf{p}_t$. The idea behind portfolio balance is that 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 supply of securities at each point in time. In the most straightforward cases, the supply of each security is treated as an exogenous quantity issued by a price-insensitive entity, such as the government. More generally, $\mathbf{X}_t$ may itself depend on prices.

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

$$p_{nt} = \mathrm{E}_t \left[M_{t,t+s} q_{nt+s} \right] \tag{1}$$

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

$$p_{nt} = E_t \left[\exp \left[- \int_0^s r_{t+v} dv \right] \right] E_t \left[q_{nt+s} \right] + \mathrm{cov}_t \left[M_{t,t+s} q_{nt+s} \right] \tag{2}$$

where r_t is the time-t instantaneous risk-free rate of interest. (In this paper, I restrict attention to discrete-time models, so r_t will be constant within each period.)

Given dynamics for the state of the economy, no-arbitrage asset-pricing models are distinguished soley by the assumptions they make about the determinants of the SDF. In order for asset quantities to affect asset prices, therefore, $M_{t,t+s}$ must depend on $\mathbf{X}_t$ in some way. While this dependence could in principle take a number of forms, 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. Instead, I consider models in which $M_{t,t+s}$ can be written as a function of the return on wealth $R_{t,t+s}$ and, possibly, of the state of the economy in time t and $t+s$. That is,

$$M_{t,t+s} = M\left(\mathbf{s}_t, \mathbf{s}_{t+s}, R_{t,t+s},\right) \tag{3}$$

where the vector $\mathbf{s}_t$ is the time-t state vector, which is assumed to be exogenous with respect to $\mathbf{X}_t$. By definition,

$$R_{t,t+s} = \frac{\mathbf{X}_t'\mathbf{q}_{t+s}}{\mathbf{X}_t'\mathbf{p}_t} \tag{4}$$

where $\mathbf{q}_{t+s}$=$\begin{pmatrix} q_{1t} & \dots & q_{Nt} \end{pmatrix}$.

Without further restrictions, the response of the price of asset n to an incremental change in the quantity outstanding of asset m is

$$\begin{aligned} \frac{\partial p_{nt}}{\partial X_{mt}} &= \mathrm{E}_t\left[\frac{\partial q_{nt+1}}{\partial X_{mt}} M_{t,t+s}\right] \\ &\quad + \mathrm{E}_t\left[\frac{q_{nt+1}}{W_t}\frac{\partial M}{\partial R_{t,t+s}}\left(q_{mt+s} - R_{t,t+s}p_{mt} + \sum_{k=1}^{N} X_{kt}\left(\frac{\partial q_{kt+s}}{\partial X_{mt}} - R_{t+1}\frac{\partial p_{kt}}{\partial X_{mt}}\right)\right)\right] \end{aligned} \tag{5}$$

While it is not possible to sign this reaction without specifying how payoffs are determined and the additional properties of the pricing kernel, it is clear that the derivative in equation (5) will not be zero in general. Indeed, a sufficient condition for quantities to matter for prices is that the covariance of $M_{t,t+s}$ and q_{kt+s} is nonzero for some k. For example, if we further assume that asset payoffs do not depend on the current portfolio allocation $\mathbf{X}_t$, then it is straightforward to show that $\partial p_{nt}/\partial X_{mt} = 0$ for all n only if $q_{mt+s}/p_{mt} = R_{t,t+s}$—that is, unless the return on asset m is equal to the return on the market portfolio in all states of the world, changing the quantity of m will affect all prices.

The Eggertsson and Woodford (2003) "neutrality proposition" mentioned in the previous section states, in essence, that $\frac{\partial p_{nt}}{\partial X_{mt}} = 0$ in a certain class of models so long as assets are valued only for their pecuniary returns. The reason that this proposition fails here is that the class of models that Eggertsson and Woodford consider are ones in which investor utility is a time-separable function of consumption. Consequently, in their analysis, $M_{t,t+s}$ depends only on consumption in periods t and $t+s$; it does not involve the market return $R_{t,t+s}$. Thus, condition (3) does not hold, and there are no portfolio-balance effects.[4] Yet, while time-separable preferences are clearly an important case to consider, there are equally important models in which (3) *does* hold. The following subsection considers some of these models that have been applied by previous studies.

2.2 Examples from the Literature

One simplification that allows portfolio-balance models to be solved analytically is to suppose that the distribution of asset payoffs is fixed. In this case (further assuming that wealth is predetermined), equation (5) reduces to

$$\frac{\partial p_{nt}}{\partial X_{mt}} = \frac{1}{W_t} \mathrm{E}_t \left[q_{nt+s} q_{mt+s} \frac{\partial M}{\partial R_{t,t+s}} \right] \tag{6}$$

Thus, for example, if the one-period risk-free rate is exogenous and the SDF is linear in $R_{t,t+s}$ with slope coefficient θ, then a change ΔX_{mt} in the quantity of asset m induces a contemporaneous price response in asset n of[5]

$$\Delta p_{nt} = \theta \mathrm{cov}_t \left[q_{nt+s} q_{mt+s} \right] \frac{\Delta X_{mt}}{W_t} \tag{7}$$

Since θ is typically negative (states of the world with higher returns are discounted by more), raising the quantity of asset m lowers the price of asset n whenever the two assets have positively correlated payoffs.

In a classic application of this type of model, Frankel (1985) considered a case in which agents solve a Markowitz portfolio choice problem over a variety of asset types:

$$\max_{\mathbf{w}_t} \mathrm{E}_t \left[\mathbf{R}'_{t+1} \mathbf{w}_t \right] - \frac{a}{2} \mathrm{var}_t \left[\mathbf{R}'_{t+1} \mathbf{w}_t \right] \tag{8}$$

where $\mathbf{R}_{t+1} = \begin{pmatrix} R_{1t,t+1} & \dots & R_{Nt,t+1} \end{pmatrix}$ is the vector of one-period asset returns, a is the coefficient of relative risk aversion, and $\mathbf{w}_t$ is the vector of dollar values allocated to each asset. The maximization is subject to $\mathbf{w}'_t \mathbf{1} \leq W_t$. Mean-variance problems of this type result in an SDF that is linear in the market return. Consequently, one can write:

$$\mathrm{E}_t \left[\mathbf{R}_{t+1} \right] = \exp \left[r_t \right] + a \mathbf{\Sigma}_t \mathbf{x}_t \tag{9}$$

[4] Eggertsson and Woodford are explicit about the consequences of this assumption: "The marginal utility to the representative household of additional income in a given state of the world depends on the household's consumption in that state, not on the aggregate payoff of its asset portfolio in that state. And changes in the composition of the securities in the hands of the public *do not* change the state-contingent consumption of the representative household..." [Emphasis in original.]

[5] The assumption that the short rate is exogenous implies $\partial E_t \left[M_{t,t+1} \right] / \partial R_{t,t+1} = 0$, which allows the product term in equation (6) to be reduced to the covariance term in equation (7).

where $\mathbf{\Sigma}_t$ is the covariance matrix of $\mathbf{R}_{t+1}$ conditional on time-t information and $\mathbf{x}_t = \begin{pmatrix} x_{1t} & \dots & x_{Nt} \end{pmatrix}$ is the vector of asset shares. (I.e., $x_{nt} \equiv X_{nt} / \sum_{m=1}^{N} X_{mt}$.) Since (9) is just an equilibrium condition, it must be true for any value of $\mathbf{x}_t$. Thus, if a policymaker can set $\mathbf{x}_t$ to an arbitrary value, he can determine the expected return on any given asset $\mathrm{E}_t [R_{nt+1}]$. In the case of bonds with known terminal payoffs, this is sufficient to determine period-t prices. Friedman (1986), Engle et al. (1995), and Reinhart and Sack (2000) are among the subsequent studies to exploit this relationship. More recently, Gromb and Vayanos (2010) illustrate the effects of quantities on prices in version of this problem in which there are only two assets, and Neeley (2010) and Joyce et al. (2011) apply it specifically to the study of central bank asset purchases.

While this approach is useful for studying certain problems, its main difficulty is that treating one-period-ahead asset payoffs as exogenous is not realistic. In most cases of practical interest, the payoff of an asset s periods ahead includes the price of that asset s periods ahead, and that price should be determined in the same way as prices today. In terms of equation (5), there is no reason to expect $\frac{\partial q_{nt+1}}{\partial X_{mt}}$ to be equal to zero, particularly if there is persistence in the value of $\mathbf{X}_t$ across periods. One way this dependence manifests itself is that the conditional covariance matrix of returns $\mathbf{\Sigma}_t$ is endogenous. But, for example, when taking these models to the data, Frankel (1985) and subsequent authors calibrate $\mathbf{\Sigma}_t$ to a covariance matrix based on the historical returns data. This is tantamount to assuming that investors do not change their beliefs about future prices when $\mathbf{x}_t$ changes, even though the model itself implies that these prices will be different. While it is possible that investors do not account for the effects of quantities on return variances when determining asset prices, doing so in the context of these models represents a violation of rational expectations. There is nothing logically wrong with performing policy hypotheticals using equation (9), but rational expectations requires us to solve for $\mathbf{\Sigma}_t$ as a function of $\mathbf{x}_t$ before doing so.

Another important class of models in which portfolio-balance effects are potentially operative are those involving recursive preferences. In particular, consider a representative investor with Epstein-Zin-Weil utility over real consumption. As Epstein and Zin (1989) showed, the SDF in this case can be written as

$$M_{t,t+s} = \left(\beta \frac{c_{t+s}}{c_t}^{-\frac{1}{\psi}} \right)^{\theta} \left(R_{t,t+s} - \pi_{t,t+s} \right)^{1-\theta} \tag{10}$$

where c_t is consumption, $\pi_{t,t+s}$ is the gross rate of inflation between periods t and $t+s$, and θ and ψ are utility parameters. This SDF has the form of (3), and so the portfolio-balance channel applies. Indeed, Piazessi and Schneider (2008) study the relationship between bond quantities and prices in this type of model. Again, however, they do not model expectations as rational. Instead, to solve the model, Piazzesi and Schneider assume that investors have adaptive expectations over the asset-return distribution, which they estimate using survey data and vector autoregressions. As they argue, there may well be good reasons to believe that investors form expectations in this way. However, it is also of interest to consider the case in which expectations are set rationally. For a given vector of asset shares $\mathbf{x}_t$, it is evidently not possible to solve analytically for the state-contingent price vector that satisfies both equation (4) and equation (10) in all states of the world.

2.3 Solution Method

As illustrated in the cases above, it is not generally possible to solve portfolio-balance models analytically for asset prices as functions of quantities under rational expectations. The special cases in which it is possible (for example, strictly exogenous one-period payoffs or risk neutrality) are typically not of practical interest. The central difficulty is that the solution for prices involves the moments of future prices, and, under rational expectations, these moments themselves must be endogenous. 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 with portfolio-balance models 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 p_t using an iterative, discrete–state approximation 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 zero lower bound.

I first make explicit that prices and quantities depend on the state of the economy. Namely, let $x_n(\mathbf{s}_t)$ describe how the quantity of asset n depends on the state. Let $\mathbf{s}_t$ be Markov on the support $\mathbf{S}$ with transition density $\tau(\mathbf{s}_{t+1}|\mathbf{s}_t)$. It is assumed that the form of the pricing kernel in equation (3), the laws of motion for the states, and the dependence of quantities on the states are known—that is, we (and investors) have knowledge of the functions $\tau(\mathbf{s}_{t+1}|\mathbf{s}_t)$, $M(\mathbf{s}_t, \mathbf{s}_{t+1}, R_{t,t+1})$, and $x_n(\mathbf{s}_t)$. We seek a vector-valued function $\mathbf{p}(\mathbf{s}_t) = \begin{pmatrix} p_1(\mathbf{s}_t) & \ldots & p_N(\mathbf{s}_t) \end{pmatrix}$ that describes how all asset prices depend on $\mathbf{s}_t$.

The price of asset n is given by

$$p_n(\mathbf{s}_t) = \int_{\mathbf{S}} \tau(\mathbf{s}'|\mathbf{s}_t) M(\mathbf{s}_t, \mathbf{s}', R_{t,t+1}) q_n(\mathbf{s}') d\mathbf{s}' \tag{11}$$

where the integral is taken over all dimensions of the state and $\mathbf{q}(\mathbf{s}_t) = (q_1(\mathbf{s}_t)$ ... $q_N(\mathbf{s}_t))$ is the vector of functions determining asset payoffs in state $\mathbf{s}_t$. This function may be known (if payoffs are strictly exogenous) or, more generally, dependent upon the unknown pricing funciton $\mathbf{p}(\mathbf{s}_t)$. While the algorithm described below applies in principle to any asset, the case of zero-coupon Treasury bonds considered in this paper is equivalent to the further assumption on the payoff function:

$$q_n(s_t) \equiv \begin{cases} p_{n-1}(\mathbf{s}_t) \text{ for } n > 1 \\ 1 \text{ for } n = 1 \end{cases} \tag{12}$$

where the index n is then interpretable the bond's maturity.

