Bond Supply and Excess Bond Returns

Robin Greenwood
Harvard Business School and NBER

Dimitri Vayanos
LSE, CEPR, and NBER

We examine empirically how the supply and maturity structure of government debt affect bond yields and expected returns. We organize our investigation around a term-structure model in which risk-averse arbitrageurs absorb shocks to the demand and supply for bonds of different maturities. These shocks affect the term structure because they alter the price of duration risk. Consistent with the model, we find that the maturity-weighted-debt-to-GDP ratio is positively related to bond yields and future returns, controlling for the short rate. Moreover, these effects are stronger for longer-maturity bonds and following periods when arbitrageurs have lost money. (*JEL* E43, E52, G11, G12, H63)

How do the supply and maturity structure of government debt affect interest rates? If, for example, the government raises the supply of long-term bonds, would this raise the spread between long and short rates? According to standard representative-agent models, there should be no effect because of Ricardian equivalence (Barro 1974). Intuitively, the consumption of the representative agent, and hence interest rates, depends on government spending but not on how spending is financed.

The irrelevance result is at odds with a view held by many policy makers and emphasized in early term-structure theories. According to the portfolio-balance theory (e.g., Tobin 1958, 1969), investors would be willing to absorb an increased supply of long-term bonds, and hence bear more risk, only if they were compensated by an increase in long rates relative to short rates. According to the preferred-habitat theory (e.g., Culbertson 1957; Modigliani and Sutch 1966), an increased supply of long-term bonds would be absorbed mainly by

a clientele of long-horizon investors. Long rates would increase, while short rates, determined mainly by short-horizon investors, might not be affected.

Determining empirically how the supply and maturity structure of government debt affect interest rates is important for informing the theory of the term structure, especially given the conflicting predictions. An empirical investigation of supply effects is also relevant from a policy viewpoint. For example, during the recent financial crisis, central banks around the world conducted unprecedented open-market purchases of intermediate- and long-term government bonds. Drawing on the portfolio-balance and preferred-habitat theories, the central banks hoped that their purchases, in a strategy also known as quantitative easing, would lower long-term interest rates and stimulate private investment.

In this paper we use time-series data to examine how the supply and maturity structure of government debt affect government bond yields and expected returns in the United States. We organize our investigation around a term-structure model in which risk-averse arbitrageurs absorb shocks to the demand and supply for bonds of different maturities. The model predicts that an increase in supply should raise bond yields and expected returns, holding the short rate constant. Moreover, these effects should be stronger for longer-maturity bonds and during times when arbitrageurs are more risk averse. The data support these predictions. Using our empirical estimates of supply effects, we calibrate the model and infer arbitrageur risk aversion.

Our theory builds on the preferred-habitat model of Vayanos and Vila (2009). We simplify that model by assuming that the demand and supply for each maturity in the absence of arbitrageurs are price-inelastic. The resulting model captures the portfolio-balance effect but abstracts away from clienteles and preferred habitats since the only agents absorbing shocks are identical arbitrageurs.

Changes in supply in our model affect bond yields and expected returns because they change the amount of interest-rate risk, or "duration risk," borne by arbitrageurs. For example, to accommodate an increase in the supply of long-term bonds, arbitrageurs must absorb more duration risk, and hence require all bonds in their portfolio to offer higher expected returns in excess of the short rate. As a consequence, prices go down for all bonds, and yields and expected returns go up. This holds even when the increase in the supply of long-term bonds is accompanied by an equal decrease in the supply of short-term bonds. Indeed, since long-term bonds are more sensitive to duration risk than short-term bonds, arbitrageurs must absorb more such risk. Therefore, prices go down for all bonds, including for short-term ones whose supply decreases.

We assume that supply is described by one stochastic factor, and allow the loadings on that factor to differ across maturities in both magnitude and sign. For example, increases in the supply factor could correspond to increases in the supply of long-term bonds and decreases in the supply of short-term

bonds. We also assume, as a normalization and without loss of generality, that increases in the supply factor correspond to increases in duration-weighted supply. Therefore, when the supply factor increases, so do the yields and expected returns of all bonds, holding the short rate constant.

Increases in the supply factor have stronger effects when arbitrageurs are more risk averse. Moreover, the effects on expected returns are stronger for long-term bonds than for short-term bonds. This is because long-term bonds are more sensitive to duration risk, and hence to changes in the price of that risk. Finally, the effects of supply on yields are increasing or hump-shaped across maturities, and are smaller than on expected returns. Both results follow from the property that the effect of a supply shock on a bond's yield is equal to the average effect on the bond's instantaneous expected return over the bond's life. This average effect can be stronger for an intermediate-term bond than for a long-term bond if the shock mean-reverts quickly. It is also smaller than the effect on the bond's current expected return for two reasons. Since the shock mean-reverts, its effect on the expected return of all bonds dies down over time. And even in the absence of mean reversion, the shock's effect on the expected return of any given bond decreases over time. This is because the bond's time to maturity decreases, and so does the bond's sensitivity to changes in the price of duration risk.

We test the predictions of our model using data on the U.S. Treasury market from 1952 to 2007. For every bond, CRSP maintains a record of bond characteristics (e.g., coupon rate and maturity) as well as monthly observations of face value outstanding. Using these data, we compute the maturity structure of aggregate payments on government debt. We also compute a dollar duration of these payments by multiplying each payment by the corresponding maturity and summing across maturities. We use this dollar duration as our main measure of supply, as suggested by our model, scale it by gross domestic product (GDP), and term it the maturity-weighted-debt-to-GDP ratio. Maturity-weighted debt to GDP is approximately the product of debt to GDP times the average maturity of debt.

We regress yields and future returns on our supply measure, controlling for the one-year yield, which we use as a proxy for the short rate. Consistent with our model, we find that supply is positively related to yields and future returns. The effects are statistically and economically significant. For example, a one-standard-deviation increase in our main measure of supply raises the yield on a long-term bond with approximate maturity twenty years by 40 basis points and its expected return over a one-year horizon by 259 basis points. We find evidence in support of the other predictions of our model as well. The effects of supply on yields and expected returns are increasing with maturity, and the effects on yields are smaller than on expected returns. Moreover, using a measure of arbitrageur wealth implied by our model, we find that both supply and the slope of the term structure become stronger predictors of future returns when arbitrageur wealth is low.

We subject our empirical results to a number of robustness tests, two of which deserve particular mention. First, we extend the time series by collecting additional data on the supply and maturity structure of government debt in a prewar 1916–1940 sample. The results in that sample are broadly similar to those in our main sample. Second, we address the concern that supply might be endogenous. For example, the government might choose maturity structure to cater to fluctuations in investor demand, mitigating and potentially even reversing any positive relationship that would otherwise obtain between supply and yields or expected returns. We instrument maturity-weighted debt to GDP by marketable Treasury debt to GDP. This is a suitable instrument because it is correlated with maturity-weighted debt to GDP, while also being driven mostly by the cumulation of past deficits rather than by changes in investor demand. In the instrumental-variables regressions, the effect of supply on expected returns remains statistically significant, and the coefficients are almost identical to their ordinary least squares (OLS) counterparts.

Last, we calibrate our model to the data. We estimate parameters for the processes governing the short rate and the supply factor. We also estimate how supply at each maturity loads on the supply factor. Combining these with our estimates of supply effects on yields and expected returns, we infer a coefficient of relative risk aversion (CRRA) for the arbitrageurs. We find that this coefficient is 57 times the ratio of arbitrageur wealth to GDP. This yields a range from 7.6, in the case where shocks to the supply of government debt are absorbed only by hedge funds, to 91.2, in the case where private pension funds, insurance companies, and mutual funds are equally active in absorbing the supply shocks.

A number of papers measure supply effects by analyzing the behavior of bond yields around specific policy events. Such events include Operation Twist, a program undertaken by the U.S. Treasury and the Federal Reserve (the Fed) during 1962–1964 with the objective to shorten the average maturity of government debt (e.g., Modigliani and Sutch 1966; Ross 1966; Wallace 1967; Swanson 2011), the 2000–2002 buybacks by the U.S. Treasury, undertaken with a similar objective (e.g., Garbade and Rutherford 2007; Greenwood and Vayanos 2010), and the recent quantitative easing (QE) programs in the United States (e.g., Gagnon et al. 2011; Krishnamurthy and Vissing-Jorgensen 2011; D'Amico, English, et al. 2012; D'Amico and King 2013) and the United Kingdom (e.g., Joyce, Lasaosa, et al. 2011).[1] An advantage of such event studies is that because the exact dates of policy events are known, it is easier to map changes in supply to changes in yields. At the same time, these events can sometimes be confounded by news about future monetary policy or the

---

[1] See also Bernanke, Reinhart, and Sack (2004) for a broader analysis of QE programs, and Joyce, Myles, et al. (2012) for a survey of the theoretical and empirical literature on QE.

broader economy, or can occur during times when arbitrageur capital is limited (Krishnamurthy and Vissing-Jorgensen 2011).

Simon (1991, 1994), Duffee (1996), and Fleming (2002) document supply effects in the cross-section of Treasury bills by correlating the supply of individual bills with the idiosyncratic component of their yields. Fleming and Rosenberg (2007) find that Treasury dealers are compensated by high excess returns when holding large inventories of newly issued Treasury securities. Lou, Yan, and Zhang (2013) document that prices of Treasury securities drop before issuance dates and then rebound predictably. We focus instead on effects at a more aggregate scale and a lower frequency.

Reinhart and Sack (2000) and Dai and Philippon (2006) find that government deficits raise the spread between long- and short-term interest rates. The latter paper also shows that the effect occurs partly through an increase in the risk premia of long-term bonds. Kuttner (2006) finds that shifts in Federal Reserve holdings of government debt toward long maturities lower the risk premia of two-, three-, four-, and five-year bonds. We examine instead how a theoretically motivated measure of the supply of Treasury debt, which includes both the level of debt and its average maturity, affects bond yields and expected returns. Beyond these findings, we also test for predictions of our model on how supply effects should manifest themselves in the cross-section and the time series.[2]

Hamilton and Wu (2012) structurally estimate a discretized version of Vayanos and Vila (VV 2009) and derive measures of supply that they then use to predict returns in the 1990–2007 sample. Li and Wei (2012) estimate an affine term-structure model with macroeconomic factors and two explicit supply factors, imposing some of the structure suggested by VV. The estimates of supply effects from these papers are broadly consistent with ours. Other papers that employ a similar theoretical framework include Hanson (2012) and Malkhozov et al. (2013), who examine how changes in the duration of mortgage-backed securities arising from prepayment options affect yields and expected returns, and Hong, Sraer, and Yu (2013), who examine how the effects of supply in the Treasury market interact with those of disagreement about future inflation.

Longstaff (2004) finds that U.S. Treasury bonds trade at a high price premium relative to bonds issued by Refcorp, a U.S. government agency, during those months of the 2000–2002 buybacks when the Treasury made large purchases. Krishnamurthy and Vissing-Jorgensen (2012) find that when government bonds are in small supply—that is, debt to GDP is low—they trade at a high price premium relative to AAA-rated corporate bonds. The findings of these papers suggest that clienteles and preferred habitats can exist not only within the Treasury market but also between Treasuries and other markets.

---

[2] Some papers document price effects of demand rather than of supply. See, for example, Park and Reinganum (1986), Ogden (1987), Fernald, Mosser, and Keane (1994), Kambhu and Mosser (2001), Sierra (2010), and Baker and Wurgler (2012).

The rest of this paper is organized as follows. Section 1 develops the theoretical framework and derives the empirical hypotheses. Section 2 describes the data and our measures of supply. Section 3 presents the empirical results. Section 4 calibrates our model to the data, and Section 5 concludes. The proofs of the theoretical results, as well as some supplementary tables and other data material, are in Appendices.