Given a distribution for the market return, $R_{t,t+1}$, (11) is a system of linear Fredholm equations of the first kind, which in principle can be discritized and solved by quadrature in one step (see Tauchen and Hussey, 1991). However, the fact that $R_{t,t+1}$ is defined as in (4) requires us to iterate by, first, solving (11) using a given distribution of $R_{t,t+1}$, and, second, given the resulting pricing fuctions finding the updated distribution of $R_{t,t+1}$. These steps can be repeated to convergence.

The specifics of this algorithm are discussed in Appendix A. Loosely speaking, it can be summarized as follows:

1. Guess a function $\mathbf{p}^i(.)$ such that $\mathbf{p}_t = \mathbf{p}^i(\mathbf{s}_t)$ on a discretization of $\mathbf{S}$.

2. Based on this function and the known law of motion for $\mathbf{s}_t$, compute (the discrete approximation to) the distribution of $R_{t,t+1}$, including its covariance with $\mathbf{s}_{t+1}$ and $\mathbf{p}_{t+1}$.

3. Using that distribution, solve for the updated function $\mathbf{p}^{i+1}(\mathbf{s}_t)$ via equation (1) and return to step 2.

As illustrated in the appendix, the algorithm converges quickly for the type of model considered in this paper. An additional modification, also discussed in the appendix, is necessary to handle the models that involve recursive preferences with endogenous consumption, since in those cases the pricing kernel depends on the state in a way that is unknown *ex ante*.

3 Portfolio Balance in the Term Structure

3.1 Preliminaries

While the above discussion applies to any set of assets in principle, I now focus the analysis on the study of zero-coupon, default-free bonds. This is an important case to consider, not just because of its policy relevance with respect to asset purchases and debt management, but also because these bonds, with known terminal payoffs and being uncontaminated by credit and liquidity risk, help to isolate the workings of the portfolio-balance channel. As noted, restricting attention to the pricing of default-free bonds does not necessarily imply that we assume that investors hold *only* those bonds.

Henceforth, I maintain assumption (12) so that p_{nt} indicates the time-t price of a bond with n periods to maturity. By definition, $p_{1t} = \exp[-r_t]$ is the price of the one-period bond, and equation (2) determines the rest of the yield cuve. In particular, perfect certainty corresponds to the "strong form" of the expectations hypothesis—long-term interest rates are the average of expected short-term interest rates (up to a Jensen's inequality term). Otherwise, the covariance term in equation (2) adds a risk premium to expected returns and a term premium to bond yields. The SDF M only appears in this covariance; thus, since I assume that r_t is set exogenously, quantity changes operate by affecting term premia in the presence of uncertainty.[6]

Restricting attention to bonds in this way, portfolio-balance effects take on a special meaning—bond quantities affect bond prices through a *duration channel*. That is, tilting the distribution of outstanding assets toward lower maturities alters the term structure by "removing duration from the market," precisely the effect that is often pointed to, in less formal terms, as the primary mechanism through which asset purchases operate. (See, for example, Gagnon et al., 2010.) To date, the only paper to capture such duration effects in a theoretical model that features both no-arbitrage restrictions and rational expectations is Vayanos and Vila (2009). In that model, arbitrage investors solve a problem like (8), but they interact with "preferred-habitat" agents who have downward-sloping demand curves for assets of particular maturities. Thus, the portfolio shares $\mathbf{x}_t$ that they hold are endogenous. Assuming a linear-Gaussian factor structure, Vayanos and Vila obtain an affine solution in which demand shocks that remove long-term debt from the hands of arbitrageurs push down long-term interest rates.

As the only model to incorporate supply effects into an otherwise standard representation of the term structure, Vayanos-Vila has been highly influential in the way that economists have designed and interpreted recent empirical studies.[7] However, the presence of the preferred-habitat agents and the endogeneity of $\mathbf{x}_t$ in that model can obscure the fundamentally simpler relationship between prices and quantities. The example below strips this mechanism to its essentials to show how duration effects work. (Appendix B shows how Vayanos and Vila's model fits into the broader portfolio-balance framework.)

3.2 Example: A linear, one-factor model

Consider a model in which the short rate r_t is the only source of stochastic variation and in which investors solve the mean-variance portfolio-choice problem (8), where the available assets consist only of Treasury bonds with maturities $1, ..., N$. The relative supply of bonds is fixed over time, $\mathbf{x}_t = \mathbf{x}$. Bond prices then solve the system of quadratic equations

$$\mathbf{p}_t = \exp\left[-r_t\right]\left(E_t\left[\mathbf{q}_{t+1}\right] - a\frac{\mathbf{\Omega}_t \mathbf{x}}{\mathbf{p}_t\prime\mathbf{x}}\right) \tag{13}$$

at each point in time, where $\mathbf{\Omega}_t$ is the conditional covariance matrix of the asset payoffs $\mathbf{q}_{t+1}$. (This is the equivalent of equation (2) for the Markowitz model.) Since the payoff on a one-period bond is always 1, the first row and column

[6] It is not necessary to assume r_t exogenous in order to generate portfolio-balance effects or to apply the solution method. However, this assumption is common in both the term-structure and monetary-policy literatures.

[7] See, for example, Doh (2010), Hamilton and Wu (2011), and Li and Wei (2012).

of $\mathbf{\Omega}_t$ are zeros. The conditional variance of the price of a one period bond, denoted ω_{1t}^2, is the second diagonal element of $\mathbf{\Omega}_t$. Since $p_{1t} = \exp[-r_t]$, this variance is a determinsitic function of the short-rate process and is a known and exogenous quantity.

In this model, the expectations hypothesis holds if $\mathbf{x} = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$. That is, eliminating all duration from investors' portfolios reduces the term premium to zero. On the other hand, as long as $a > 0$ and $x_1 < 1$, we have $p_{nt} < p_{1t}E_t[p_{nt+1}]$, for $n > 1$, so that all risky assets require a discount and all risky returns include a risk premium. Indeed, since r_t is the only source of variation, the term premium in this example reflects *only* the compensation for bearing duration risk. Specifically, there is a single market price of risk given by the Sharpe ratio

$$\lambda_t = \frac{\mathrm{E}_t\left[R_{nt,t+1}\right] - R_{1t,t+1}}{\sigma_{nt}} \tag{14}$$

where σ_{nt} is the standard deviation of the return on asset n. This, in turn, implies that all asset prices and returns are perfectly correlated, regardless of the value of $\mathbf{x}$ or the dynamics of r_t. In particular, there exist coefficients A_{nt} and B_{nt} such that we can write $R_{nt,t+1} = A_{nt} + B_{nt}p_{1t+1}$ for any n.[8] The coefficient B_{nt} represents the response of the return on a risky asset to the risk-free rate—the analogue of the "spot-rate duration" that appears in continuous-time term-structure models. The standard deviation of the return on the market portfolio can be expressed as $\omega_{1t}\mathbf{x}'\mathbf{B}_t$, where $\mathbf{B}_t$ is the vector of B_{nt} coefficients. Combining this with equation (14), we have

$$\lambda_t = a\omega_{1t}\mathbf{x}'\mathbf{B}_t \tag{15}$$

analogous to a result produced by Vayanos and Vila (2009). Reducing the quantity $\mathbf{x}'\mathbf{B}_t$ is the model's version of "removing duration from the market." Doing so will lower both the total risk of the Treasury portfolio and the price of that risk.[9]

To implement the model numerically, take periods to be one year in length and suppose that there are $N = 30$ bonds. Suppose that the short rate follows a linear process

$$r_t = \phi_0 + \phi_1 r_{t-1} + \varepsilon_t \tag{16}$$

where ε_t has variance σ^2. For parsimony, I approximate the maturity structure of government debt with an exponential distribution:

$$x_n \propto \exp\left[-zn\right] \tag{17}$$

[8] Specifically, $B_{nt} \equiv \sigma_{nt}/\omega_{1t}$ and $A_{nt} \equiv (1 - p_{2t}B_{nt})/p_{1t}$. Because r_t follows an AR(1) process, the function B_{nt} has a similar shape across n at each value of the short rate as in the Vasicek (1977) model.

[9] Since (13) is the result of a Markowitz portfolio problem it also has a conditional CAPM representation. Specifically, since $\mathrm{E}_t[R^w_{t,t+1}] = \mathrm{E}_t[\mathbf{q}_{t+1}]'\mathbf{x}/\mathbf{p}_t'\mathbf{x}$ is the expected return on wealth, we can write

$$E_t[\mathbf{R}_{t+1}] - R_{1t,t+1}\mathbf{1} = \boldsymbol{\beta}_t\left(E_t[R_{t,t+1}] - R_{1t,t+1}\right)$$

where $\boldsymbol{\beta}_t = \mathbf{\Omega}_t\mathbf{x}/\mathbf{x}'\mathbf{\Omega}_t\mathbf{x}$. The expected excess return on the market portfolio is $E_t[\mathbf{R}_{t+1}] - R_{1t,t+1} = a\mathbf{x}'\mathbf{\Omega}_t\mathbf{x}$.

where the parameter z is the average maturity outstanding. The exponential approximation is likely to be a good one because the distribution of government liabilities typically consists overwhelmingly of shorter maturities. For example, as shown in the top panel of Figure 1, liabilities with less than five years of duration (including currency and reserves) account, on average, for about 80 percent of total face value. The exponential functional form implies that, as in the data, the distribution is highly skewed toward securities with very low duration.[10] However, as discussed further below, the exact shape of the distribution beyond its first moment is actually of little importance.

Let $\phi_0 = 0.003$, $\phi_1 = 0.95$, $\sigma = 0.015$, and $a = 8$. These values are calibrated roughly to match the dynamics of the short-rate and the average value of the 15-year yield in the data. The black line in Figure 2 demonstrates how the duration of the Treasury portfolio affects the behavior of asset prices in the model by plotting the equilibrium values of the (negative) price of risk λ_t across possible values of z in this model. The relationship between duration and risk prices is monotonic but nonlinear. Near the historical mean duration of 2.7 years (see the bottom panel of Figure 1), increasing duration by one year raises the price of risk by about 50%.

Figure 3 illustrates how these risk prices translate into yields. I compare two cases: $z = 2.7$ years and $z = 2.0$ years. As discussed in Section 5, the difference of -0.7 years is approximately equivalent to the effects of the cumulative asset purcahses by the Fed through 2012. The difference in the maturity distribution of securities is illustrated by the red bars in the top panel of the figure. The effect on the yield curve, evaluated at the sample average value of the short rate, is shown in the middle panel. For the calibration used here, the ten-year yield is 56 basis points lower under the lower-duration distribution. Again, this is due to reductions both in the price of risk illustrated in the previous figure and in risk itself. The latter is shown in the bottom panel of Figure 3, which plots the standard deviation of one-year returns (the square-root of the diagonal of Σ_t) under the two supply distributions.

Finally, we can demonstrate the earlier claim that the shape of the distribution $\mathbf{x}$ is of little importance for the shape of the yield curve. Analytically, equations (13) and (15) show that the individual asset share x_n does not matter for the individual asset price p_{nt}. Only the weighted sum of asset shares $\mathbf{x}'\mathbf{B}_t$ is relevant, and it affects all prices in the same way. Consequently, this model cannot produce local-supply effects from large quantity gluts or shortages in particular sectors of the market. To illustrate this, consider a stark alternative to the exponential shape used for the maturity distribution above. In particular, suppose that the distribution was nearly degenerate around the same mean of $z = 2.7$, as shown in the top panel of Figure 4. This adjustment does slightly increase the duration risk of the portfolio, primarily because it completely eliminates the risk-free one-year bonds from investors' portfolios. However, after recalibrating a to match the average slope of the previous yield curve, there

[10] The data in this figure are compiled from CRSP, taking the duration of all Treasury debt in the hands of the public, weighted by its face value. In addition, I include the monetary base, as reported by the Federal Reserve, as a "Treasury" security with zero duration.

is virtually no difference between the term structures generated under the two distributions. This can be seen in the bottom panel of the figure, where the solid blue line is the same yield curve shown in Figure 3, and the dashed red line is the (recalibrated) yield curve at the same value of the short rate under the degenerate supply distribution. Similar results obtain for other possible choices of the shape of x.