1. Theoretical Predictions

A theoretical framework helps organize our empirical investigation of supply effects on the term structure. The theory builds on the preferred-habitat model of Vayanos and Vila (VV 2009), in which arbitrageurs absorb shocks to the demand and supply for bonds of different maturities. Arbitrageurs integrate the markets for different maturities, rendering the term structure arbitrage-free. Because, however, they are risk averse, demand and supply shocks affect bond prices. We focus on the two-factor version of VV, with a short rate and a supply factor, and simplify the model by assuming that the demand and supply for each maturity in the absence of arbitrageurs are price-inelastic. This allows us to derive closed-form solutions and compute the equilibrium for a broader range of parameters than VV. Using the closed-form solutions, we determine how the supply of government debt affects bond prices, and derive our empirical hypotheses.

1.1 Model

The model is set in continuous time. The term structure at time t consists of a continuum of zero-coupon bonds with maturities in the interval $(0, T]$ and face value one. We denote by $P_t^{(\tau)}$ the price of the bond with maturity τ at time t, and by $y_t^{(\tau)}$ the bond's yield (i.e., the spot rate for maturity τ). The yield is related to the price through

$$y_t^{(\tau)} = -\frac{\log P_t^{(\tau)}}{\tau}. \tag{1}$$

We denote by r_t the short rate, which is the limit of $y_t^{(\tau)}$ when τ goes to zero.

Bonds are issued by a government and are traded by arbitrageurs and other investors. We model explicitly only the arbitrageurs, and treat the demand and supply coming out of the government and the other investors as exogenous and price-inelastic. We assume that arbitrageurs choose a bond portfolio to trade off the instantaneous mean and variance of changes in wealth. Denoting their time-t wealth by W_t and their dollar investment in the bond with maturity τ by $x_t^{(\tau)}$, their budget constraint is

$$dW_t = \int_0^T x_t^{(\tau)} \frac{dP_t^{(\tau)}}{P_t^{(\tau)}} d\tau + \left(W_t - \int_0^T x_t^{(\tau)} d\tau \right) r_t dt. \tag{2}$$

668

The first term in (2) is the arbitrageurs' return from investing in bonds, and the second term is their return from investing their remaining wealth in the short rate. The arbitrageurs' optimization problem is

$$\max_{\{x_t^{(\tau)}\}_{\tau \in (0,T]}} \left[E_t(dW_t) - \frac{a}{2} Var_t(dW_t) \right], \tag{3}$$

where a is a risk-aversion coefficient. One interpretation of the preferences in (3) is that arbitrageurs form overlapping generations, each of which starts with the same level of wealth, lives for a short period, and maximizes expected utility of final wealth. Introducing long-lived arbitrageurs would complicate the optimization problem. Wealth would generally become a state variable, and arbitrageurs could have a hedging demand in addition to the myopic one generated by (3). Within our simple specification (3), we can derive wealth effects as comparative statics by identifying changes in wealth with changes in a. We can also introduce a hedging motive by allowing arbitrageurs to care not only about mean and variance but also about the covariance between changes in wealth and the risk factors. In Appendix B we show that the hedging demand generated by this covariance does not affect our main results.

We assume that the net supply coming out of the government and the other investors is described by a one-factor model: the dollar value of the bond with maturity τ supplied to arbitrageurs at time t is

$$s_t^{(\tau)} = \zeta(\tau) + \theta(\tau)\beta_t, \tag{4}$$

where $\zeta(\tau)$ and $\theta(\tau)$ are deterministic functions of τ, and β_t is a stochastic supply factor. The factor β_t follows the Ornstein-Uhlenbeck process

$$d\beta_t = -\kappa_\beta \beta_t dt + \sigma_\beta dB_{\beta,t}, \tag{5}$$

where $\kappa_\beta > 0$ and $\sigma_\beta > 0$ are constants, and $B_{\beta,t}$ is a Brownian motion. The assumption $\sigma_\beta > 0$ is without loss of generality since we can switch the sign of $B_{\beta,t}$.

Since the supply factor β_t has mean zero, the function $\zeta(\tau)$ measures the average supply for maturity τ. The function $\theta(\tau)$ measures the sensitivity of that supply to β_t. We assume that $\theta(\tau)$ has the following properties.

Assumption 1. The function $\theta(\tau)$ satisfies:

  (i) $\int_0^T \theta(\tau)d\tau \geq 0$.
  (ii) There exists $\tau^* \in [0,T)$ such that $\theta(\tau) < 0$ for $\tau < \tau^*$ and $\theta(\tau) > 0$ for $\tau > \tau^*$.

Part (i) of Assumption 1 requires that an increase in β_t does not decrease the total dollar value of bonds supplied to arbitrageurs. This is without loss of generality since we can switch the sign of β_t. Part (ii) of Assumption 1 allows for the possibility that the supply for some maturities decreases when

β_t increases, even though the total supply does not decrease. The maturities for which supply can decrease are restricted to be at the short end of the term structure. As we show in Section 1.3, Parts (i) and (ii) together ensure that an increase in β_t makes the portfolio that arbitrageurs hold in equilibrium more sensitive to movements in the short rate. This increase in sensitivity is what generates a positive effect of β_t on yields and expected returns.

Assumption 1 includes many cases of interest. One polar case is that an increase in β_t increases supply for each maturity and hence total supply. This case can be derived by setting the threshold τ^* to zero so that $\theta(\tau) > 0$ for all τ. Another polar case is that an increase in β_t leaves total supply constant, but only shifts weight from short maturities to long maturities. This case can be derived by setting $\int_0^T \theta(\tau) d\tau$ to zero.

We treat the short rate r_t as exogenous, but motivated in part by the data, allow it to depend on the supply factor β_t. We assume that r_t follows the Ornstein-Uhlenbeck process

$$dr_t = \kappa_r(\bar{r} - r_t - \gamma\beta_t)dt + \sigma_r dB_{r,t} + \sigma_{r\beta} dB_{\beta,t}, \tag{6}$$

where $\bar{r}$, $\kappa_r > 0$, $\sigma_r > 0$, γ, $\sigma_{r\beta}$ are constants, and $B_{r,t}$ is a Brownian motion independent of $B_{\beta,t}$. The assumption $\sigma_r > 0$ is without loss of generality since we can switch the sign of $B_{r,t}$. The constants γ and $\sigma_{r\beta}$ introduce correlation between r_t and β_t. We mainly focus on the case where r_t and β_t are independent, thus setting $\gamma = \sigma_{r\beta} = 0$, because the independent case is simple and yields the main intuitions. We sketch the analysis of the correlated case at the end of Section 1.3.

1.2 Equilibrium term structure
The two risk factors in our model are the short rate r_t and the supply factor β_t. We next examine how shocks to these factors influence the bond prices $P_t^{(\tau)}$ that are endogenously determined in equilibrium. We solve for equilibrium in two steps: first solve the arbitrageurs' optimization problem for equilibrium bond prices of a conjectured form, and second use market clearing to verify the conjectured form of prices. We conjecture that equilibrium bond yields are affine functions of the risk factors. Bond prices thus take the form

$$P_t^{(\tau)} = e^{-\left[A_r(\tau)r_t + A_\beta(\tau)\beta_t + C(\tau)\right]} \tag{7}$$

for three functions $A_r(\tau)$, $A_\beta(\tau)$ and $C(\tau)$ that depend on maturity τ. The functions $A_r(\tau)$ and $A_\beta(\tau)$ characterize the sensitivity of bond prices to the short rate r_t and the supply factor β_t, respectively, where sensitivity is defined as the percentage price drop per unit of factor increase. Applying Ito's lemma to (7) and using the dynamics of β_t and r_t in (5) and (6) for $\gamma = \sigma_{r\beta} = 0$, we find that the instantaneous return of the bond with maturity τ is

$$\frac{dP_t^{(\tau)}}{P_t^{(\tau)}} = \mu_t^{(\tau)} dt - A_r(\tau)\sigma_r dB_{r,t} - A_\beta(\tau)\sigma_\beta dB_{\beta,t}, \tag{8}$$

670

where

$$\mu_t^{(\tau)} \equiv A_r'(\tau)r_t + A_\beta'(\tau)\beta_t + C'(\tau) + A_r(\tau)\kappa_r(r_t - \bar{r}) + A_\beta(\tau)\kappa_\beta\beta_t + \frac{1}{2}A_r(\tau)^2\sigma_r^2$$

$$+ \frac{1}{2}A_\beta(\tau)^2\sigma_\beta^2 \tag{9}$$

denotes the instantaneous expected return. Substituting bond returns (8) into the arbitrageurs' budget constraint (2), we can solve the arbitrageurs' optimization problem (3).

Lemma 1. The arbitrageurs' first-order condition is

$$\mu_t^{(\tau)} - r_t = A_r(\tau)\lambda_{r,t} + A_\beta(\tau)\lambda_{\beta,t}, \tag{10}$$

where for $i = r, \beta$,

$$\lambda_{i,t} \equiv a\sigma_i^2 \int_0^T x_t^{(\tau)} A_i(\tau)d\tau. \tag{11}$$

According to (10), a bond's instantaneous expected return in excess of the short rate, $\mu_t^{(\tau)} - r_t$, is a linear function of the bond's sensitivities $A_r(\tau)$ to the short rate and $A_\beta(\tau)$ to the supply factor. The coefficients $\lambda_{r,t}$ and $\lambda_{\beta,t}$ of the linear function (which are the same for all bonds) are the prices of short-rate and supply risk, respectively. These coefficients measure the expected excess return per unit of sensitivity to each factor. While we derive (10) from the optimization problem of arbitrageurs with mean-variance preferences, this equation is a more general consequence of the absence of arbitrage: the expected excess return per unit of factor sensitivity must be the same for all bonds, otherwise it would be possible to construct arbitrage portfolios.

Absence of arbitrage imposes essentially no restrictions on the prices of risk $\lambda_{r,t}$ and $\lambda_{\beta,t}$. We determine these instead from market clearing. Equation (11) shows that the price of risk $\lambda_{i,t}$ for factor $i = r, \beta$ depends on the overall sensitivity $\int_0^T x_t^{(\tau)} A_i(\tau)d\tau$ of the arbitrageurs' portfolio to that factor. Intuitively, if arbitrageurs are highly exposed to a factor, they require that any asset they hold yields high expected return per unit of factor sensitivity. The portfolio that arbitrageurs hold in equilibrium is determined from the market-clearing condition

$$x_t^{(\tau)} = s_t^{(\tau)}, \tag{12}$$

which equates the arbitrageurs' dollar investment $x_t^{(\tau)}$ in the bond with maturity τ to the bond's dollar supply $s_t^{(\tau)}$. Substituting $\mu_t^{(\tau)}$ and $x_t^{(\tau)}$ from (4), (9), and (12) into (10), we find an affine equation in r_t and β_t. Setting linear terms in r_t and β_t to zero yields two ordinary differential equations (ODEs) in $A_r(\tau)$ and $A_\beta(\tau)$, respectively. Setting constant terms to zero yields an additional ODE in $C(\tau)$. We solve the three ODEs in Theorem 1.

Theorem 1. The functions $A_r(\tau)$ and $A_\beta(\tau)$ are given by

$$A_r(\tau) = \frac{1 - e^{-\kappa_r \tau}}{\kappa_r}, \tag{13}$$

$$A_\beta(\tau) = \frac{Z}{\kappa_r}\left(\frac{1 - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta} - \frac{e^{-\kappa_r \tau} - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta - \kappa_r}\right), \tag{14}$$

respectively, where

$$Z \equiv a\sigma_r^2 I_r,$$

$$I_r \equiv \int_0^T \frac{1 - e^{-\kappa_r \tau}}{\kappa_r}\theta(\tau)d\tau, \tag{15}$$

and $\hat{\kappa}_\beta$ solves

$$\hat{\kappa}_\beta = \kappa_\beta - a^2\sigma_r^2\sigma_\beta^2 I_r \int_0^T \frac{1}{\kappa_r}\left(\frac{1 - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta} - \frac{e^{-\kappa_r \tau} - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta - \kappa_r}\right)\theta(\tau)d\tau. \tag{16}$$

Equation (16) has a solution if a is below a threshold $\bar{a} > 0$. The function $C(\tau)$ is given by (A.10) in Appendix A.

In the proof of Theorem 1 we show that (16) has an even number of solutions, possibly zero. If the number of solutions is zero, then equilibria with affine yields fail to exist. Otherwise, equilibria exist and are in even number. The mechanism causing the multiplicity is reminiscent of that in DeLong, Summers, Shleifer, and Waldmann (1990), and Spiegel (1998). If yields are highly sensitive to shocks to the supply risk factor β_t, then bonds become highly risky for arbitrageurs. Hence, arbitrageurs absorb supply shocks only if they are compensated by large changes in yields, making the high sensitivity of yields to shocks self-fulfilling.

Equilibria exist if the arbitrageurs' risk-aversion coefficient a is below a threshold $\bar{a} > 0$. We focus on that case, and select the equilibrium corresponding to the largest solution of (16). This equilibrium is well behaved in the sense that when a converges to zero, it converges to the unique equilibrium that exists for $a = 0$.

1.3 Effects of debt supply
We next derive the implications of Theorem 1 for how shocks to the supply factor β_t affect yields and expected returns. We derive our empirical hypotheses from these implications.

Proposition 1 (Supply and Yields). A shock to the supply factor β_t moves the yields of all bonds in the same direction as the shock. Moreover, the effect is either increasing or hump-shaped across maturities.

Proposition 2 (Supply and Expected Returns). A shock to the supply factor β_t moves the instantaneous expected returns of all bonds in the same direction as the shock. Moreover, the effect is increasing across maturities.

That an increase in supply raises the yields and instantaneous expected returns of all bonds appears intuitive: the price of a bond must drop so that risk-averse arbitrageurs are induced to hold the bond's increased supply. Implicit in this explanation, however, is that the increase in supply concerns all bonds. Our definition of the supply factor is more general: Assumption 1 requires that an increase in the supply factor correspond to a (weak) increase in total supply, but allows for the possibility that the supply of short-term bonds decreases.

Why do the prices of short-term bonds decrease when their supply decreases? A bond's price and supply can move in the same direction because supply effects operate not locally, but globally through changes in the prices of risk. Local effects are made global through the activity of arbitrageurs, who integrate the markets for different maturities. Following an increase in the supply factor, the portfolio that arbitrageurs hold in equilibrium becomes more sensitive to changes in the short rate. This is so even when the supply of short-term bonds decreases because overall supply (weakly) increases and long-term bonds are more sensitive to changes in the short rate than short-term bonds. Because arbitrageurs become more exposed to short-rate risk, they become less willing to bear that risk, and that risk's price increases. Since all bonds load positively on short-rate risk, in the sense of experiencing a price drop when the short rate increases, their instantaneous expected return increases. Therefore, the price of all bonds—both short- and long-term—decreases and their yield increases.[3]

The increase in instantaneous expected returns is largest for long-term bonds because they are the most sensitive to risk. The increase in yields, however, can be larger for intermediate-term bonds than for long-term bonds. Intuitively, the effect of a supply shock on a bond's yield is equal to the average effect on the bond's instantaneous expected return over the bond's life. This average effect can be largest for intermediate-term bonds if the shock mean-reverts quickly. Regardless of mean-reversion, however, supply shocks have small effects on the yields and expected returns of short-term bonds. Intuitively, short-term bonds are close substitutes to investing in the short rate, and arbitrageurs can tie their yields closely to current and expected future short rates.

The effect of supply on instantaneous expected returns, derived in Proposition 2, is larger than the effect on yields, derived in Proposition 1. This follows from the property that the effect of a supply shock on a bond's

---

[3] Recent evidence suggests that supply effects can have a significant local component. For example, in September 2011 the Federal Reserve announced its intention to buy Treasury securities with maturities from six to 30 years and sell an equal amount of securities with maturities up to three years. Upon announcement short-term yields increased, contrary to our model. One way to generate more local effects of supply is to assume that the demand and supply for each maturity in the absence of arbitrageurs are price-elastic. See Vayanos and Vila (2009).

yield is equal to the average effect on the bond's instantaneous expected return over the bond's life. The average effect on the bond's expected return is smaller than the effect on the current expected return for two reasons. Since the shock mean-reverts, its effect on the expected return of all bonds dies down over time. And even in the absence of mean reversion, the shock's effect on the expected return of any given bond decreases over time because the bond's time to maturity decreases and so does the bond's sensitivity to changes in the price of short-rate risk.

Proposition 3 (Expected Returns vs. Yields). A shock to the supply factor β_t has a larger effect on instantaneous expected returns than on yields.

Supply can affect prices only when arbitrageurs are risk averse. Indeed, when arbitrageurs are risk neutral, they require no compensation for absorbing supply shocks, and these shocks do not affect prices. More generally, supply has stronger effects when the arbitrageurs' risk aversion coefficient a increases— that is, not only when comparing risk-averse ($a > 0$) to risk-neutral ($a = 0$) arbitrageurs, but also when comparing across any different values of a.

Proposition 4 (Arbitrageur Risk Aversion). The effect of the supply factor β_t on instantaneous expected returns is increasing in the arbitrageurs' risk-aversion coefficient a.

To turn Propositions 1–4 into testable hypotheses, we need to construct empirical measures of supply. Changes in supply in our model are fully described by the single factor β_t, and are hence perfectly correlated across maturities. In practice, however, the correlation might be imperfect and multiple factors might be needed to describe supply. Despite this limitation, our model provides some guidance on suitable measures of supply. Indeed, supply affects the equilibrium through the prices of risk $\lambda_{r,t}$ and $\lambda_{\beta,t}$. Substituting (12) into (11), we find $\lambda_{i,t} = a\sigma_i^2 \Delta_{i,t}$ for $i = r, \beta$, where

$$\Delta_{i,t} \equiv \int_0^T s_t^{(\tau)} A_i(\tau) d\tau. \tag{17}$$

The price of risk for factor $i = r, \beta$ is thus proportional to $\Delta_{i,t}$, the factor sensitivity of the total supply available to arbitrageurs. The quantities $\Delta_{r,t}$ and $\Delta_{\beta,t}$ fully characterize the effects of supply. Suppose, in particular, that supply is described by multiple factors, but one factor suffices to describe the joint dynamics of $\Delta_{r,t}$ and $\Delta_{\beta,t}$. Then the equilibrium is the same as in our model, in which supply is described by one factor. Using the calibration of our model based on postwar U.S. data in Section 4, we can compute the functions $A_r(\tau)$ and $A_\beta(\tau)$, and evaluate the correlation between $\Delta_{r,t}$ and $\Delta_{\beta,t}$ in the data. This correlation is 98%, suggesting that our one-factor description of supply is a good approximation.

The quantities $\Delta_{r,t}$ and $\Delta_{\beta,t}$ are similar to dollar duration. Indeed, the dollar duration of the supply available to arbitrageurs is $\Delta_t \equiv \int_0^T s_t^{(\tau)} \tau d\tau$, the sum of

supply for each maturity τ, weighted by τ. The quantities $\Delta_{r,t}$ and $\Delta_{\beta,t}$ are similar weighted sums, with the weighting function τ being replaced by $A_r(\tau)$ and $A_\beta(\tau)$, respectively. In our empirical analysis, we use dollar duration Δ_t as the basis for our main measure of supply since it is similar in spirit to $\Delta_{r,t}$ and $\Delta_{\beta,t}$, and simpler to construct. Using our calibration, we find that Δ_t has correlation 99% and 97%, respectively, with $\Delta_{r,t}$ and $\Delta_{\beta,t}$ in the data. Hence, it approximates well both quantities.

The dollar duration Δ_t concerns the supply $s_t^{(\tau)}$ available to arbitrageurs, which we do not observe. We proxy this supply by the supply of bonds issued by the government. This proxy is accurate for our empirical purposes if shocks to the supply of government debt affect the bond portfolios held by arbitrageurs and other investors in a proportional manner. Propositions 1–3 yield empirical hypotheses 1–3, respectively.

Hypothesis 1. A regression of bond yields on the dollar duration of government bond supply, controlling for the short rate, has a positive coefficient. This coefficient is either increasing or hump-shaped across maturities.

Hypothesis 2. A regression of future bond returns on the dollar duration of government bond supply, controlling for the short rate, has a positive coefficient. This coefficient is increasing across maturities.

Hypothesis 3. The regression coefficient in Hypothesis 2 is larger than the one in Hypothesis 1.

The regression coefficients in Hypotheses 1 and 2 correspond to the effects of β_t derived in Propositions 1 and 2. These effects are comparative statics, holding the short rate r_t constant. To identify these effects in a regression, we control for the short rate. This control is not necessary when the short rate is independent of supply, but becomes necessary when the two are correlated.

An additional empirical hypothesis follows from Proposition 4, which shows that the effect of supply on instantaneous expected returns is increasing in the arbitrageurs' risk-aversion coefficient a. In our model a is constant over time, and Proposition 4 is a comparative statics result. Stepping outside of the model, however, we can interpret Proposition 4 as concerning the effects of time variation in a. If, in particular, a is decreasing in arbitrageur wealth, then it increases in periods when arbitrageurs lose money. Identifying such periods requires a measure of arbitrageur returns. We use a measure that is implied by our model and is simple to construct. Specifically, arbitrageurs in our model hold large long positions in bonds when β_t is high, and in that case their portfolio is highly sensitive to changes in the short rate and the supply factor. Moreover, when either factor increases, bond prices decrease, especially for long-term bonds. Thus, arbitrageurs lose money when high values of β_t are followed by underperformance of long-relative to short-term bonds. By a similar argument,

they also lose money when low values of β_t are followed by overperformance of long- relative to short-term bonds. We can identify high values of β_t by high values of dollar duration of government bond supply. Using Proposition 1, we can also identify high β_t by high yields of intermediate- and long-term bonds relative to short-term bonds.

Hypothesis 4. The regression coefficient in Hypothesis 2 is decreasing in arbitrageur wealth. Arbitrageur wealth is low in the following two cases:

- Long-term bonds underperform short-term bonds, and this happens following times when the term structure slopes up or the dollar duration of government bond supply is high.
- Long-term bonds overperform short-term bonds, and this happens following times when the term structure slopes down or the dollar duration of government bond supply is low.

Our analysis so far focuses on the case where the short rate r_t and the supply factor β_t are independent. The independent case is derived by setting $\gamma = \sigma_{r\beta} = 0$ in the specification of the short-rate process (6). We can also consider the correlated case, allowing γ and $\sigma_{r\beta}$ to be non-zero. When $\sigma_{r\beta} \neq 0$, supply shocks affect the current short rate. When $\gamma \neq 0$, supply shocks affect expected future short rates, holding the current short rate constant. In Section 4 we find that γ is positive in the data, meaning that an increase in supply lowers expected future short rates. We derive an equilibrium with affine yields in the correlated case in Appendix B. Within this equilibrium, we can show that a positive γ reinforces the result of Proposition 3, shown in the independent case, that supply has a smaller effect on yields than on instantaneous expected returns.

2. Data

2.1 Supply of government debt
Our main sample covers the period from June 1952 to December 2007. We also use a second sample covering the period from June 1916 to June 1940, to evaluate the robustness of our findings. We omit the period between the two samples because the U.S. Federal Reserve was pegging bond yields across the term structure, so variation in yields was limited.[4] We end our main sample in 2007 because we forecast three-year returns, which go until 2010.

To construct our main sample, we collect data from the CRSP historical bond database on every U.S. government bond issued between 1940 and 2007. CRSP provides data on bond characteristics (issue date, coupon rate, maturity, callability features) as well as monthly observations of face value outstanding. As in Doepke and Schneider (2006), we break the stream of each bond's cash

---

[4] See, for example, Hetzel and Leach (2001) and D'Amico, English, et al. (2012) for a description of the Fed's policy during that period.

flows into principal and coupon payments. Consider, for example, the seven-year bond issued in February 1969 (CRSP ID 19760215.206250) with a coupon payment of 6.25%. On the last day of March 1972, investors holding the bond were expecting eight more coupon payments of $3.125 per $100 of face value, starting in August 1972 and ending in February 1976 (the maturity of the bond), with the full principal to be repaid in February 1976. CRSP reports a total face value of $882 million outstanding as of March 1972. Thus, as of the last day of March 1972 there were eight coupon payments of $27.56 million and the principal payment of $882 million.