3.3 Effect of the Zero Lower Bound

Since risk prices in portfolio-balance models 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 zero 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. Consequently, investors in all bond maturities face less near-term interest-rate risk and therefore demand lower risk premiums. Furthermore, equation (15) shows that the effect on risk prices of removing duration from the market (in the sense of reducing the weighted average $\mathbf{x}'\mathbf{B}_t$) is directly proportional to the variance of the one-period bond ω_{1t}. Thus, the reduction in the volatility of short-term interest rates induced by the ZLB will also dampen duration effects.

To illustrate these outcomes in the simple one-factor model above, suppose that, rather than allowing the short rate to follow an unrestricted AR process, we append the condition

$$\Pr\left[\varepsilon_t < 0\right] = 0 \quad \forall t \tag{18}$$

to equation (16). That is, let the distribution of the short-rate shock ε_t be a truncated normal, where the truncation point varies with r_t in such as way that it always enforces the zero lower bound. The conditional distribution of the short rate under this process is illustrated in Figure 5. The reduction in volatility near zero is evident.

Refering back to Figure 2, the blue line illustrates risk prices calculated at the ZLB for different values of duration; on average, they are roughly half of the linear case considered previously. Furthermore, the slope of the blue line is considerably flatter than that of the black line, indicating that a given change in duration has a smaller effect on λ_t when the ZLB binds. To see these results in terms of the yield curve, the solid lines in the top panel Figure 6 correspond to the same maturity distributions that were used in Figure 3. (2.7 years for the blue line and 2.0 years for the red.) The same duration–reduction experiment that caused a 56 basis point reduction when the short rate was at 5.8% causes only a 29 basis point reduction when the short rate is at 0% with a binding ZLB.

The reduction in duration effects becomes even more pronounced if the short rate is anticipated to stay near zero with a high degree of certainty for multiple periods, because this results in a larger reduction in the variance of rates across the term structure. This type of thought experiment is relevant because the forward-guidance language that has been included in most FOMC statements since 2008 is widely viewed as having essentially committed the Fed to keeping rates near zero for a significant time into the future. To illustrate, the dashed lines in the figure show the model's outcomes when r_t is constrained to equal 0 for both periods t and t+1 (and investors know this with certainty at time t) and only then follows its truncated AR(1) process. The forward guidance itself has a large effect on long-term rates; in this example, it reduces the 10-year yield by 52 basis points (the difference between the solid and dashed blue lines). However, with the forward guidance in place, the effect of the reduction in duration on the ten-year yield is only 23 basis points, rather than 29. The reason is that, as illustrated in the bottom panel, forward guidance in the one-factor model implies that the returns on all bonds in period $t+1$ are known with certainty. Consequently, the risk premium for the period over which the forward guidance is in place is zero, regardless of the amount of duration outstanding—i.e., ω_{1t} is zero, and so, by equation (15), the price of risk λ_t is zero. Thus, although they both serve to reduce longer-term yields, forward guidance and asset purchases are somewhat offsetting, at least if duration effects are the only channel through which the supply distribution matters for prices.

4 Multi-Factor Models

While the simple one-factor model is useful for illustrating portfolio-balance effects, it is not particularly realistic. This section considers several extensions of that model. In particular, I consider models that include inflation risk, recursive preferences over consumption, and the possibility that investors maintain wealth in assets other than Treasury securities. In addition, I allow the duration of Treasury securities itself to be a stochastic variable over which investors must form expectations. The purpose of analyzing these models is to get a sense of the range of possible magnitudes for the effects of Treasury supply changes, as well as to illustrate the flexibility of the general approach.

To more cleanly compare the effects of different assumptions about the pricing kernel, I take the dynamics of the state of the economy to be identical across all five models. Specifically, I assume that inflation and the short-term interest rate follow a VAR(1) process truncated at zero for the short rate. I also assume that the distribution of outstanding Treasury securities changes over time. To allow for such changes in a parsimonious way, I maintain the exponential maturity distribution (17), but I allow the parameter z to follow a first-order autoregressive process. This induces (nonlinear) dynamics into all elements of the asset-share vector $\mathbf{x}_t$.

4.1 Model Specification

4.1.1 Model 1 - Quadratic nominal preferences with Treasury-only portfolio

The first model retains the quadratic nominal preferences from the example of the previous section (maintaining the pricing equation (13)); it differs only in the dimension and dynamics of the state vector. In this model, investors do not care about inflation *per se*, since their objective function is over nominal returns. However, inflation may still be important to the extent that the nominal short rate responds to it. The VAR dynamics I assume for the short rate and inflation allow for this possibility. In the other models below, inflation itself will play a direct role in investors' payoffs, and so this process maintains comparability across models in the exogenous dynamics of the economy.

Model 1 is similar in most respects to that of Greenwood and Vayanos (2014). The principal differences are the addition of the inflation term and the imposition of the ZLB.[11]

4.1.2 Model 2 - Quadratic real preferences with Treasury-only portfolio

Model 2 allows investors to care about real returns, rather than nominal. In this case, maintaining quadratic preferences, the investors' decision problem becomes

$$\max_{\mathbf{w}_t} \mathrm{E}_t \left[\mathbf{R}'_{t,t+1}\mathbf{w}_t - \pi_{t,t+1}\right] - \frac{a}{2}\mathrm{var}_t \left[\mathbf{R}'_{t,t+1}\mathbf{w}_t - \pi_{t,t+1}\right] \tag{19}$$

Prices then solve

$$\mathbf{p}_t = \exp\left[-r_t\right] \left(\mathrm{E}_t \left[\mathbf{q}_{t+1}\right] - a\frac{\mathbf{\Omega}_t \mathbf{x}_t}{\mathbf{p}_t \prime \mathbf{x}_t} - cov_t \left[\mathbf{q}_{t+1}, \pi_{t+1}\right]\right) \tag{20}$$

and the nominal pricing kernel is conditionally linear in the real return $(R_{t,t+1} - \pi_{t,t+1})$. Note that changes in $\mathbf{x}_t$ have no direct effect on inflation-risk premia through portfolio rebalancing. While they may change the covariance of prices with inflation through their effect on the overall volatility of asset prices, these effects will generally be second-order.

4.1.3 Model 3 - Recursive preferences with Treasury-only portfolio

I next consider a model with Epstein-Zin-Weil preferences over consumption (Model 3). Such models have the pricing kernel given in equation (10). Unlike

[11] There are three other, minor differences between this model and Greenwood-Vayanos. First, the Greenwood-Vayanos model is in continuous time. Second, I use the exponential form in (17) for the supply distribution, whereas they use a linear-factor structure. Finally, they assume that market-value weights w_{nt} are exogenous, while I assume that the par values x_{nt} are exogenous. None of these differences in approach is likely to have substantial effects on the model outcomes.

the quadratic preferences of Models 1 and 2, this speficiation involves consumption growth. While one could adopt an exogenous process for consumption, it turns out that the exogenous dynamics for the state variables r_t, π_t, and $\mathbf{x}_t$ that we have already specified are sufficient to pin down the dynamics of the quantity $\beta \frac{c_{t+s}}{c_t}^{-\frac{1}{\psi}}$ in equation (10), if we assume that consumption can be approximated as a trend-stationary process. Again, proceeding in this way has the advantage that it maintains the same state vector and dynamics across all models, isolating the differences that are due to assumptions about the pricing kernel alone. The solution requires an additional step the solution algoritm, which is described in Appendix B. It does not require us to take a stand on the values of β and ψ or the precise form of the consumption process.

4.1.4 Model 4 - Quadratic real preferences with diversified portfolio

A major criticism of all of the above models is that they assume that government debt is the only asset available to the public. Furthermore, to the extent that participants in the bond market also bear the tax burden of repaying the bonds, they may not function as assets at all, as discussed in Barro (1974) and Wallace (1981). A more flexible and realistic model should allow for Treasuries to receive a weight of less than 1 in the representative investor's portfolio. Models 4 and 5 consider an extension of the real-quadratic-preference and recursive-preference models, respectively, that allows for this possiblity.

In particular, suppose that Treasuries constitute a fraction v of the market value of investor assets. Then, rather than the definition (4), we have

$$R_{t,t+1} = v(\frac{\mathbf{x}_t\prime\mathbf{q}_{t+1}}{\mathbf{x}_t\prime\mathbf{p}_t}) + (1-v)R^w_{t+1} \tag{21}$$

where R^w_{t+1} is the return on the non-Treasury portfolio. For simplicity, I assume that R^w_{t+1} is iid. Thus, R^w_t acts as a fourth state variable in the models in which it appears, but its time-t level contains no information about its time $t+1$ level. Consequently, the yield curve is not affected by shocks to R^w_{t+1}, even though the term premium does provide compensation for the risk associated with these shocks to the extent that they are correlated with bond returns. The previous models, which implicitly take $v = 1$, essentially represent an upper bound on the potential size of portfolio-balance effects in the term structure. For any given calibration, as v approaches zero, those portfolio-balance effects must approach zero as well.

4.1.5 Model 5 - Recursive preferences with diversified portfolio

The final model also takes the Treasury share in the portfolio to be $v < 1$ so that the return on wealth is given by (21), but it assumes recursive preferences as in Model 3.

4.2 Calibration

The parameters of the truncated VAR process for inflation and the short rate are calibrated to the dynamics of the annual U.S. data over the period 1971 - 2008 and are shown in Table 1.[12] Estimating the AR(1) process for z_t on the data shown in Figure 1 gives an intercept of 0.16, an autoregressive parameter of 0.95, and a standard error of 0.14 (also shown in the table), and these are the values used in the calibration. The assumption that z_t is contemporaneously and dynamically uncorrelated with r_t and π_t is both analytically convenient — since we will want to isolate the effects of exogenous shocks to z_t — and realistic. In the data, those correlations are very close to zero.

In both the quadratic and EZW models, the pricing kernel depends on a single parameter (either a or θ). In each case, I calibrate the value of this parameter such that the 1- to 15-year slope of the yield curve is 130 basis points on average, given the observed values of inflation and the short rate over the 1971 – 2008 period, consistent with the average yield-curve slope observed in the data during that period.

Models 4 and 5 require a calibration for the Treasury-wealth share v. Given the various considerations that might affect segmention and exposure to Treasuries, it is unclear what the appropriate value of v is. Since models 1 through 3 give us results at one extreme (with v assumed equal to 1), for illustration in models 4 and 5, I calibrate v to a relatively low value of 0.1. I calibrate the distribution of R^w to the results of Lustig et al. (2013), who estimate that the excess annual return on wealth averages 2.4% with a standard deviation of 9.2%.

4.3 Results

4.3.1 Model moments

Table 3 reports some summary statistics for the five models. Given the calibration to the average level of 15-year yields, the models with quadratic preferences fit the short end of the yield curve better than the models with recursive preferences, which imply too much curvature. Both types of preferences generate too little volatility in long-term yields, a manifestation of the "excess volatility" puzzle for bonds, although the recursive preference models do slightly better in this dimension than the quadratic-preference models. Partly because of low volatility, both types of preferences understate the average excess returns on bonds.

In the data, yields themselves are only weakly correlated with the average duration of outstanding government debt. All five models also display fairly weak correlation with duration in yields. Annual bond returns, on the other hand, are approximately 30% correlated with duration in the data. The models also broadly replicate this fact, with the quadratic-preference models having

[12] Yield data are the Gurkaynak et al. (2007) zero coupon yields maintained at the Federal Reserve Board. The sample period begins in 1971 because that is the date at which 15-year yields become available. Inflation is measured by the core PCE index.

slightly higher correlations than the recursive-preference models. The weakest correlations of duration with both yields and returns are produced by Model 5. Intuitively, assuming a low share of Treasury bonds in the portfolio implies that the total Treasury duration that investors must bear is unimportant. Model 3 also assumes a small share of Treasuries in the overall portfolio, but, in that model, risk aversion is calibrated to a very high value of $a = 75$ (see Table 1) which boosts the correlation between duration and returns. Indeed, the reason for the large calibrated value of a in Model 3 is that the low amount of duration risk that investors bear in that model requires them to be very risk averse in order to generate a sufficient term premium to match the average empirical yield-curve slope. (The recursive-preference model does not require this because it allows the consumption process to adjust to create additional risk for the investor.)