Despite generally complete data from CRSP, there are some reporting gaps in face values. When these occur, we fill in with the face value outstanding at the end of the previous month. In early years, face values are reported only occasionally. By the early 1950s, face values are reported consistently. We further check the accuracy of the CRSP data by comparing aggregate face values in selected months with releases of the Monthly Statement of the Public Debt.

For a large fraction of securities, CRSP reports both the entire face value and the face value held by the public. The latter measure nets out Federal Reserve and interagency holdings, so it seems a better proxy for the supply of bonds available to arbitrageurs. The face value held by the public, however, is reported only sporadically for some bonds, and tends to be missing for bills until the 1990s. We thus use the entire face value, although we explore corrections for Fed holdings, which we report in our robustness tests in Section 3.3. Simple measures of the average maturity of Fed holdings correlate strongly with the average maturity of all outstanding bonds. Moreover, the size of the Fed's portfolio is positively correlated with the debt-to-GDP ratio, and fluctuated between 4% and 7% of GDP during the 1952–2007 sample period. Taken together, these facts suggest that variation in Fed holdings should generate only small variation in the supply of bonds held by the public prior to 2007.

We construct the maturity structure of government debt at a given date by aggregating cash flows across individual bonds. Total payments due τ years from date t are

$$D_t^{(\tau)} = PR_t^{(\tau)} + C_t^{(\tau)} = \sum_i PR_{it}^{(\tau)} + \sum_i C_{it}^{(\tau)},$$

where $PR_t^{(\tau)}$ are total principal payments, derived by summing over bonds the principal payment $PR_{it}^{(\tau)}$ that each bond i is due to make τ years from date t, and $C_t^{(\tau)}$ are total coupon payments, derived by summing over bonds the coupon payment $C_{it}^{(\tau)}$ that each bond i is due to make τ years from date t. Figure 1 shows the time-series average maturity structure of total payments scaled by GDP. The figure marks principal and coupon payments separately.

Following the theoretical discussion in Section 1, we construct our main measure of the supply of government debt based on dollar duration. Our main
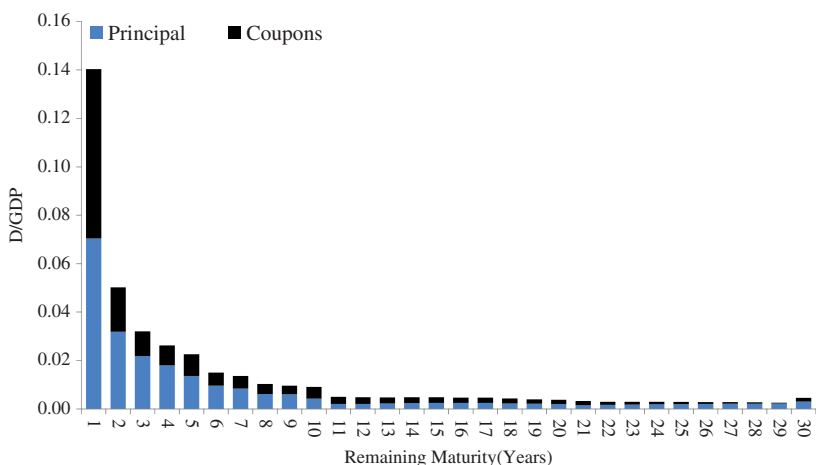
Figure 1
Principal and coupon payments
The time-series average, for each maturity year, of total payments on bills, bonds, and notes scaled by GDP. The bottom bars denote principal payments. The darker top bars denote coupon payments. Any payments beyond 30 years are included in the 30-year bucket. The data are based on the CRSP bond database and cover the period June 1952–December 2007.

measure is the maturity-weighted-debt-to-GDP ratio

$$\left(\frac{MWD}{GDP} \right)_t = \frac{\sum_{0<\tau\leq30} D_t^{(\tau)}\tau}{GDP_t},$$

computed by multiplying the payments $D_t^{(\tau)}$ for each maturity τ times τ, summing across maturities, and scaling by GDP. Maturity-weighted debt is similar to dollar duration of debt, except that we express the payments $D_t^{(\tau)}$ in face-value rather than market-value terms. Following Krishnamurthy and Vissing-Jorgensen (2012), we also measure supply by the long-term-debt-to-GDP ratio. We compute this ratio

$$\left(\frac{LTD}{GDP} \right)_t = \frac{\sum_{10\leq\tau\leq30} D_t^{(\tau)}}{GDP_t},$$

by summing payments $D_t^{(\tau)}$ across all maturities τ longer than ten years and scaling by GDP. This measure is similar in spirit to our main measure, except that the weighting function is zero for maturities below ten years and one for maturities above.

We express debt payments in face-value rather than market-value terms to avoid an endogeneity problem: bond yields and returns, our dependent variables, have a mechanical effect on supply, our independent variable, if the latter is computed using market values. This effect tends to generate a spurious negative relationship between supply and yields or returns. For example, a

decrease in the demand for long-term bonds by investors would lower bond prices, and raise yields and expected returns. It would also lower maturity-weighted debt and long-term debt if these are computed using market values, thus generating a negative relationship. As we argue in Section 3.2, endogeneity concerns are not entirely avoided when our measures are computed using face values. To address these concerns, we perform instrumental-variables regressions. In Section 3.3 we also reestimate our regressions with market-value counterparts of our supply measures, and show that our main results are robust.[5]

A useful decomposition of maturity-weighted debt to GDP is

$$\left(\frac{MWD}{GDP}\right)_t = M_t \cdot \frac{\sum_{0 < \tau \leq 30} D_t^{(\tau)}}{GDP_t}, \tag{18}$$

where

$$M_t = \frac{\sum_{0 < \tau \leq 30} D_t^{(\tau)} \tau}{\sum_{0 < \tau \leq 30} D_t^{(\tau)}}.$$

The variable M_t is dollar-weighted average maturity, constructed by weighting each maturity τ by the fraction that the corresponding payments $D_t^{(\tau)}$ represent of total payments. The variable $\sum_{0 < \tau \leq 30} D_t^{(\tau)} / GDP_t$ is total debt payments divided by GDP. It differs from the standard debt-to-GDP ratio (such as described in Bohn 2008) because it includes coupon payments but does not include non-marketable debt such as intragovernmental obligations. Despite these differences, it is highly correlated with debt to GDP: the correlation is 92% in the 1952–2007 sample. Equation (18) thus implies that maturity-weighted debt to GDP can be thought of intuitively as the product of average maturity times debt to GDP.

While variation in maturity-weighted debt to GDP can be decomposed into two distinct components, variation in average maturity and in debt to GDP, these two components are strongly positively correlated: the correlation between dollar-weighted average maturity and debt to GDP is 60% in the 1952–2007 sample. The positive correlation reflects the fact that as the U.S. government increased the size of its debt, it issued a larger fraction of it long-term to reduce the risk of having to refinance large amounts of short-term debt at high rates (Greenwood, Hanson, and Stein 2010). The strong correlation makes it somewhat difficult to discern whether the effects of debt supply are driven by maturity-weighted debt to GDP, or by average maturity, or by debt to GDP. Nevertheless, in our main return-forecasting regressions, maturity-weighted debt to GDP drives out either of the other two variables in horse races, as

---

[5] Our results are also robust to scaling by household net worth from Table B100 of the Flow of Funds instead of by GDP. Either scaling can be viewed a simple way to adjust for time-series variation in the risk-bearing capacity of arbitrageurs. Finally, our results would not be affected by taking bond callability into account. This is because there are few callable bonds, and most of them are callable close to their maturity date.
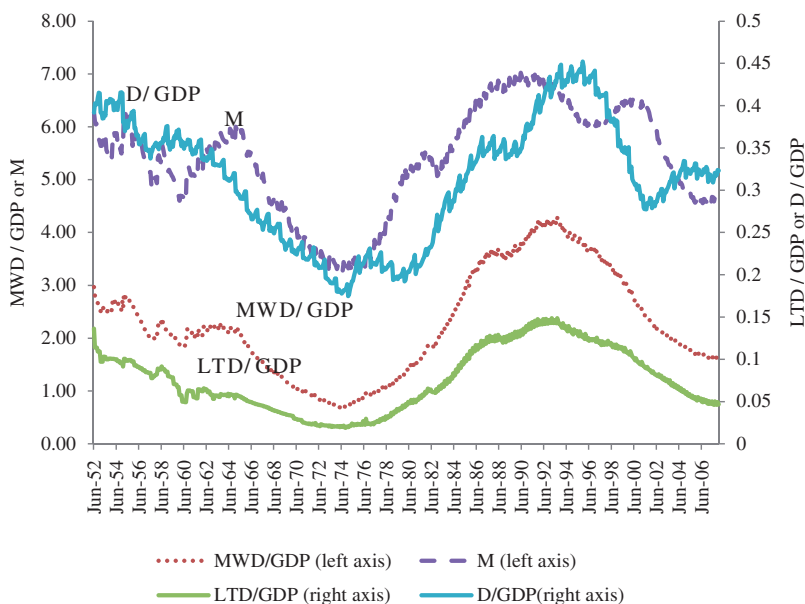
Figure 2
Bond supply, 1952–2007
MWD/GDP is the maturity-weighted-debt-to-GDP ratio, computed by multiplying each debt payment by the corresponding maturity,summing across maturities, and scaling by GDP. LTD/GDP is the long-term-debt-to-GDP ratio, computed by summing all debt payments with maturity beyond ten years, and scaling by GDP. M is the dollar-weighted average maturity expressed in years. D/GDP is the ratio of the aggregate principal payments of all Treasury securities to GDP. MWD/GDP, LTD/GDP, and M are computed using aggregate principal and coupon payments.

we show in Table C.4 in online Appendix C. Thus, average maturity brings useful additional information relative to debt to GDP in forecasting returns, and conversely, debt to GDP brings useful information relative to average maturity.

Figure 2 plots dollar-weighted average maturity, debt to GDP, and our two measures of debt supply for the 1952–2007 sample. Panel A of Table 1 reports summary statistics for these variables. Figure 2 shows that maturity-weighted debt to GDP varies significantly over time. For example, it decreased sharply from the mid-1960s to the mid-1970s, to a minimum value of 79.9%, and increased sharply from the mid-1970s to the early 1990s, to a maximum value of 463%. These movements were driven by variation in both average maturity and debt to GDP.

The sharp drop in average maturity from the mid-1960s to the mid-1970s, and the subsequent rise, were partly driven by the 4.5% regulatory ceiling on bonds' coupon rates. Because of the ceiling, the Treasury did not issue bonds between 1965 and 1973, leading to a decline in average maturity. Maturity started increasing in 1976, when Congress raised the maturity of notes, to which the ceiling did not apply, to ten years. The ceiling was eliminated in 1988. An additional driver of the rise in average maturity during the 1980s

Table 1
Summary Statistics

	Mean	Median	SD	Min	Max
Panel A. Main sample 1952–2007					
Measures of Debt Maturity and Debt Supply:					
MWD/GDP	2.284	2.142	0.997	0.671	4.275
MWD/GDP (MV-based)	1.050	1.117	0.444	0.338	2.127
LTD/GDP	0.077	0.070	0.038	0.019	0.149
LTD/GDP (MV-based)	0.047	0.046	0.025	0.011	0.116
M	5.387	5.511	1.016	3.269	7.024
D/GDP	0.313	0.322	0.073	0.175	0.452
Returns and Yields:					
$y^{(LT)}$	0.063	0.060	0.025	0.026	0.138
$y^{(LT)} - y^{(1)}$	0.008	0.008	0.012	−0.032	0.041
$y^{(1)}$	0.055	0.053	0.029	0.006	0.158
r_1	0.062	0.049	0.093	−0.187	0.434
r_3	0.189	0.187	0.165	−0.187	0.680
r_5	0.320	0.312	0.232	−0.166	1.100
Macroeconomic Conditions and Other Controls:					
Output gap	0.000	0.000	0.016	−0.048	0.038
Output growth	0.033	0.033	0.024	−0.031	0.091
Inflation	0.038	0.031	0.028	−0.009	0.138
Inflation risk	0.002	0.002	0.001	0.001	0.006
Interest-rate risk	0.007	0.006	0.006	0.001	0.035
Stock-market risk	0.007	0.006	0.004	0.002	0.049
MWD/GDP (Fed adjusted)	2.154	2.044	1.001	0.555	4.134
Arbitrageur Wealth:					
ΔW_1^{Arb}	0.000	0.000	0.001	−0.004	0.007
ΔW_2^{Arb}	0.040	0.006	0.225	−0.673	0.958
Panel B. Prewar sample 1916–1940					
MWD/GDP	3.062	3.500	0.974	0.552	4.264
LTD/GDP	0.128	0.151	0.045	0.022	0.181
M	12.857	12.307	3.572	9.153	23.377
D/GDP	0.250	0.238	0.104	0.024	0.426
$y^{(LT)}$	0.035	0.036	0.008	0.022	0.055
$y^{(1)}$	0.023	0.029	0.019	0.000	0.058
r_1	0.000	0.000	0.000	0.000	0.001
r_3	0.031	0.035	0.010	0.006	0.043

Panel A summarizes the main sample 1952–2007. MWD/GDP is the maturity-weighted-debt-to-GDP ratio. Its market-value-based version multiplies the market value of each bond by Macaulay duration, sums across bonds, and scales by GDP. LTD/GDP is the long-term-debt-to-GDP ratio. Its market-value-based version sums the market values of all bonds with maturity beyond ten years, and scales by GDP. M is the dollar-weighted average maturity expressed in years. D/GDP is the ratio of the aggregate principal payments of all Treasury securities to GDP. $y^{(LT)}$ is the yield of a long-term bond with approximate maturity twenty years, and $y^{(1)}$ is the one-year yield. r_1, r_2, and r_3 are holding-period returns for the long-term bond over one-, three-, and five-year horizons, respectively. Output gap is the residual from a Hodrick-Prescott filter of log GDP. Output growth is the difference between log real GDP in the most recent quarter t and log real GDP in quarter $t-4$. Inflation risk is the standard deviation of monthly inflation over the past year. Interest-rate risk is the standard deviation of the monthly short-term Treasury-bill yield over the past year. Stock-market risk is the standard deviation of daily CRSP value-weighted stock returns over the past month. Change in arbitrageur wealth ΔW_1^{Arb} is the product of the one-year lagged spread between the long-term and the one-year yield, times the subsequent one-year excess return of the long-term bond. Change in arbitrageur wealth ΔW_2^{Arb} is the product of the one-year lagged MWD/GDP, times the subsequent one-year excess return of the long-term bond. Panel B summarizes the prewar 1916–1940 sample.

was the expansion of government debt. Indeed, the Treasury issued at long maturities to reduce the risk of having to refinance large amounts of short-term debt at high interest rates.[6] The sharp increase in debt during the 1980s and early 1990s was driven by a combination of tax cuts and increased military spending.

To construct our second sample, we collect data from Banking and Monetary Statistics (BMS). BMS reports the maturity structure of government debt in six- and twelve-month intervals beginning in June 1916. Using these data, we construct maturity-weighted debt to GDP. We use the same definition as for the 1952–2007 sample, but because the data are coarser, we construct our measure slightly differently. BMS groups bonds into maturity buckets and reports total face value within a bucket. We assign the average maturity of each bucket to all bonds in that bucket—for example, all bonds in the five- to ten-year bucket are assigned maturity 7.5 years. Moreover, we do not take into account coupon payments because they are not reported in the BMS data.

Figure 3 plots dollar-weighted average maturity, debt to GDP, and our two measures of debt supply as a function of time for the 1916–1940 sample. Panel B of Table 1 reports summary statistics of the same variables. Figure 3 shows that maturity-weighted debt to GDP varies significantly over time. Its movements parallel those of debt to GDP, as average maturity is approximately flat in most of the sample. Debt to GDP was nearly zero in 1916, then rose sharply during World War I, and then declined during the 1920s. It rose sharply again during the early 1930s, as GDP decreased during the Great Depression and spending on social programs increased.

2.2 Bond yields and returns

We use the Fama-Bliss discount bond database to obtain yields and holding-period returns for one-, two-, three-, four-, and five-year zero-coupon bonds for the 1952–2007 sample period. Beyond five years, yields are not available for most maturities. However, Ibbotson Associates provides yields and returns for a bond with an approximate maturity of twenty years, and we use this to obtain a long-term yield and return. For the 1971–2007 period, we use the zero-coupon curves provided by Gurkaynak, Sack, and Wright (2007) to obtain yields and returns for bonds with maturities of up to fifteen years. For the 1916–1940 period, we use Global Financial Data to obtain the yield and return of a bond with an approximate maturity of ten years. We do not have one-year zero-coupon yields (the Fama-Bliss database starts in 1952), so we use instead monthly Treasury-bill yields rolled over one year as our measure of the one-year yield.

Yields and returns are computed in logs. We denote by $y_t^{(\tau)}$ the yield of the τ-year bond at date t. (This is consistent with the notation in our model for the

---

[6] A detailed discussion of the variation in the maturity of government debt from 1952 onward is in Garbade (2007).
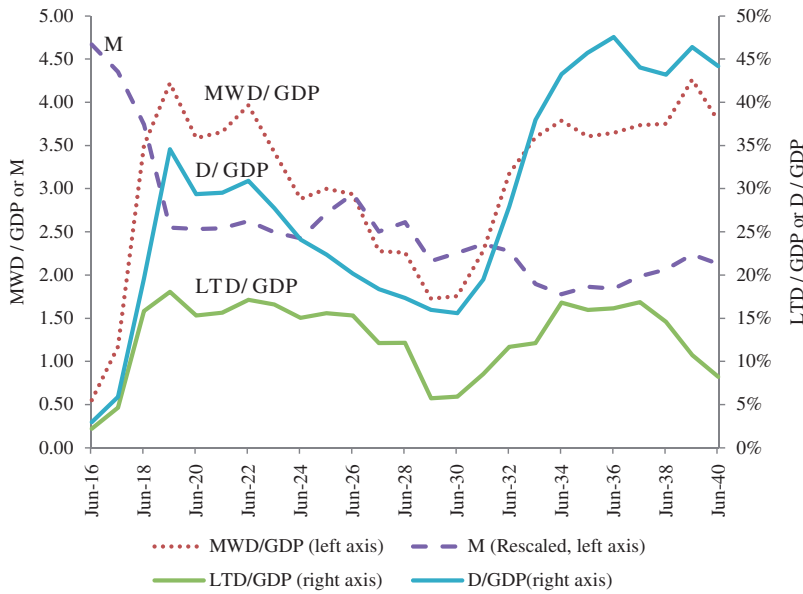
Figure 3
Bond supply, 1916–1940
MWD/GDP is the maturity-weighted-debt-to-GDP ratio, computed by multiplying each debt payment by the corresponding maturity, summing across maturities, and scaling by GDP. LTD/GDP is the long-term-debt-to-GDP ratio, computed by summing all debt payments with maturity beyond ten years, and scaling by GDP. M is the dollar-weighted average maturity expressed in years and divided by five (to fit into the picture). D/GDP is the ratio of the aggregate principal payments of all Treasury securities to GDP. MWD/GDP, LTD/GDP, and M are computed using aggregate principal payments.

one- to five-year bonds because they are zero-coupon, and for simplicity we also use this notation for the long-term coupon bond.) We denote by $r_{t+1}^{(\tau)}$ the return of the τ-year bond during the year following date t, and by

$$r_{t+k,k}^{(\tau)} \equiv \sum_{k'=1}^{k} r_{t+k'}^{(\tau-k'+1)}$$

the bond's cumulative return during the k years following date t.

3. Results

3.1 Basic tests

Table 2 shows regressions of yields and future returns on our two measures of government debt supply: maturity-weighted debt to GDP, and long-term debt to GDP. The yield regression is

$$y_t^{(\tau)} = a + bX_t + cy_t^{(1)} + u_t, \tag{19}$$

where $y_t^{(\tau)}$ is the yield on the τ-year bond, X_t is the measure of supply, and $y_t^{(1)}$ is the one-year yield that we use as a control for the short rate. Observations are

683

Table 2
Bond supply, bond yields, and bond returns

	X = MWD/GDP							X = LTD/GDP						
	b	[t NW]	[t AR]	c	[t NW]	[t AR]	R^2	b	[t NW]	[t AR]	c	[t NW]	[t AR]	R^2
Yield spreads:														
Yield 2-yr bond	0.001	[2.597]	[2.363]	0.981	[50.113]	[70.970]	0.987	0.029	[2.476]	[2.293]	0.982	[49.381]	[69.955]	0.987
Yield 3-yr bond	0.002	[2.510]	[1.881]	0.951	[29.510]	[36.999]	0.968	0.044	[2.364]	[1.811]	0.952	[29.150]	[36.591]	0.968
Yield 4-yr bond	0.002	[2.497]	[1.805]	0.932	[22.657]	[30.645]	0.949	0.058	[2.356]	[1.772]	0.934	[22.419]	[30.362]	0.948
Yield 5-yr bond	0.002	[2.358]	[1.580]	0.913	[19.528]	[23.621]	0.933	0.064	[2.258]	[1.601]	0.914	[19.387]	[23.506]	0.932
Yield LT bond	0.004	[2.682]	[1.719]	0.795	[12.167]	[12.993]	0.379	0.107	[2.610]	[1.822]	0.797	[12.234]	[13.253]	0.374
Returns:														
1-year return 2-yr bond	0.004	[1.979]	[1.438]	1.114	[12.201]	[11.290]	0.774	0.116	[2.176]	[1.590]	1.118	[12.214]	[11.245]	0.776
1-year return 3-yr bond	0.007	[1.860]	[1.512]	1.134	[6.751]	[6.847]	0.507	0.191	[2.013]	[1.627]	1.140	[6.750]	[6.804]	0.509
1-year return 4-yr bond	0.010	[1.964]	[1.774]	1.157	[4.864]	[4.535]	0.358	0.266	[2.084]	[1.867]	1.166	[4.855]	[4.495]	0.360
1-year return 5-yr bond	0.011	[1.902]	[1.852]	1.145	[3.897]	[4.172]	0.263	0.308	[2.012]	[1.913]	1.154	[3.897]	[4.132]	0.265
1-year return LT bond	0.026	[3.097]	[3.462]	1.212	[2.846]	[3.181]	0.190	0.685	[3.196]	[3.468]	1.229	[2.860]	[3.142]	0.189
3-year return LT bond	0.065	[4.200]	[4.121]	3.737	[4.971]	[4.587]	0.506	1.786	[4.200]	[4.284]	3.795	[5.039]	[4.627]	0.516
5-year return LT bond	0.094	[5.421]	[3.580]	6.139	[5.401]	[4.650]	0.658	2.625	[5.340]	[4.068]	6.235	[5.612]	[5.062]	0.675

Monthly time-series regressions of the form:

$$y_t^{(\tau)} = a + bX_t + cy_t^{(1)} + u_t$$

$$r_{t+k,k}^{(\tau)} = a + bX_t + cy_t^{(1)} + u_{t+k}$$

The dependent variable is the yield or the one-year, three-year, or five-year return of the τ-year bond. The independent variable X_t is MWD/GDP, the maturity-weighted-debt-to-GDP ratio, or LTD/GDP, the long-term-debt-to-GDP ratio. The regressions control for the one-year yield. The first set of t-statistics, reported in brackets, is based on Newey-West standard errors, with 36 lags in the case of the yield and one-year return regressions, and 54 and 90 lags in the case of the three- and five-year return regressions. The second set of t-statistics is based on modeling the error process as AR(1) for the yield regressions, and as ARMA(1, k) for the return regressions where k denotes the number of months in the return cumulation (e.g., twelve for the one-year return).

monthly. We include the short-rate control in all our regressions for yields and returns because the short rate in the data is negatively correlated with supply. In Table C.1 in online Appendix C we show that omitting this control and using as dependent variables yield spreads instead of yields and excess returns instead of returns does not affect our results.