4.3.2 Supply effects on term premia

Figure 7 displays impulse-response functions for the yield curve in all five models, following a one-standard-deviation positive shock to duration (increasing the outstanding duration by 0.14 years). In each of these graphs, maturity is plotted along the lower-left axis, and calendar time is ploted along the lower-right axis. Because the models involve nonlinearities, initial conditions matter somewhat for these outcomes, and in each case they are evaluated at the unconditional means of the state variables implied by the parameters in Table 1. Note that since the short rate and inflation do not respond to Treasury supply by assumption, there is no effect of a shock to z on the expected path of rates. Thus, the responses shown in the figure are due exclusively to changes in term premia.

In all of the models, the response functions have the same basic shape, increasing relatively sharply at shorter maturities and peaking in the middle of the curve. Because z_t is very persistent, these responses decay only slowly over the twenty years following the hypothetical shock, and at about the same rate across all models. Magnitudes vary somewhat across models. Responses for the quadratic models with preferences over real returns (models 2 and 4) are nearly identical and are a bit smaller than the model with nominal preferences. This reflects the fact that, in the nominal model, duration risk is the only source of risk. With real preferences, inflation risk matters, and duration shocks have less volatility to account for. (This is why the risk aversion coefficient for Model 1 is calibrated to 9.5, while that for Model 2 is calibrated to 7.2.) Nonetheless, the responses in Models 1 - 4 are quantiatively similar (around 3 basis points for the initial response of the ten-year yeild), while the response in Model 5 is an order of magnitude smaller.

In general, the response of the yield curve to stochastic duration shocks is not monotonic across maturities. The reason is that it conflates two effects. On the one hand, investors' duration exposure—and therefore the price of duration risk—goes up in the period of the shock, raising the risk premium required on long-term bonds by more than that on short-term bonds. On the other hand, since z_t is mean-reverting, duration is expected to be lower in future periods. Thus, long-term bonds are likely to face lower risk prices on average over their lifetimes than short-term bonds. (In the one-factor model, the duration shock illustrated in Figures 3 and 6 was assumed to be permanent, and the response therefore *was* monotonic across maturities.) Figure 8 illustrates the consequences of the duration persistence assumption in Model 1 by considering the effect of the same shock to duration when the autoregressive coefficient for z_t is calibrated to be zero instead of 0.95 (holding the unconditional mean of the z_t process constant). The response of the two-period yield is the same in both cases, since two-period bondholders are unaffected by the amount of duration in the market beyond the current period. However, overall the response to the shock when z_t is iid is much lower at all other maturities than when z_t is persistent. In addition, the magnitude of the response peaks at about the four-year maturity, as the anticipated mean-reversion dominates the temporary increase in duration risk beyond relatively short horizons.

Consequently, it is not just the *size* of a shock to Treasury duration that matters. The anticipated persistence of that shock is also very important for both the magnitude and the shape of the term-structure response. This point could be relevant for assessing the effects of central bank asset purchases, which may have a lifetime that is shorter than the typical shock to Treasury maturity structure.

Table 5 puts these results into further perspective by applying the models to the most-recent historical episode in which there was a large change in Treasury maturity structure. In particular, as was illustrated in Figure 1, between the years 1976 and 1988, Treasury duration (including base money) in the hands of the public increased from about 1.5 years to about 3.5 years. The first column of the table shows what each of the five models predict for the ten-year term premium, given a shock to z_t of this amount, assuming that inflation and the short rate are at their average values over the 1976-1988 period. In all cases except Model 5, the predicted response is about a 50 basis point increase in the ten-year term premium. Using the Kim and Wright (2005) model, Rudebusch et al. (2007) do indeed document an increase in term premiums estimated from unobseved-factor models around this time.[13] Although their estimates point to an even larger increase (about 100 basis points on net over this period), the results in Table 5 suggest that the increase in Treasury duration could have been a significant contributing factor.

[13] Other models reduced-form term-structure models, such as that of Adrian et al. (2013) also generate an increasing term premium during this period, as do estimates based on survey data. See, for example,
http://libertystreeteconomics.newyorkfed.org/2014/05/treasury-term-premia-1961-present.html#.VFEFwBawXv4.

5 Duration Effects of the Federal Reserve's Asset Purchases

I now consider the effects of the Federal Reserve's large-scale asset purchases (LSAPs) that took place between 2008 and 2012 within the context of the five models developed in the previous section. While calculating the overall effect of these operations on the oustanding duration of government debt is not straightforward, I approximate the magnitude based on the aspects of the programs are summarized in Table 6. One difficulty is that a large component of the Federal Reserve's purchases consisted of agency mortgage-backed securities. Throughout the paper I have ignored agency MBS when calculating the duration of government liabilities, but this exclusion may be inappropriate given that these securities have typically been perceived to carry an implicit government guarantee. Two technical measurement problems arise when trying to extend the analysis to agency MBS. The first is that comprehensive data on the maturity structure of outstanding MBS are not available. The second is that, unlike the case of zero-coupon Treasuries, duration for MBS depends on economic conditions—in particular, it depends on the level of interest rates through the negative convexity induced by the prepayment option. However, according to the Barclays MBS index, the average duration of MBS since 1989 is about three years, only slightly lower than that of Treasury debt. (See Hanson, 2012.) In performing the calculations below, I assume that asset purchases did not materially change the average duration of MBS outstanding, but I do incorporate the amount of MBS outstanding, so that Fed purchases swap their duration for zero-duration reserves.[14]

The net effect of the LSAP programs, as of December 2012, was to reduce the outstanding supply of Treasury and MBS securities by $2 trillion and increase reserves by a similar amount. Moreover, the Treasury securities removed from the market were, after the duration twist induced by the Maturity Extension Program (MEP), almost all of maturity greater than five years. Consequently, the average duration of Treasury securities in the hands of the public after all was said and done was about 0.5 years lower than it otherwise would have been. Taking all of these facts together, we can estimate how investors' duration changed as a result of the programs. Specifically, as of Q4 2008, Treasury debt held by the U.S. public was $5.3 trillion, agency MBS was $5.0 trillion, and the monetary base was $1.4 trillion. The average duration of Treasuries (excluding the monetary base) in the hands of the public was 3.1 years as of the beginning of the program. Therefore, assuming that MBS have an average duration of three years (and again counting the base as zero duration), the weighted average of Treasuries, MBS, and the monetary base in the hands of the public was 2.7 years when the first LSAP was announced. Holding all else constant, this value would have fallen to 2.0 years as a result of the programs.[15]

[14] Data on MBS quantities come from SIFMA (www.sifma.org/research/statistics.aspx).

[15] I ignore the effect of the Fed's $165 billion of agency debt purchases on the outstanding duration of government securities, although the resulting increase in reserves is included.

The final column of Table 6 presents the range of values for the empirical effects of the LSAPs on the ten-year term premium, culled from the literature that has examined this question. In particular, this range draws on estimates from Gagnon et al. (2011), Krishnamurthy and Vissing-Jorgensen (2011), Ihrig et al (2012), D'Amico et al. (2012), and Rosa (2013). (In some cases, additional minor calculations were required to make the results comparable.) Of course, all of these estimates are subject to a high degree of uncertainty surrounding both parameter values and specification. Furthermore, since most of these papers employ an event-study methodology using program-announcement dates, they are more likely to understate than overstate the effects, to the extent that they do not adequately capture expectations of purchases prior to announcement dates or any additional effects that may be priced in when the purchases actually occurred. Nonetheless, they provide a rough guide to what we should expect for the combined effects of the programs—likely on the order of 100 to 250 basis points altogether. To be clear, this is (according to the authors of the studies) only the effect on the term premium, controlling for changes in expectations of the short rate, and it only reflects the initial impact of the programs, ignoring any subsequent dynamic effects.

How do these estimates compare to what would be predicted by portfolio-balance models? The second column of Table 5 shows the effect of removing 0.7 years of duration from the hands of investors, starting from a level of duration that is equal to what was observed prior to the LSAP's introduction (2.7 years) and a configuration of short rates and inflation that approximates the situation on average since that time. Specifically, inflation is set at a level of 1.4 percent (its average over the period 2009 – 2012), and the short-term interest rate is taken to be near the zero lower bound with two years of forward guidance in place. Although the market's perception of the length of time that the ZLB would bind fluctuated over this period, the assumption of two years is likely close to the average that prevailed (see Femia et al., 2013).

The effect of the LSAPs on the ten-year yield of the LSAP shock in this environment is at most 13 basis points, in the model with quadratic, nominal preferences (Model 1). In the recursive-preference model with diversification (Model 5), it is a mere 1 basis point. Note that this is the initial impact of the shock—given the estimated dynamics of z_t, the effect would decay to zero over time, similarly to the response shown in Figure 7.[16] Of course, as noted above, if investors anticipated LSAP shocks to be less persistent that conventional duration shocks, the responses would be even smaller than this.

Because this exercise involved some approximation and educated guessing, one might worry that the apparent weak impact of LSAPs simply reflects inaccuracies in the model inputs. To address this possibility, I consider adjusting the model in a number of ways that might make these effects larger. First, I double the size of the LSAP shock (assuming a reduction from 3.05 years to 1.65 years). Second, I weaken the effects of forward guidance by assuming that

[16] The decay in the effect of the LSAP shock over time is qualitatively consistent with the evidence presented in Ihrig et al. (2012) and Wright (2012).

investors anticipate the short rate to stay at zero for only one year, rather than two. Third, I allow for the possibility that non-Treasury asset-return uncertainty may have permanently increased in the wake of the crisis by doubling the standard deviation of R^v (in the models where that quantity is relevant). Finally, I allow for the possiblity that risk aversion was higher than usual during the financial crisis by doubling the value of the preference parameter (a or θ) in each model. Any one of these modifications by itself would generally increase the impact of the LSAP shock. To ensure that I am indeed capturing an upper bound on the plausible magnitude of the duraiton effects of LSAPs, I consider all four of them simultaneously. As shown in the final column of Table 5, these modifications collectively do increase the impact of the LSAP shock by a factor of between 3 and 5. However, even the largest is still only about half of the bottom end of the range of empirical estimates. These results suggest that the Federal Reserve's programs could not have generated effects on long-term yields of the magnitude that has been observed in empirical studies through duration effects alone.

6 Conclusion

This paper has presented a new method for studying the ways in which the term structure of interest rates and other asset prices are affected by fluctuations in the quantities of assets outstanding. The type of model considered is a rational-expectations version of portfolio-balance models that have been in use, with varying degrees of formality, for decades. It may be viewed as a generalization of the preferred-habitat model of Vayanos and Vila (2009), with investors that potentially have a broader class of objective functions and in which nonlinearities may be important. The key aspect of the model class that allows portfolio-balance effects to exist under no-arbtrage is that the pricing kernel depends on the return on wealth.

I examined portfolio-balance effects on the yield curve under a variety of assumptions about investor objectives. Some sets of assumptions can generate sizeable effects of Treasury maturity structure on term premia in some states of the world. However, these effects can be attenuated by various factors, including nonlinearities associated with the zero lower bound, diversificaiton of investor portfolios into assets that are uncorrelated with Treasury returns, and weak persistence of duration shocks. Many of these factors seem to be in play when considering the Federal Reserve's asset purchases. Even under rather extreme assumptions, I was unable to come close to even the lower bound of the empirical effects of these programs through the duration channel alone. I interpret these results, not as evidence that the asset purchases were ineffective, but as evidence that they probably had their effects primarily through mechanisms other than the removal of duration risk. It seems possible that phenomena not easily captured by a no-arbitrage model, such as market dislocations, liquidity shocks, and capital constraints, were important during and after the financial crisis and that LSAPs had additional effects through those channels. This reading is broadly consistent with the empirical conclusions of Krishnamurthy and Vissing-Jorgensen (2011), Cahill et al. (2013), and D'Amico and King (2013). More structural modeling of the behavior driving such scarcity effects is needed to determine whether they would significantly alter the results presented above.