The results of the yield regression are in the first five rows of Table 2. Because yields can depend on persistent variables other than supply and the short rate (e.g., expected future short rates), the regression residuals are serially correlated and t-statistics must be adjusted accordingly. We report a first set of t-statistics using Newey and West (1987) standard errors and allowing for 36 months of lags. Allowing for more lags does not seem to affect the results. We also report a second set of t-statistics computed by estimating an AR(1) process for the regression residuals. These are lower than in Table 2, but the positive relationship between supply and yields remains statistically significant (at the one-sided 5% level) in most specifications. Our findings thus support Hypothesis 1.

To evaluate the economic significance of the yield results, we note, for example, that when supply is measured by maturity-weighted debt to GDP, it carries a coefficient of 0.004 in the regression of the long-term yield. Thus, an increase in supply by one standard deviation (0.997 from Table 1), holding the one-year yield constant, is associated with an increase of 40 basis points (bps) in the long-term yield. This is about one-half of the long-term yield's standard deviation conditional on the one-year yield. Measuring supply by long-term debt to GDP yields similar results.

Our time-series estimates of the link between supply and yields are somewhat smaller than estimates of the price impact of recent quantitative easing (QE) programs undertaken in the United Kingdom and the United States. During a first QE program in 2009–2010, the U.S. Federal Reserve bought $300 billion of Treasury securities and about $1 trillion of other securities such as agency and mortgage-backed. It bought an additional $600 billion of Treasury securities during a second QE program in 2010–2011. The average maturity of Treasury securities purchased by the Fed was approximately 6.5 years (Figure 1 of D'Amico and King 2013). Taking the corresponding duration to be five years and GDP to be $14 trillion, the reduction in maturity-weight debt to GDP was $0.9 \times 5/14 = 0.32$ if only Treasury securities are included, and $1.9 \times 5/14 = 0.68$ if the other securities are also included. Gagnon et al. (2011), D'Amico, English, et al. (2012), and Li and Wei (2012) estimate that the two QE programs together lowered the ten-year Treasury yield by 90–100 bps.

Between 2009 and mid-2010, the Bank of England bought £200 of Treasury securities. The average maturity of the purchased bonds was approximately 14.5 years (Chart 4 of Joyce, Lasaosa, et al. 2011). Taking the corresponding duration to be ten years and GDP to be £1.5 trillion, the reduction in maturity-weight debt to GDP was $0.2 \times 10/1.5 = 1.33$. Joyce, Lasaosa, et al. (2011) estimate that the QE program lowered the ten-year Treasury yield by about 100 bps.

Our estimate that a unit decrease in maturity-weighted debt to GDP lowers the long-term yield by 40 bps is somewhat smaller than the QE estimates. This could be because of the higher risk aversion during the QE period and the financial crisis. Moreover, part of the QE effect was due to a decrease in expected future short rates (about one-third in the United States according to Krishnamurthy and Vissing-Jorgensen 2012), while in our time-series analysis decreases in bond supply are associated with increases in expected future short rates.

We next turn to the results on returns. Panel A of Figure 4 plots maturity-weighted debt to GDP (horizontal axis) against the subsequent three-year excess return of the long-term bond. The series are sampled annually, in December. The figure shows a positive correlation. We complement the figure with the return regression

$$r_{t+k,k}^{(\tau)} = a + bX_t + cy_t^{(1)} + u_{t+k}, \tag{20}$$

where $r_{t+k,k}^{(\tau)}$ is the future k-year return of the τ-year bond, X_t is supply, and $y_t^{(1)}$ is the one-year yield. Observations are monthly. We perform this regression for one-year returns for all bonds in our sample, and for three- and five-year returns for the long-term bond.

The results of the return regression are in the last seven rows of Table 2. As in the case of yields, t-statistics must be adjusted for serial correlation in the regression residuals. In the case of returns, serial correlation arises from two sources. First, persistent variables other than supply and the short rate can affect expected returns (e.g., macroeconomic variables, investor demand). Second, because returns are measured over one or more years but are sampled monthly, measurement periods overlap. Sampling returns annually eliminates the overlap problem in the case of one-year returns. Results for annual sampling, shown in Table 5, are similar to those in Table 2.

We report a first set of t-statistics in Table 2 using Newey and West (1987) standard errors and allowing for $\min\{36, 1.5k\}$ months of lags, where k is the forecast horizon in years (thus, 36 months for one-year returns, 54 months for three-year returns, and 90 months for five-year returns). Allowing for more lags does not seem to affect the results. We also report a second set of t-statistics obtained by estimating a parametric process for the regression residuals. As Cochrane (2008) points out, a plausible process in the case of one-year returns and annual sampling is ARMA(1,1): such a process would arise if we assume that the annual residuals are the sum of a white-noise component and an expected-return component that is AR(1). Under monthly sampling, the same assumptions on the monthly residuals generate instead an ARMA(1,12k) process for k-year returns. The t-statistics based on the ARMA(1,12k) process tend to be slightly lower, but the positive relationship between supply and future returns remains statistically significant in most specifications. For example, for three-year returns the t-statistic is 4.200 under Newey and West, and 4.203 under ARMA(1,36); for five-year returns it is 5.381 under Newey and West, and 3.824 under ARMA(1,60). Our findings thus support Hypothesis 2.
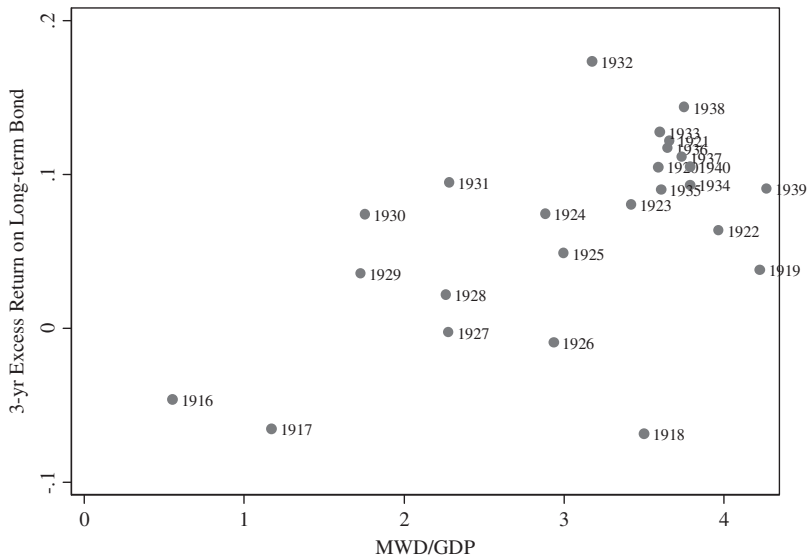
Panel A. 1952—2007
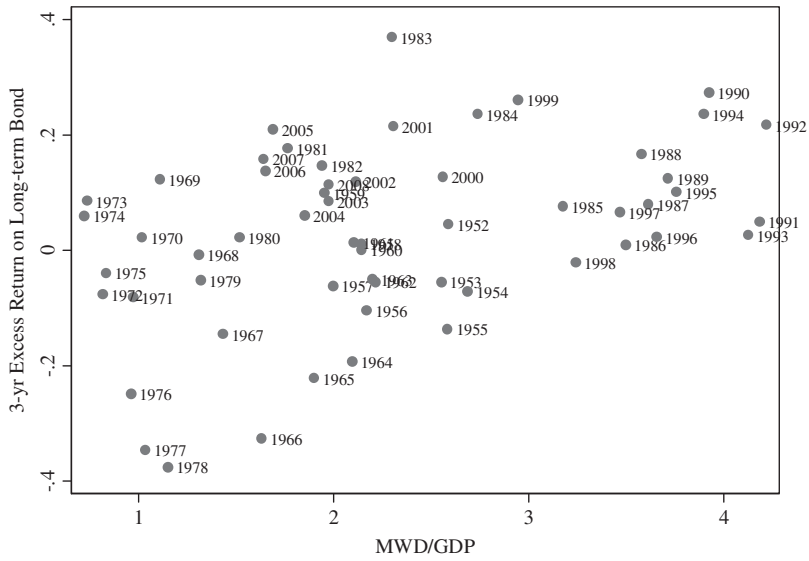
Panel B. 1916—1941

Figure 4
Bond supply and excess bond returns
Plots of three-year holding-period excess return on long-term government bonds (vertical axis) against the maturity-weighted-debt-to-GDP ratio (horizontal axis). Panel A shows the 1952–2007 period. Panel B shows the 1916–1941 period.

As an additional robustness check for our yield and return regressions, we use the bootstrap approach suggested by Bekaert and Hodrick (2001) and Bekaert, Hodrick, and Marshall (2001). We compute bootstrapped *p*-values by comparing the Newey and West *t*-statistic to the distribution of bootstrapped *t*-statistics. To preserve the time-series dependence of the original data, we create pseudo time series using the stationary block bootstrap of Politis and Romano (1994). We repeat this exercise, varying the block size between 12 months and 24 years (288 months). When supply is measured by maturity-weighted debt to GDP, *p*-values range between 3.8% and 7.2% for the yield of the long-term bond and between 0.6% and 2.3% for that bond's one-year return (Table C.3 in online Appendix C).[7]

To evaluate the economic significance of the return results, we note, for example, that when supply is measured by maturity-weighted debt to GDP, it carries a coefficient of 0.026 in the regression of the one-year return of the long-term bond. Thus, an increase in supply by one standard deviation, holding the one-year yield constant, is associated with an increase of 259 bps in the expected one-year return of the long-term bond. This is about one-third of that return's standard deviation conditional on the one-year yield. Measuring supply by long-term debt to GDP yields similar results.

The effects of supply on yields are smaller than on expected returns: the coefficients in the yield regression in the first five rows of Table 2 are smaller than their counterparts in the return regression in the next five rows of the table. This is consistent with Hypothesis 3. As we point out in Section 1, a smaller effect of supply on yields than on returns is to be expected for two reasons. First, supply shocks in the data are negatively correlated with changes in future short rates, and this dampens the effect that they have on yields through expected returns. Second, even in the absence of correlation, the effect of a supply shock on a bond's expected return dies down over time both because (i) the shock mean-reverts, and (ii) the bond's time to maturity decreases and so does the bond's sensitivity to changes in the price of short-rate risk. The shock's mean-reversion can be caused by mean-reversion in the supply of bonds by the government, or by entry of new capital in the market to absorb the shock. Since the shock's effect on a bond's expected return dies out over time, the shock's effect on the bond's yield, which is the average effect on the bond's expected return over the bond's life, is smaller than the effect on the bond's current expected return.

According to Hypothesis 2, the coefficient of supply in the return regression should be increasing with maturity because longer-maturity bonds are more sensitive to changes in the price of short-rate risk. Moreover, according to

---

[7] An additional concern related to statistical significance is that the coefficient of the return regression may be biased if innovations in the forecasting variable, i.e., supply, are correlated with innovations in returns (Mankiw and Shapiro 1986; Stambaugh 1986). This bias is small in our data and works against us because it lowers the regression coefficient.

Table 3
Bond supply, bond yields, and bond returns: Regression coefficients as a function of maturity

	Yields				1-yr returns			
	b	[t NW]	[t AR]	R^2	b	[t NW]	[t AR]	R^2
$\tau =$								
2-yr	0.0009	[2.051]	[1.567]	0.985	0.005	[2.143]	[1.484]	0.752
3-yr	0.0014	[2.016]	[1.360]	0.961	0.008	[2.134]	[1.659]	0.489
4-yr	0.0017	[2.045]	[1.303]	0.938	0.011	[2.223]	[1.881]	0.339
5-yr	0.0020	[2.087]	[1.280]	0.916	0.014	[2.340]	[2.088]	0.254
6-yr	0.0023	[2.133]	[1.271]	0.898	0.016	[2.459]	[2.267]	0.203
7-yr	0.0025	[2.178]	[1.270]	0.882	0.019	[2.569]	[2.421]	0.171
8-yr	0.0026	[2.222]	[1.275]	0.868	0.022	[2.663]	[2.555]	0.151
9-yr	0.0028	[2.263]	[1.286]	0.856	0.024	[2.740]	[2.671]	0.137
10-yr	0.0029	[2.301]	[1.301]	0.846	0.027	[2.800]	[2.772]	0.128
11-yr	0.0030	[2.334]	[1.319]	0.836	0.029	[2.845]	[2.858]	0.121
12-yr	0.0030	[2.363]	[1.338]	0.827	0.032	[2.878]	[2.933]	0.115
13-yr	0.0031	[2.386]	[1.356]	0.819	0.034	[2.901]	[2.998]	0.111
14-yr	0.0032	[2.403]	[1.373]	0.812	0.037	[2.916]	[3.054]	0.107
15-yr	0.0032	[2.414]	[1.388]	0.805	0.039	[2.926]	[3.104]	0.104

Monthly time-series regressions of the form:

$$y_t^{(\tau)} = a + b_\tau (MWD/GDP)_t + cy_t^{(1)} + u_t, \text{ and}$$

$$r_{t+k,k}^{(\tau)} = a + b_\tau (MWD/GDP)_t + cy_t^{(1)} + u_{t+k}.$$

for each maturity τ between two and fifteen years, during the 1971–2007 period. The dependent variable is the yield or the one-year return of the τ-year bond. The independent variable is *MWD/GDP*, the maturity-weighted-debt-to-GDP ratio. The regressions control for the one-year yield. The first set of *t*-statistics, reported in brackets, is based on Newey-West standard errors with 36 lags. The second set of *t*-statistics is based on modeling the error process as AR(1) for the yield regressions, and as ARMA(1,12) for the return regressions.

Hypothesis 1, the coefficient of supply in the yield regression should be increasing or hump-shaped. Table 2 shows an increasing pattern for the five maturities that are available in our 1952–2007 sample. To examine whether this pattern holds for a larger set of maturities, we focus on the subsample 1971–2007 for which Gurkaynak, Sack, and Wright (2007) provide zero-coupon yields for all maturities between one and 15 years. Table 3 reports the coefficients from our yield and return regressions in the subsample, in the case of one-year returns and maturity-weighted debt to GDP. Figure 5 plots these coefficients as a function of maturity. The table and the figure confirm the increasing pattern, while also showing that the effects of supply on returns remain significant in the subsample.

3.2 Instrumental variables tests

One concern with our analysis is that our measures of supply could be endogenous and affected by variables that also affect bond yields and returns. As we point out in Section 2, one endogeneity problem arises when supply is measured in market-value terms because it is then affected mechanically by bond yields and returns. But endogeneity could arise even when supply is measured in face-value terms. Suppose, for example, that the government chooses maturity structure to minimize the expected interest payments on its debt. Then, an increase in the demand for long-term bonds by investors
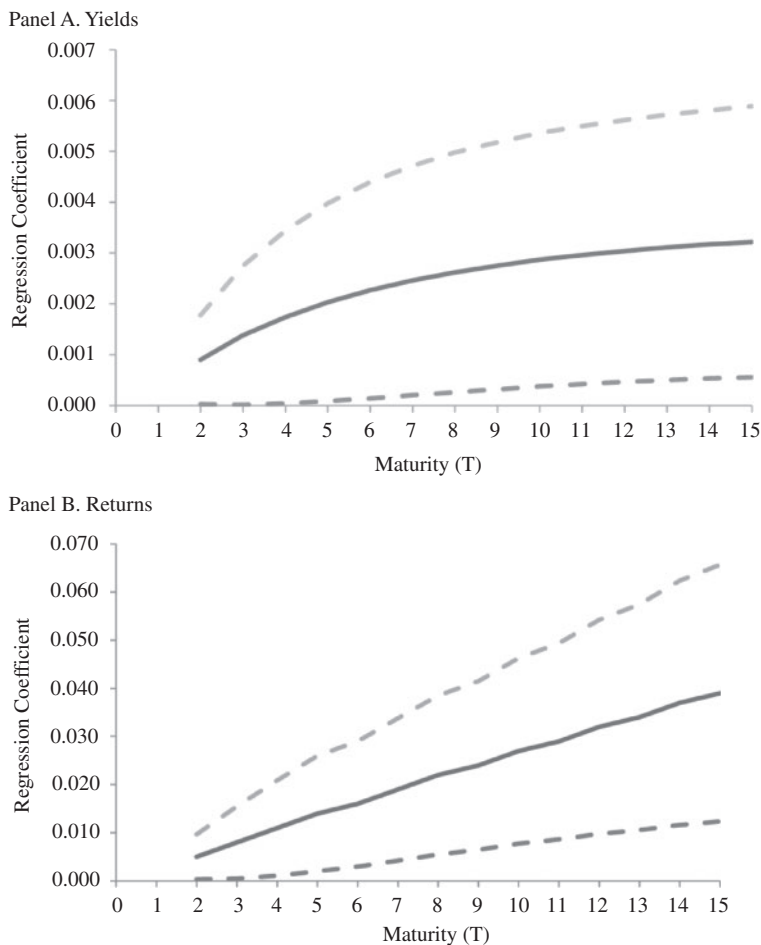
Panel A. Yields

Panel B. Returns

Figure 5
Bond supply, bond yields, and bond returns: Regression coefficients as a function of maturity
We use the data provided by Gürkaynak, Sack, and Wright (2007) to obtain yields and one-year returns for zero-coupon bonds with maturities between two and fifteen years during the November 1971–December 2007 period. For each maturity τ, we estimate monthly time-series regressions of the form:

$$y_t^{(\tau)} = a + b_\tau (MWD/GDP)_t + cy_t^{(1)} + u_t, \text{ and}$$

$$r_{t+k,k}^{(\tau)} = a + b_\tau (MWD/GDP)_t + cy_t^{(1)} + u_{t+k}.$$

In the first equation the yield is regressed on the maturity-weighted-debt-to-GDP ratio controlling for the one-year yield. In the second equation the dependent variable is instead the one-year return. The figures below show the coefficients b_τ as a function of τ, together with 95% confidence intervals.

would lower their yields and induce the government to shift the issuance of its debt toward long maturities.[8] The average maturity of government debt would then increase, and so would maturity-weighted debt to GDP, which is (approximately) the product of average maturity times debt to GDP. This would generate a negative relationship between supply and yields or returns, and would bias our analysis toward finding smaller effects.

To address the possible endogeneity of the maturity structure of government debt, we follow the instrumental variables (IV) approach of Krishnamurthy and Vissing-Jorgensen (KVJ 2012). KVJ use debt to GDP, as well as its square and cube, as instruments for maturity structure. Debt to GDP is a suitable instrument because it is driven mostly by the cumulation of past deficits rather than by changes in investor demand. It causes variation in maturity-weighted debt to GDP through two channels. The first is mechanical: holding average maturity constant, maturity-weighted debt to GDP varies because it is (approximately) the product of average maturity times debt to GDP. Second, and as pointed out in Section 2, an increase in debt to GDP induces governments to issue a larger fraction of their debt long-term, hence raising average maturity.

Table 4 shows IV regressions for the yield and for the one- and three-year returns of the long-term bond. We measure supply by maturity-weighted debt to GDP, and use the ratio of marketable Treasury debt to GDP as our instrument. Marketable debt includes the bonds in the CRSP database as well as Inflation Protected Securities (TIPS), and is measured in face-value terms. Adding the square and cube of marketable debt to GDP as instruments, as in KVJ, does not affect our results. Our results are also not affected if we use only the bonds in CRSP instead of all marketable Treasury debt. If we use total debt (Bohn 2008), which includes also non-marketable debt, then our results weaken somewhat. The results for the latter two instruments are shown in Tables C.5 and C.6 in online Appendix C.

The top panel of Table 4 shows the first-stage regression of maturity-weighted debt to GDP on marketable debt to GDP, controlling for the short rate. The R^2 is 83.5%, confirming that much of the variation in maturity-weighted debt to GDP is driven by debt to GDP. The bottom panel shows the IV regressions. The t-statistics are computed using Newey and West (1987) standard errors and allowing for 36 months of lags for the yield and the one-year return, and 54 months for the three-year return. The effect of supply on the yield is not statistically significant. The effect on the one- and three-year returns is, however, and the coefficients are almost identical to their OLS counterparts in

---

8　Guibaud, Nosbusch, and Vayanos (2013) study issuance policy in the presence of investor clienteles and demand shocks. They show that a welfare-maximizing government tailors the maturity structure of its debt to the clientele mix, e.g., issues a larger fraction of its debt long-term when the fraction of long-horizon investors increases. A supply response to demand shocks could also be generated by the private sector. Koijen, Van Hemert, and Van Nieuwerburgh (2009) show that households are more likely to take fixed-rate mortgages (effectively issuing long-term bonds) when long-term bonds are expensive relative to short-term bonds. Greenwood, Hanson, and Stein (2010) show that private firms issue a larger fraction of their debt long-term when the supply of long-term bonds by the government is low.

Table 4
Instrumental-variables tests

First-stage regressions: $X_t = (MWD/GDP)_t = \alpha + \beta(D/GDP)_t + \delta y_t^{(1)} + u_t$

	Instrument = D/GDP
D/GDP	14.186
	[12.33]
$y^{(1)}$	12.809
	[4.08]
R^2	0.835

Second-stage regressions: $y_t^{(LT)} = a + b\hat{X}_t + y_t^{(1)} + u_t$, and

$$r_{t+k,k}^{(LT)} = a + b\hat{X}_t + y_t^{(1)} + u_{t+k}$$

	Yield	1-year return	3-year return
MWD/GDP (IV)	0.002	0.028	0.067
	[1.129]	[2.885]	[3.493]
$y^{(1)}$	0.845	1.223	3.748
	[11.600]	[2.868]	[4.923]
R^2	0.849	0.189	0.506

We instrument for the maturity-weighted-debt-to-GDP ratio by the Treasury-debt-to-GDP ratio. Both the first-and second-stage regressions are monthly and include a control for the one-year yield. The dependent variable in the second-stage regression is the yield or the one- or three-year return of the long-term bond. *t*-statistics, reported in parentheses, are based on Newey-West standard errors, with 36 lags in the case of the yield and one-year return regressions, and 54 lags in the case of the three-year return regression.

Table 2 (0.028 for IV and 0.026 for OLS in the case of the one-year return; 0.067 for IV and 0.065 for OLS in the case of the three-year return).

3.3 Robustness tests

In addition to the IV tests shown in Table 4, we perform a number of other robustness tests. Some of the tests concern our main sample and others our second sample, 1916–1940. Table 5 reports tests on our main sample for the one- and three-year returns of the long-term bond. We first compute our measures of supply in market-value rather than face-value terms. The market-value counterpart of maturity-weighted debt to GDP is

$$\left(\frac{MWD}{GDP}\right)_{MV,t} = \frac{\sum_{0<\tau\leq30} e^{-\tau y_t^{(\tau)}} D_t^{(\tau)} \tau}{GDP_t},$$

where $y_t^{(\tau)}$ is the yield of a zero-coupon bond with maturity τ at date t. Because there is not enough information to compute an accurate term structure of zero-coupon yields for the entire 1952–2007 sample period, we use the approximation

$$\left(\frac{MWD}{GDP}\right)_{MV,t} \approx \frac{\sum_i MV_{i,t} Dur_{i,t}}{GDP_t}, \tag{21}$$

where the summation is over all bonds in the CRSP database, $MV_{i,t}$ is the market value of bond i at date t, and $Dur_{i,t}$ is the bond's Macaulay duration. We can compute Macaulay duration because CRSP reports yield to maturity for each bond. The approximation (21) is exact when the term structure is flat.