Appendices

A. Solution Algorithm

Let $\mathbf{p}_d^i\ (\mathbf{s}_t)$ be a proposal for the 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 $i = 0, \ldots, I$ indexes iterations, and let $\mathbf{q}_d^i$ be the corresponding discritiztion of $\mathbf{q}(\mathbf{s}_t)$. 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

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

$\mathbf{p}_d^i$ and $\widehat{\tau}$ are used to generate a proposal for the joint distribution of period $t+1$ states and prices. That distribution, in turn, generates an updated pricing function $\mathbf{p}_d^{i+1}\ (\mathbf{s}_t)$ through the analogue to equation (11)—that is, by solving

$$p_{dn}^{i+1}\left(\mathbf{d}_j\right) = \sum_{g=1}^{G}\widehat{\tau}(\mathbf{d}_g|\mathbf{d}_j) M\left(\mathbf{d}_j, \mathbf{d}_g, \frac{\mathbf{x}\left(\mathbf{d}_j\right)' \mathbf{q}_d^i\left(\mathbf{d}_g\right)}{\mathbf{x}\left(\mathbf{d}_j\right)' \mathbf{p}_d^i\left(\mathbf{d}_j\right)}\right) p_{dn-1}^{i+1}(\mathbf{d}_g) \qquad \text{(A1)}$$

for the vector $\mathbf{p}_d^{i+1}\ (\mathbf{d}_j)$ at each node $j = 1, ..., G$.

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 $\mathbf{D}$, $\mathbf{p}_d^i\ (\mathbf{d}_j) \longrightarrow \mathbf{p}_d\ (\mathbf{d}_j)\ \forall\ \mathbf{d}_j \in \mathbf{D}$, where $\mathbf{p}_d$ is the (unique) pricing function that obtains if $\widehat{\tau}$ is the data-generating process. But continuity of τ ensures that, for any node j,

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

i.e., in the limit, the pricing function solves the no-arbitrage condition (1).

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

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

Once the algorithm has converged, the quantities on the right-hand side are all known. Thus, securities can be priced in at any point in $\mathbf{S}$.

Figure A1 displays some results on the convergence of the solution algorithm for the one-factor model discussed in Section 3. The top panel shows the computed 5-, 10-, 15-, and 30-year yields, shown for a short rate at its average value of 5.8%, across the first 50 iterations ($i = 1, \ldots 50$). The algorithm is initialized at a price vector $\mathbf{p}_d^0(\mathbf{d}_j) = (1, \ldots, 1)$ for all values of $\mathbf{d}_j$ and uses $G = 65$ nodes across the state space. It is evident from this figure that, for each maturity n, the solution converges very quickly once $i > n$.

The middle panel shows convergence in the number of gridpoints by displaying the computed yield curve (after $I = 50$ iterations), again using $r_t = 0.058$ for illustration. Yield curves are shown for $G = 5$, 9, 17, 33, and 65, in each case spaced equally across possible values of p_{1t}. The space is assumed bounded between 0 and 0.80, so this partitioning corresponds roughly to increments of between 38 basis points and 5 percentage points. While 5 nodes is clearly too few to achieve convergence, the solutions using 17 or more nodes are indistinguishable from each other.

For brevity, these results were shown for the average value of the 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 $p_{1t} = 0.80$ (corresponding to a risk-free return of 25%). 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 p_{1t} is extended over the entire range $(0, 1)$, again at the average value of the short rate. (The latter computation used $G = 100$.) The two curves are virtually identical, differing by less than 1 basis point across maturities.

As noted in the text, the model with Epstein-Zin preferences involves the additional variable of consumption growth, and, under a further assumption of trend-stationarity, this can be solved for within the context of the model itself, via an additional step in the algorithm. In particular, suppose consumption has a deterministic long-run trend with growth rate g, and denote the time-t level of consumption relative to this trend as $\widetilde{c}(\mathbf{s}_t)$. Then, we can rewrite the first term in equation (10) as

$$\left(\beta (c_{t+1}/c_t)^{-1/\psi}\right)^{\theta} = \frac{1}{\lambda^i} \frac{\gamma^i(\mathbf{s}_{t+1})}{\gamma^i(\mathbf{s}_t)}$$

where $\lambda^i \equiv \left(\beta g^{-1/\psi}\right)^{-\theta}$ and $\gamma^i(\mathbf{s}_t) \equiv \widetilde{c}(\mathbf{s}_t)^{-\theta/\psi}$.

At iteration i of the algorithm, define $\mathbf{A}^i$ as the $G \times G$ matrix whose h, j^{th} element is

$$a^i_{h,j} = \widehat{\tau}(\mathbf{d}_h|\mathbf{d}_j) \left(\frac{\mathbf{x}(\mathbf{d}_j)' \, \mathbf{q}^i_d(\mathbf{d}_h)}{\mathbf{x}(\mathbf{d}_j)' \, \mathbf{p}^i_d(\mathbf{d}_j)} - \pi(\mathbf{d}_h) \right)^{1-\theta}$$

From equation (10), λ^i is an eigenvalue of $\mathbf{A}^i$, with corresponding eigenvector $\boldsymbol{\gamma}^i = \begin{pmatrix} \gamma^i(\mathbf{d}_1) & ... & \gamma^i(\mathbf{d}_G) \end{pmatrix}$. But, since $\mathbf{A}^i$ is necessarily a non-negative square matrix, the Perron-Frobenius Theorem implies that the eigenvector corresponding to the (unique) maximum eigenvalue is the only eigenvector that has all real and positive elements. Consequently, it is the only viable solution. Thus, for given pricing and payoff functions $\mathbf{p}_d^i(\mathbf{d}_j)$ and $\mathbf{q}_d^i(\mathbf{d}_h)$, it is straightforward to find λ^i and $\boldsymbol{\gamma}^i$. With these in hand, equation (A1) for the Epstein-Zin model is

$$p_{dn}^{i+1}(\mathbf{d}_j) = \sum_{g=1}^{G} \widehat{\tau}(\mathbf{d}_g|\mathbf{d}_j) \frac{1}{\lambda^i} \frac{\gamma^i(\mathbf{d}_{t+1})}{\gamma^i(\mathbf{d}_t)} \left(\frac{\mathbf{x}(\mathbf{d}_j)' \mathbf{q}_d^i(\mathbf{d}_h)}{\mathbf{x}(\mathbf{d}_j)' \mathbf{p}_d^i(\mathbf{d}_j)} - \pi(\mathbf{d}_h) \right)^{1-\theta} p_{dn-1}^{i+1}(\mathbf{d}_g)$$

Note that we need not solve explicitly for consumption, nor the parameters β and ψ, in order to compute asset prices or find the distribution of M within this model.

B. Preferred Habitat

In the models considered in the text, the supply of debt that must be held by investors is exogenous. More generally, this distribution itself could depend on bond prices, as in the model of Vayanos and Vila (2009). This appendix briefly shows how the class of models discussed in the text can incorporate this type of endogeneity.

Apart from its affine structure, the essential feature of Vayanos-Vila that differs from the models considered in the text is that investors ("arbitraguers" in their terminology) face an elastic supply curve at each maturity. This supply curve is assumed to arise from the presence of preferred-habitat agents, each of whom deals in debt of only one maturity. Specifically, for each bond, preferred-habitat agents demand a (market-value) quantity that is a function h of maturity and price:

$$p_{nt}\xi_{nt} = h(n, p_{nt}) \tag{B1}$$

where ξ_{nt} is the par value demanded. The par value left in the hands of the arbitrageurs is simply $x_{nt} - \xi_{nt}$. The models in the text effectively assumed $h(n, p_{nt}) = 0$ for all n. Equation (B1) can be substituted directly into (4) to give the return on the arbitrageurs' portfolio:

$$R_t = \frac{\sum\limits_{n=1}^{N} \overline{x}_{nt}(p_{nt}) p_{n-1t+1}}{\sum\limits_{n=1}^{N} \overline{x}_{nt}(p_{nt}) p_{nt}}$$

where

$$\overline{x}_{nt}(p_{nt}) \equiv x_{nt} - \frac{h(n, p_{nt})}{p_{nt}} \tag{B2}$$

For any given assumption of the form of $h(.)$, and an assumption on the pricing kernel, the algorithm can solve for asset prices under this modified definition of the return on wealth.

In the Vayanos-Vila model, bonds are in zero net supply—a positive position by preferred-habitat agents implies a negative position for arbitrageurs, and vice versa. In the models discussed here, Treasury bonds are assumed to be issued by the government and, from the point of view of investors, constitute net wealth. That is, they are in positive net supply at all maturities. (One rationale for this assumption is that only the government, not private investors, can create perfectly default-free debt.) We can further impose that both arbitrageurs and preferred-habitat agents can only hold positive quantities of bonds by modifying (B2) as follows:

$$\overline{x}_{nt}(p_{nt}) \equiv \min\left[x_{nt}, \max\left[0, x_{nt} - \frac{h\left(n, p_{nt}\right)}{p_{nt}}\right]\right]$$

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.

Barro, R. J., 1974. "Are Government Bonds Net Wealth?" Journal of Political Economy 82(6): 1098-117.

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.

Bernanke, B., 2010. The Economic Outlook and Monetary Policy. Remarks delivered at Jackson Hole, Wyoming, August.

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).

Chabot, B. and Herman, G., 2013. "A History of Large-Scale Asset Purchases before the Fed." FRB Chicago *Economic Perspectives* 37, Q4.

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.

Doh, T., 2010. "The Efficacy of Large-Scale Asset Purchases at the Zero Lower Bound." FRB Kansas City *Economic Review* (Q2): 5-34.

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.

Femia, K.; Friedman, S.; and Sack, B., 2013. "The Effects of Policy Guidance on Perceptions of the Fed's Reaction Function." FRB New York Staff Report 652 (November).

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.*

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.

Gurkaynak, R. S. and Wright, J. H., 2012. "Macroeconomics and the Term Structure." *Journal of Economic Literature* 50(2): 331-67.

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., 2012. "Mortgage Refinancing, the Quantity of Risk, and Bond Risk Premia." Harvard Business School, Working paper (September).

Ihrig, J.; Klee, E.; Li, C.; Schulte, B.; and Wei, M., 2012. "Expectations about the Federal Reserve's Balance Sheet and the Term Structure of Interest Rates." FEDS working paper 2012-57.

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.

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.

Markowitz, H. M., 1952. “Portfolio Selection.” *Journal of Finance* 7(1): 77-91.

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

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

Rogers, J. H.; Scotti, C.; and Wright, J. H., 2014. "Evaluating Asset-Market Effects of Unconventional Monetary Policy: A Cross-Country Comparison." Federal Reserve Board, International Finance Discussion Paper 1101 (March).

Rosa, C., 2012. “How ‘Unconventional’ Are Large-Scale Asset Purchases? The Impact of Monetary Policy on Asset Prices.” FRB NY Staff Report 560, May.

Rudebusch, G. D.; Sack, B. P.; and Swanson, E. T., 2007. “Macroeconomic Implications of Changes in the Term Premium.” FRB St. Louis *Review* 89(4): 241-69.

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

Vasicek, O., 1977. “An Equilibrium Characterisation of the Term Structure.” *Journal of Financial Economics* 5(2): 177-88.

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

Wallace, N., 1981. “A Modigliani-Miller Theorem for Open-Market Operations.” *American Economic Review* 71(3): 267-74.

Woodford, M., 2012. "Methods of Policy Accommodation at the Interest-Rate Lower Bound." Paper presented at the Federal Reserve Bank of Kansas City Jackson Hole Symposium, 16 September.

Wright, J. H., 2012. "What Does Monetary Policy Do to Long-Term Interest Rates at the Zero Lower Bound?" *Economic Journal* 122: F447-66.

Table 1. Investor Assumptions for Portfolio-Balance Models

Model	Preferences	Treas. weight (v)	Preference calibration
1	Quadratic - Nominal returns	1	$a = 9.5$
2	Quadratic - Real returns	1	$a = 7.3$
3	EZW	1	$\theta = -7.3$
4	Quadratic - Real returns	0.1	$a = 74.9$
5	EZW	0.1	$\theta = -7.6$

Notes: The table shows the salient features of the five three-factor portfolio-balance models described in the text. Models are characterized by (1) the form of investors preferences (quadratic over real or nominal returns or Epstein-Zin-Weil), (2) the weight v that Treasury bonds receive in their overall portfolio, and (3) a single preference parameter, which is calibrated to match the average 1- to 15-year yield curve slope in the data.

Table 2. Calibration of Economic Dynamics

	π_t	r_t	z_t
coef			
π_{t-1}	0.83	0.26	0
r_{t-1}	0.04	0.72	0
z_{t-1}	0	0	0.96
intercept	0.004	0.006	0.15
stdev err	0.013	0.017	0.14

Notes: The table shows the parameters of the VAR dynamics for the three state variables–inflation, the short-term interest rate, and the average Treasury duration ()—used in the five portfolio-balance models described in the text. Parameter values are estimated based on data over the 1971-2008 period. The dynamics further assume that the short rate is truncated at zero.

Table 3. Yield Curve Statistics

Means

Maturity	Data	Model 1	Model 2	Model 3	Model 4	Model 5
5y	7.1	7.0	6.9	8.5	7.0	8.5
10y	7.5	7.4	7.4	8.0	7.4	8.1
15y	7.7	7.7	7.7	7.7	7.7	7.7

Standard deviations

Maturity	Data	Model 1	Model 2	Model 3	Model 4	Model 5
5y	2.6	2.2	2.3	3.3	2.3	3.4
10y	2.4	1.7	1.7	2.3	1.7	2.5
15y	1.5	1.3	1.3	1.6	1.3	1.8

Correlation with Treasury duration

Maturity	Data	Model 1	Model 2	Model 3	Model 4	Model 5
5y	0.09	0.09	0.08	0.08	0.08	0.05
10y	0.07	0.13	0.12	0.10	0.12	0.05
15y	0.07	0.17	0.15	0.12	0.15	0.05

Notes: The table shows certain summary statistics for nominal yields at various maturities generated by the five three-factor portfolio-balance models described in the text and compares them to the values observed in the data between 1971 and 2008. The model results are calculated by feeding in the actual time series for the three state variables (core PCE inflation, the short-term interest rate, and average Treasury duration outstanding) over this period.

Table 4. Excess Return Statistics

Means

Maturity	Data	Model 1	Model 2	Model 3	Model 4	Model 5
5y	2.2	1.9	1.8	3.1	1.9	3.2
10y	3.8	3.0	3.0	2.6	3.1	2.8
15y	4.9	3.6	3.7	2.4	3.8	2.7

Standard deviations

Maturity	Data	Model 1	Model 2	Model 3	Model 4	Model 5
5y	6.2	5.6	5.5	8.4	5.5	8.5
10y	12.1	9.3	9.3	13.2	9.3	13.6
15y	17.7	11.4	11.3	14.9	11.4	15.4

Correlation with Treasury duration

Maturity	Data	Model 1	Model 2	Model 3	Model 4	Model 5
5y	0.29	0.25	0.24	0.09	0.24	0.05
10y	0.31	0.28	0.27	0.18	0.27	0.15
15y	0.30	0.29	0.28	0.19	0.28	0.16

Notes: The table shows certain summary statistics for returns on bonds of various maturities, in excess of the one-year return, generated by the five three-factor portfolio-balance models described in the text and compares them to the values observed in the data between 1971 and 2008. The model results are calculated by feeding in the actual time series for the three state variables (core PCE inflation, the short-term interest rate, and average Treasury duration outstanding) over this period.

Table 5. Term Premium Experiments

	1970s-80s Duration rise	Asset purchases Baseline	Extreme values
Experiment	$z_0 = 1.5$	$z_0 = 2.7$	$z_0 = 3.05$
	$z_1 = 3.5$	$z_1 = 2.0$	$z_1 = 1.65$
Assumptions:			
State values	$\pi = 0.058$	$\pi = 0.014$	$\pi = 0.014$
	$r = 0.091$	$r = 0.0025$	$r = 0.0025$
"Forward guidance"	None	2 years	1 year
Preference param.	baseline	baseline	double baseline
Other asset vol.	baseline	baseline	double baseilne
Effect on term premium (pp)			
Model 1	0.56	-0.13	-0.37
Model 2	0.43	-0.10	-0.48
Model 3	0.52	-0.11	-0.33
Model 4	0.44	-0.10	-0.49
Model 5	0.06	-0.01	-0.05

Notes: The table shows the outcomes of experiments in which the duration of outstanding Treasury debt is shifted within each of the five three-factor portfolio-balance described in the text. The top rows show the amount by which duration is shifted relative to its starting value (z_0), the middle rows show the assumed initial values of the other state variables and certain other modeling assumptions, and the bottom rows show the response of the ten-year term premium in the period when the shift occurs. In all cases the change in duration is assumed to be unanticipated and to subsequently follow the dynamic process described in the text. The first experiment is meant to replicate the increase in outstanding duration that occurred roughly between 1976 and 1988. The second experiment is meant to replicate the duration extracted from the market by the Federal Reserve's asset purchases during the period 2008 - 2012. For the latter, I consider both a baseline case and a case that uses more-extreme assumptions to maximize term-premium effects.

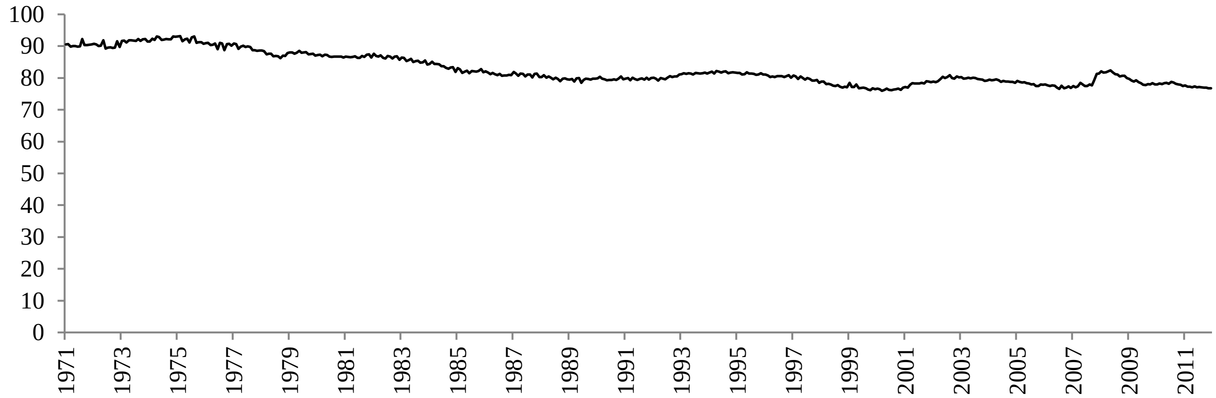
Table 6. Summary of Federal Reserve Asset Purchase Programs

	Dates	Net quantity of Treasuries purchased ($bil)	Net quantity of MBS purchased ($bil)	Net quantity of reserves created ($bil)	Assumed change in average Treasury duration outstanding	Empirical effect on ten-year term premium
LSAP I	5/08-5/10	$300	$941	$1,318	0	40-100 bp
MBS reinvestment	8/10-10/11	$285	$0	$285	0	10-25 bp
LSAP II	11/10-7/11	$600	$0	$600	0	15-55 bp
MEP	10/11-12/12	$0	$0	$0	-0.5	30-65 bp
Total		$1,185	$941	$2,203	-0.5	95-245 bp

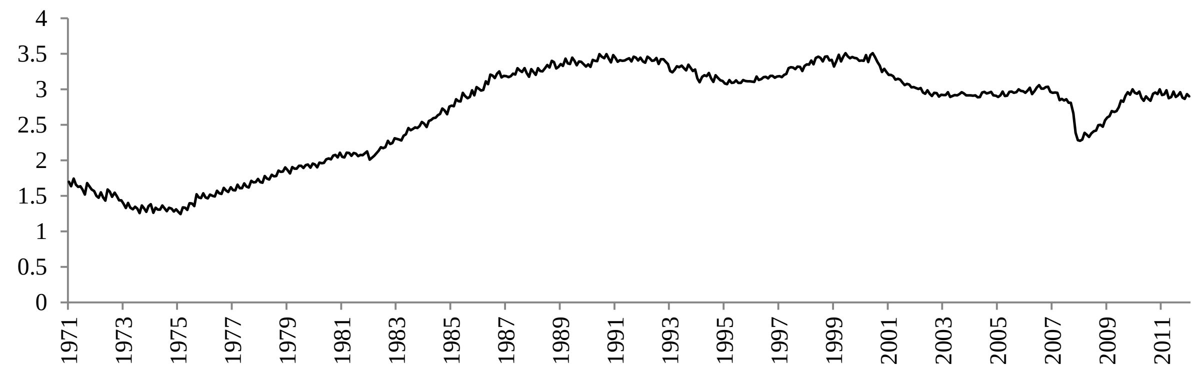
Notes: All quantities are net of redemptions and principal payments through December 2012. The assumed duration change is only that in Treasury securities (i.e., excluding agency debt, MBS, and the monetary base). The empirical effect of each program on the ten-year nominal Treasury yield is taken from the literature mentioned in Section 5 of the text.

Figure 1. Characteristics of U.S. government liabilities in public hands

A. Percentage with duration < 5 years

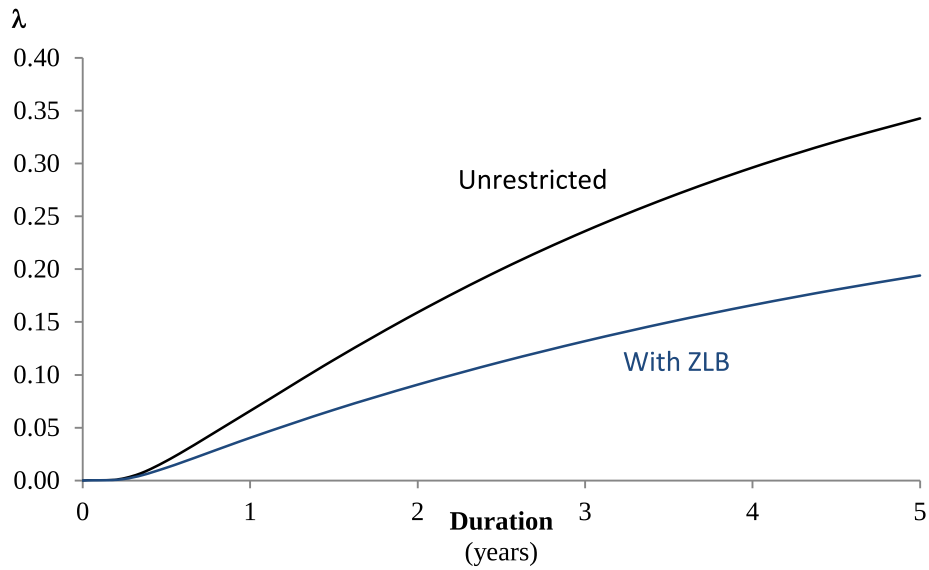


B. Average duration



Notes: Includes coupon notes, bonds, and bills issued by the U.S. Treasury, less the amount held in the Federal Reserve's SOMA portfolio, plus reserves and currency in circulation, which are assumed to have a duration of zero. Sources: CRSP, Federal Reserve.

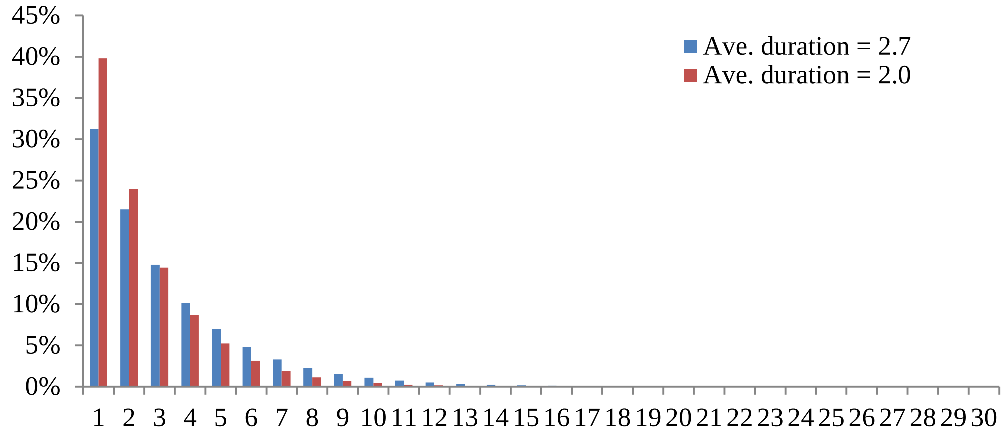
Figure 2. Risk prices in the one-factor model



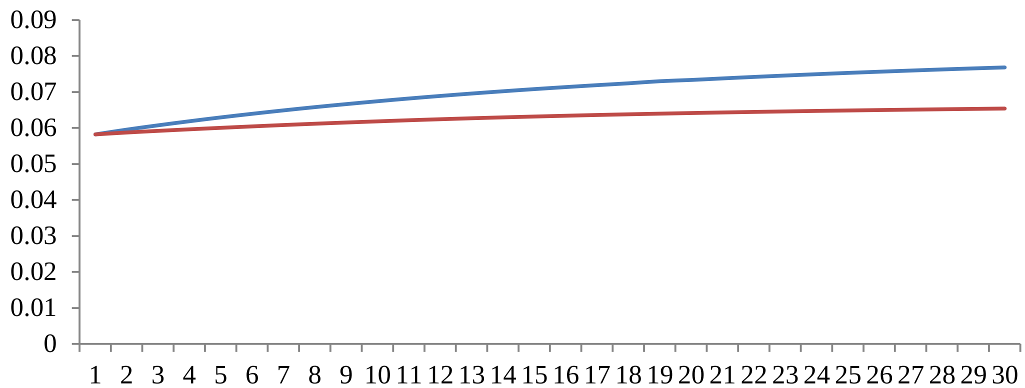
Notes: The figure shows the price of short-rate risk (expected excess returns divided by their standard deviation) in the one-factor model for various values of aggregate duration (z). The black line corresponds to a model in which the short rate follows a simple AR(1) process. The blue line corresponds to a model in which that process is truncated at zero.

Figure 3. A duration shift in the one-factor model

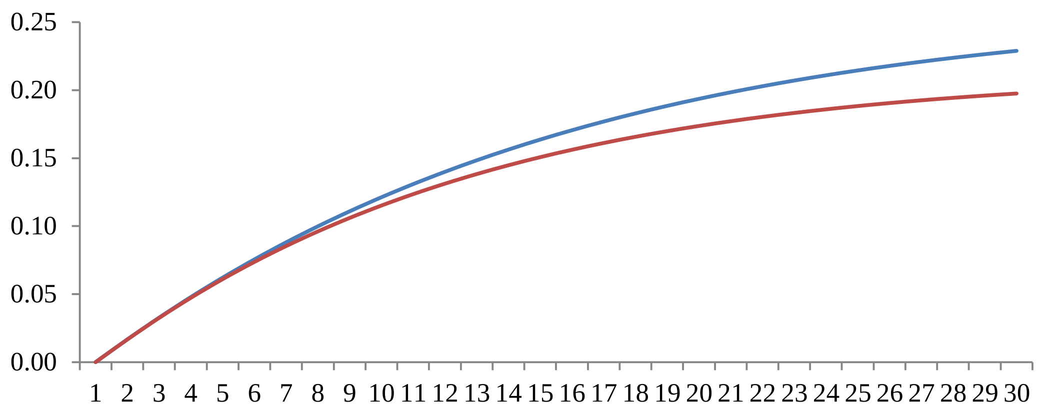
Duration distribution of outstanding debt



Yield curve



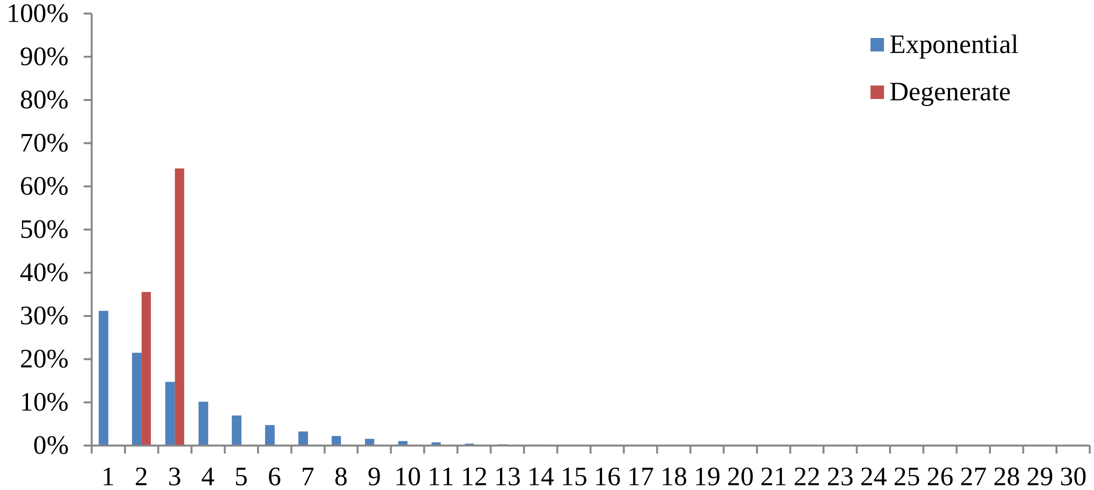
Return Volatility (One-Year)



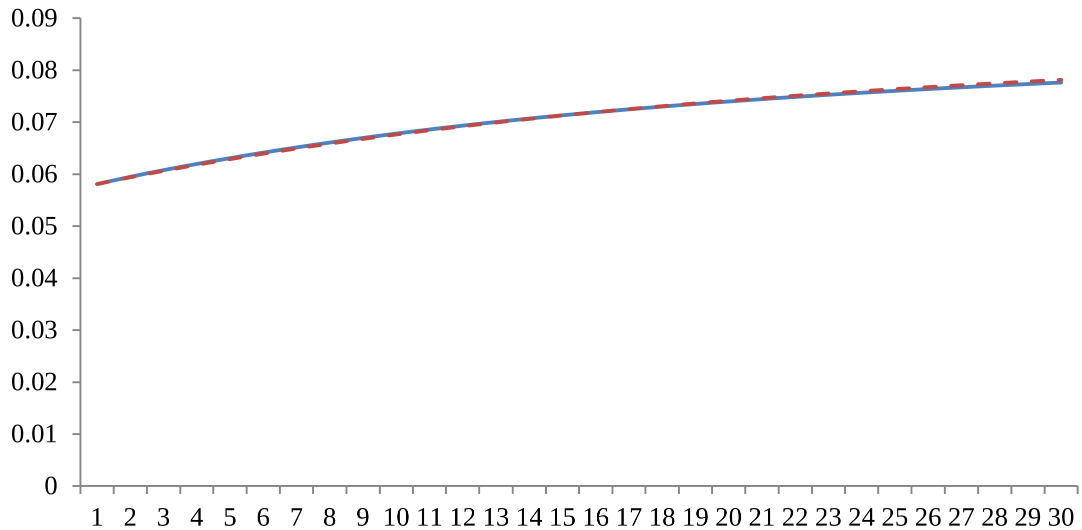
Notes: The figure shows the effect of moving from an average duration of 2.7 to 2.0 years in the one-factor model, evaluated at a short rate of 5.8% (the sample average).

Figure 4. Effect of the maturity-distribution shape in the one-factor model

Duration distribution of outstanding debt

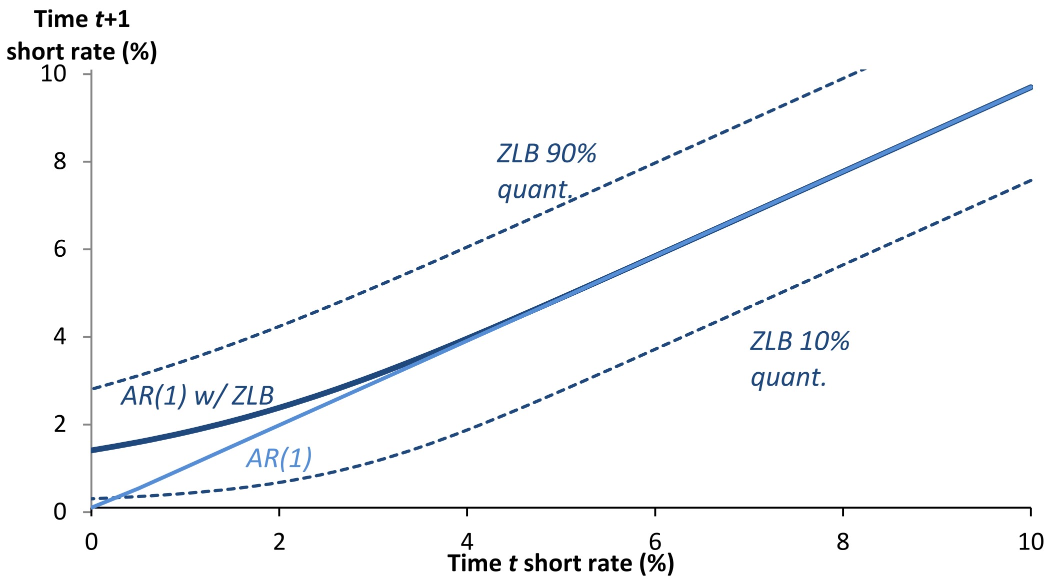


Yield curve



Notes: The top panel shows two possible distributions for the maturity of outstanding Treasury debt, one exponential and one nearly degenerate, both with means of 2.7 years. The bottom panel shows the corresponding yield curves generated by the one-factor model when the short rate is at its average value of 5.8%. In both cases, the risk-aversion coefficient is calibrated to match the average value of the 15-year yield.

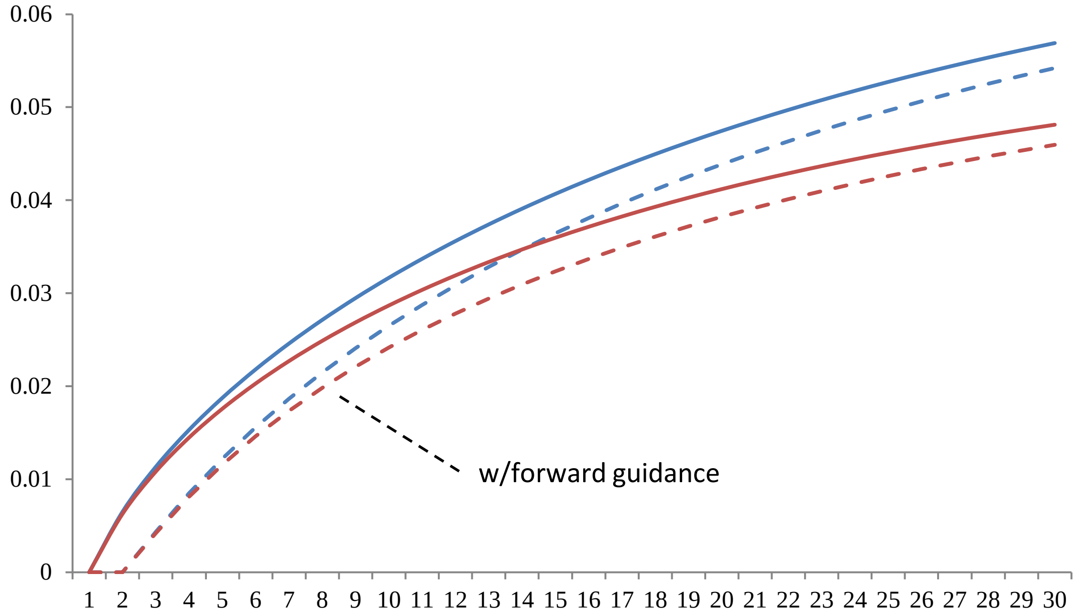
Figure 5. Short-rate processes



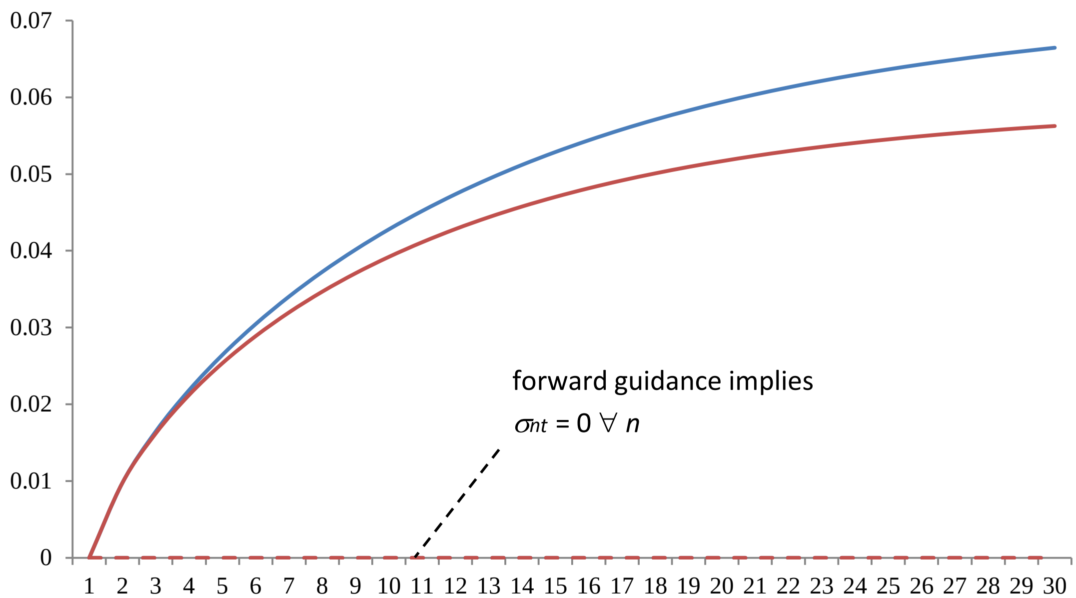
Notes: The figure shows the conditional mean of r_{t+1}, as a function of r_t near the value of zero under both a standard AR(1) process (light-blue solid line) and the modified process enforcing the zero lower bound, given by equation (18) (dark-blue solid line). 10% and 90% quantiles of the modified process are shown by the dashed lines. Frequency is annual.

Figure 6. A duration shift in the one-factor model at the zero lower bound

Yield curve

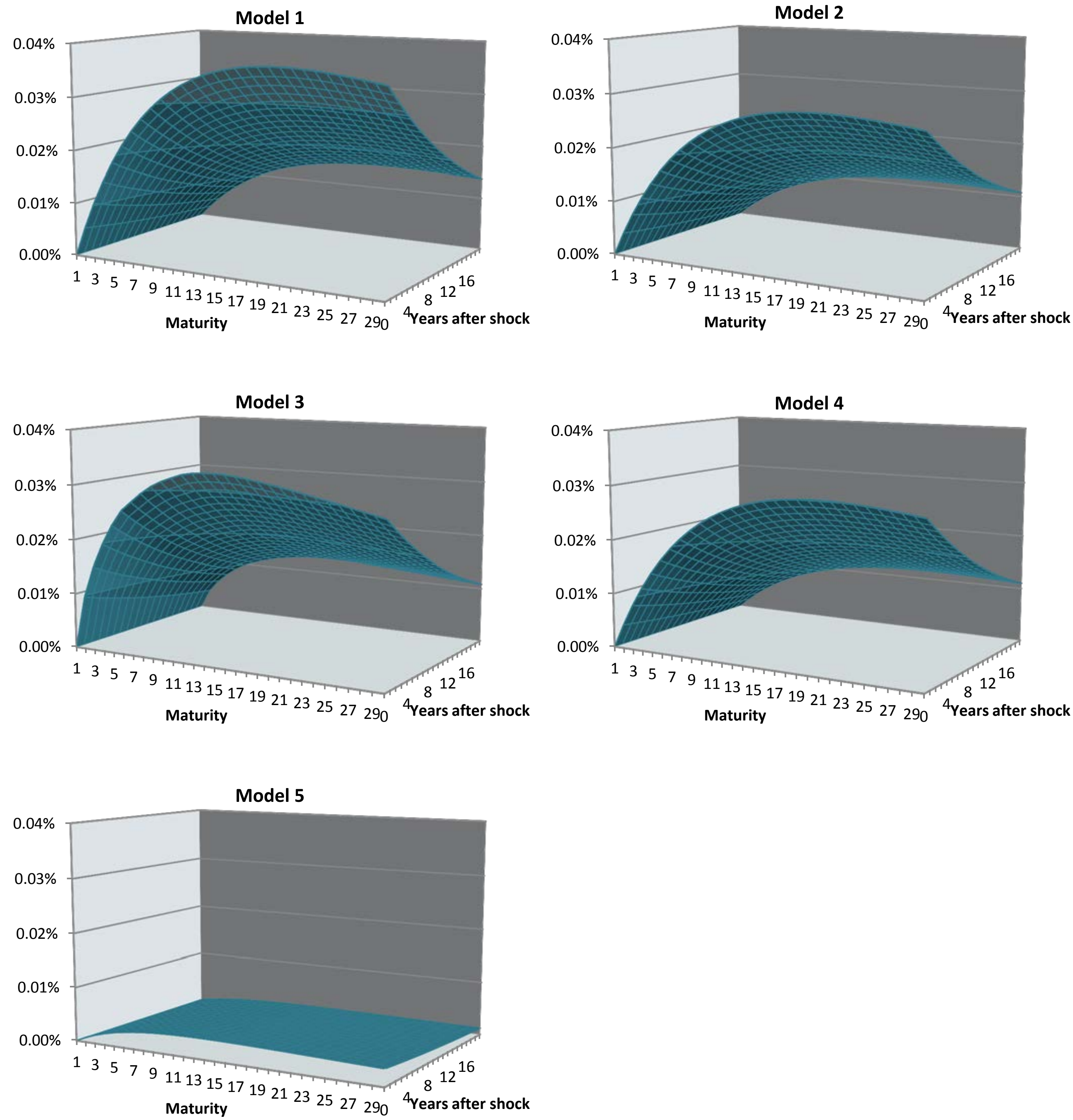


Return Volatility (One-Year)



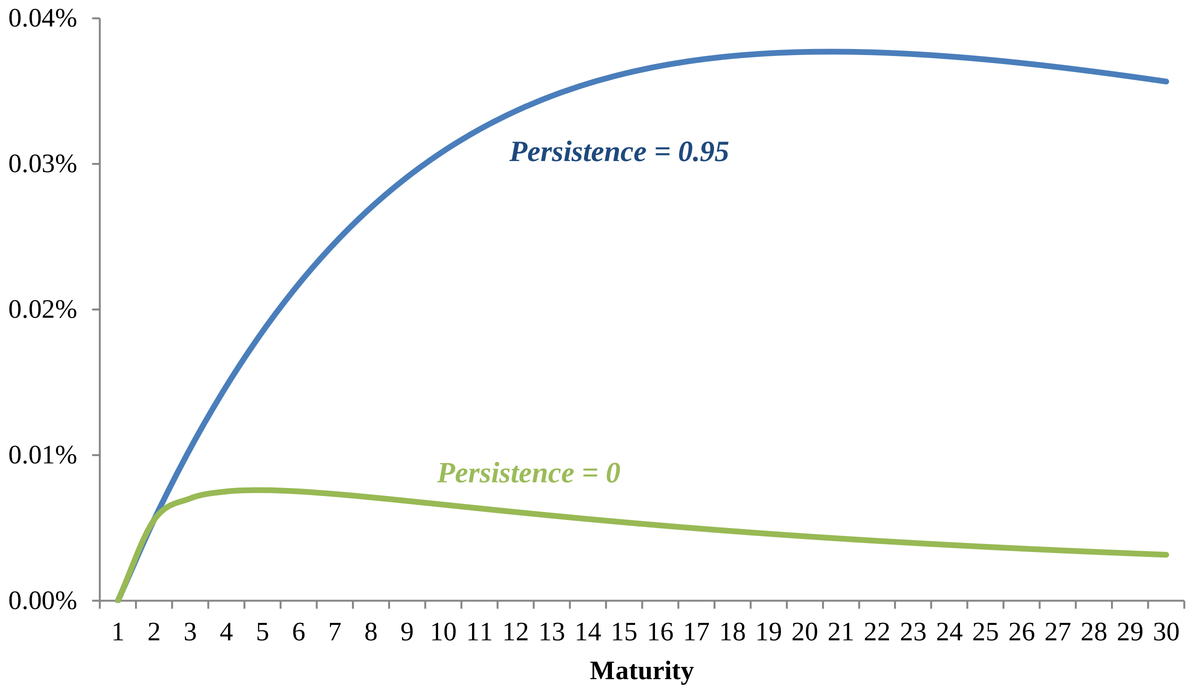
Notes: The figure shows the effect of moving from an average duration of 2.7 years (blue) to 2.0 years (red) in the one-factor model, evaluated at a short rate of 0%. The solid lines represent the case in which the short rate is expected to evolve according to its usual truncated-AR(1) process. The dashed lines ("forward guidance") represent the case in which the short rate is anticipated to remain at 0% with certainty for one year and only then to follow its usual truncated-AR(1) process.

Figure 7. Impulse-response functions for the three-factor models



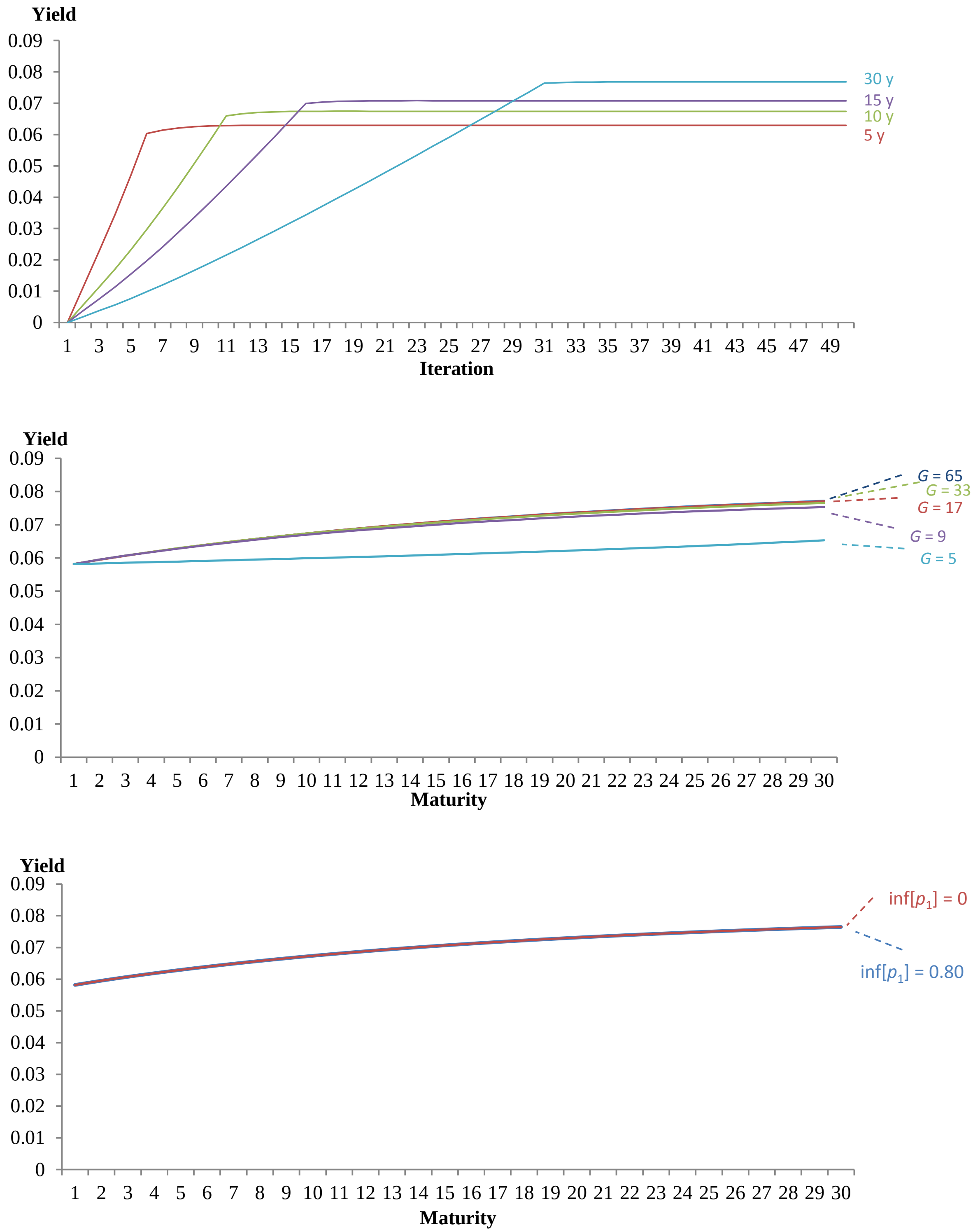
Notes: The figures show the response of the yield curve, evaluated initially at the unconditional means of the three state variables, to a one-standard-deviation positive shock to average Treasury duration (z_t) in each of the five portfolio-balance models described in the text.

Figure 8. Effect of duration persistence on the response of yields



Notes: The figure shows the response of the yield curve in Model 1 to one-standard-deviation positive shock to average outstanding Treasury duration (z_t), in the period when the shock occurs, under two assumptions for the persistence of the z_t process. The blue line shows the response when the autoregressive coefficient for z_t is equal to 0.95 (same as in Figure 7), and the green line shows the response when this autoregressive coefficeint is zero. The functions are evaluated at the unconditional means of the state variables.

Figure A1. Solution convergence in the one-factor model



Notes: For the one-factor portfolio-balance model, the graphs show how the solution algorithm converges in the number of iterations (top), number of nodes in the grid (middle), and truncation point of the state space (bottom). All calculations are illustrated at the average value of the short rate.