Table 5
Robustness tests

		1-year return			3-year return		
		b	[t]	R^2	b	[t]	R^2
(1)	Base case: control for short rate only	0.026	[3.097]	0.190	0.065	[4.200]	0.506
	Market-value-based measures of supply:						
(2)	MWD/GDP MV (OLS)	0.003	[1.748]	0.136	0.011	[2.269]	0.426
(3)	MWD/GDP MV (IV)	0.007	[2.671]	0.114	0.016	[3.110]	0.409
(4)	LTD/GDP MV (OLS)	0.894	[2.586]	0.155	2.887	[3.728]	0.489
(5)	LTD/GDP MV (IV)	1.395	[2.837]	0.142	3.317	[3.448]	0.486
	Macroeconomic and other controls:						
(6)	Add control for term spread	0.014	[1.892]	0.281	0.046	[2.422]	0.582
(7)	Add control for output gap	0.026	[3.399]	0.207	0.065	[4.308]	0.536
(8)	Add control for output growth	0.026	[3.165]	0.188	0.066	[4.376]	0.522
(9)	Add control for inflation	0.020	[2.680]	0.210	0.061	[4.511]	0.509
(10)	Add control for inflation risk	0.027	[3.124]	0.191	0.076	[4.809]	0.588
(11)	Add control for interest-rate risk	0.026	[3.126]	0.189	0.068	[4.358]	0.519
(12)	Add control for stock-market risk	0.026	[3.101]	0.194	0.064	[4.324]	0.538
(13)	Time trend control	0.020	[2.343]	0.215	0.045	[2.892]	0.594
(14)	Future changes in debt maturity	0.027	[3.198]	0.202	0.067	[4.402]	0.524
(15)	Adjust MWD for Fed holdings	0.027	[3.120]	0.199	0.660	[4.242]	0.516
(16)	Annual sampling (September)	0.028	[3.085]	0.183	0.072	[4.106]	0.505
(17)	Annual sampling (December)	0.023	[2.524]	0.139	0.060	[3.538]	0.456

OLS and instrumental-variables monthly time-series regressions of the form:

$$r_{t+k,k}^{(LT)} = a + b(MWD/GDP)_t + cy_t^{(1)} + dZ_t + u_{t+k}$$

where Z denotes a control that includes the term spread, output gap, output growth, inflation, inflation risk, interest-rate risk and stock-market risk. The dependent variable is the one- or three-year return of the long-term bond. t-statistics, reported in parentheses, are based on Newey-West standard errors, with 36 lags in the case of the one-year return regression, and 54 lags in the case of three-year return regression.

We compute a market-value counterpart of long-term debt to GDP by adding the market value of all bonds whose remaining maturity exceeds ten years and scaling by GDP.

Rows (2)–(5) of Table 5 show OLS and IV regressions for our market-value-based measures of supply. For the IV regressions we use the same instrument and methodology as in Table 4. The coefficients of our supply measures are larger in the IV regression than in the OLS regression, especially for one-year returns (0.007 for IV and 0.003 for OLS in the case of maturity-weighted debt to GDP; 1.395 for IV and 0.894 for OLS in the case of long-term debt to GDP). This is consistent with the endogeneity problem mentioned in Section 2: measuring supply in market-value terms tends to induce a mechanical negative relationship between supply and yields or returns, biasing the OLS estimates downward. The IV coefficients are statistically significant, with t-statistics similar to those derived when measuring supply in face-value terms (Table 4). The OLS coefficients have smaller t-statistics than their face-value counterparts (Table 2), but they remain statistically significant except in one case.

We next add to our regressions a number of macroeconomic controls. Macroeconomic variables, such as output growth and inflation, have been shown to forecast bond returns (e.g., Ferson and Harvey 1991; Baker,

Greenwood, and Wurgler 2003; Ludvigson and Ng 2009). If the same variables affect maturity structure and debt to GDP, then our supply measures could forecast bond returns regardless of any causal effect of supply.

Rows (6)–(12) report the results of including macroeconomic controls. Supply in these rows and the rest of Table 5 is measured by maturity-weighted debt to GDP computed in face-value terms. Our first control is the term spread, defined as the yield spread between the long-term and the one-year bond. According to our model and the evidence in Table 2, the term spread is affected by supply and can hence subsume some of the supply effect. Yet, Table 5 shows that supply remains significant when controlling for the term spread. Our remaining controls are output gap, output growth, inflation, inflation risk, volatility in short-term interest rates, and stock market volatility. (Details of how each of these controls is computed are in Table 1.) Table 5 shows that the positive relationship between supply and future returns remains significant when including any of these controls. This is not entirely surprising: macroeconomic risk premia mainly vary at business-cycle frequency, but supply captures a lower-frequency component of expected returns.

Rows (13)–(17) show that our results remain significant after the following controls and adjustments. Row (13) controls for a time trend. Row (14) controls for future changes in debt supply by adding the variable $(MWD/GDP)_{t+k} -$ $(MWD/GDP)_t$ to the regression. Since debt supply affects yields, as implied by our model and the evidence in Table 2, changes in supply can explain part of returns. Controlling for them reduces estimation noise, and indeed our results strengthen somewhat. Row (15) nets out Federal Reserve holdings to derive a better proxy for the supply of bonds available to arbitrageurs. We compute Fed holdings using data from Banking and Monetary Statistics between 1952 and 1970 and from issues of the *Federal Reserve Bulletin* after 1970. These sources report holdings by maturity buckets rather than for individual bonds. We construct a measure of maturity-weighted debt held by the Fed by assigning the average maturity of each bucket to all bonds in that bucket. We then subtract that measure from our main measure of maturity-weighted debt. Rows (16) and (17) sample the data annually in September and December, respectively. For these annual regressions, the Newey-West standard errors are based on three years of lags in the case of one-year returns, and five years of lags in the case of three-year returns.

An important robustness test is whether our results hold in another time period. We consider the period 1916–1940, and omit the period between 1941 and the beginning of our main sample in 1952 for reasons explained in Section 2. Panel B of Figure 4 plots our main result for the 1916–1940 sample. Maturity-weighted debt to GDP (horizontal axis) is positively correlated with the subsequent three-year excess return of the long-term bond. Table 6 shows our yield and return regressions for the long-term bond. Yields are positively correlated with supply, but the correlation is not statistically significant. Future

Table 6
Bond supply, bond yields, and bond returns, 1916–1940

	Yield	Returns	
		1-year return	3-year return
MWD/GDP	0.001	0.008	0.029
	[1.310]	[0.988]	[5.528]
$y^{(1)}$	0.379	0.035	1.169
	[5.216]	[0.103]	[2.778]
R^2	0.698	0.043	0.448

Annual time-series regressions of the form:

$$y_t^{(LT)} = a + b MWD/GDP_t + cy_t^{(1)} + u_t$$

$$r_{t+k,k}^{(LT)} = a + b MWD/GDP_t + cy_t^{(1)} + u_{t+k}$$

during the 1916–1940 period (with one-year returns extending to 1941 and three-year returns extending to 1943). The data are sampled at the end of June. The dependent variable is the yield or the one- or three-year return of the long-term bond. *t*-statistics, reported in parentheses, are based on Newey-West standard errors, with three lags in the case of the yield and one-year return regressions, and five lags in the case of the three-year return regression.

returns are also positively correlated with supply. This correlation is not statistically significant for one-year returns but becomes highly significant for three-year returns. The latter correlation remains significant even after excluding the years 1916 and 1917, which appear in Figure 4 to play a large role in driving the correlation.

3.4 Arbitrageur wealth and bond returns

When arbitrageurs become more risk averse, they demand higher compensation to accommodate changes in bond supply, and bond supply has a stronger effect on expected returns. In this section we explore time-series implications of this idea under the assumption that risk aversion increases following losses. Based on Hypothesis 4, we construct two measures of the change in arbitrageur wealth during the year that precedes date t:

$$\Delta W_{1t}^{Arb} = (y_{t-1}^{(\tau)} - y_{t-1}^{(1)}) \cdot rx_t^{(\tau)}, \tag{22}$$

$$\Delta W_{2t}^{Arb} = \left(\frac{MWD}{GDP}\right)_{t-1} \cdot rx_t^{(\tau)}, \tag{23}$$

where $rx_t^{(\tau)} \equiv r_t^{(\tau)} - y_{t-1}^{(1)}$ denotes the excess return of the bond with maturity τ during that year. The first measure is the product of the yield spread between maturities τ and one at date $t-1$ times excess bond returns during the following year. This measure identifies arbitrageurs' past positions based on the slope of the term structure. The second measure is the product of maturity-weighted debt to GDP at date $t-1$ times excess bond returns during the following year. This measure identifies arbitrageurs' past positions based on the supply of

government debt.[9] We assume that the yield spread and excess returns in the definition of the measures concern the long-term bond.

We examine whether our measures of arbitrageur wealth influence the relationship between supply and future returns, documented in Section 3.1, and between the slope of the term structure and future returns. The latter relationship has been documented by Fama and Bliss (1987), and arises naturally in our model. Indeed, in periods when supply is high, expected bond returns are high (Proposition 2), and the term structure slopes up in the sense that intermediate- and long-term bonds have high yields relative to short-term bonds (Proposition 1). As with supply, slope has predictive power only when arbitrageurs are risk averse, so we would expect more predictive power the higher risk aversion is.

Table 7 reports results from regressing the return of the long-term bond, over both a one- and a three-year horizon, on (i) maturity-weighted debt to GDP and its interaction with either of our two measures of arbitrageur wealth, or on (ii) the long-term yield spread and its interaction with either of our two measures. In all cases we control for the short rate, so the regression equation is

$$r_{t+k,k}^{(LT)} = a + b\left(\frac{MWD}{GDP}\right)_t + c(y_t^{(LT)} - y_t^{(1)}) + d\,\Delta W_t^{Arb}\left(\frac{MWD}{GDP}\right)_t$$
$$+ e\,\Delta W_t^{Arb}(y_t^{(LT)} - y_t^{(1)}) + f\,y_t^{(1)} + u_{t+k}.$$

According to Hypothesis 4, the interaction terms should have a negative coefficient: supply and slope predict returns positively, and more so when arbitrageur wealth decreases. The results confirm this prediction in the case of one-year returns. Indeed, the interaction terms between supply and our two measures of arbitrageur wealth have a negative and statistically significant coefficient, and the same is true for the interaction terms between slope and arbitrageur wealth. In the case of three-year returns, the interaction terms have a negative coefficient, but only one out of the four is statistically significant. Controlling for past returns does not affect the statistical significance of the interaction terms.

The coefficients of the interaction terms are economically significant. Consider, for example, the interaction term between supply and our first measure of arbitrageur wealth, (22). This term has a coefficient of –4.436 for one-year returns. From Table 1, the standard deviation of the wealth measure is

---

[9] An alternative version of measure (23), which is more in the spirit of Hypothesis 4, is

$$\Delta W_{2t}^{Arb} = \left(\left(\frac{MWD}{GDP}\right)_{t-1} - \overline{\left(\frac{MWD}{GDP}\right)}\right) \cdot rx_t^{(\tau)}, \tag{24}$$

where $\overline{\left(\frac{MWD}{GDP}\right)}$ denotes the time-series average of maturity-weighted debt to GDP. Under both (24) and Hypothesis 4, arbitrageurs are short bonds when the dollar duration of government bond supply is low. Under (23) instead, arbitrageurs hold a small long position. The choice of (23) or (24) does not matter for our results.

Table 7
Bond supply, bond returns, and arbitrageur wealth

	1-year return						3-year return					
MWD/GDP	0.026	0.031	0.034				0.065	0.067	0.069			
	[3.102]	[3.408]	[3.634]				[4.210]	[4.252]	[4.268]			
$y^{(LT)} - y^{(1)}$				3.443	4.26	4.114				6.46	7.076	6.819
				[5.136]	[5.190]	[5.633]				[4.710]	[4.611]	[4.375]
$\Delta W_1^{Arb} \cdot MWD/GDP$		−4.729						−1.971				
		[−3.386]						[−1.221]				
$\Delta W_2^{Arb} \cdot MWD/GDP$			−0.029						−0.013			
			[−3.114]						[−1.422]			
$\Delta W_1^{Arb} \cdot (y^{(LT)} - y^{(1)})$					−675.811						−509.326	
					[−2.774]						[−2.404]	
$\Delta W_2^{Arb} \cdot (y^{(LT)} - y^{(1)})$						−5.226						−2.792
						[−2.954]						[−1.153]
$y^{(1)}$	1.214	1.146	1.130	1.894	1.838	1.989	3.741	3.713	3.703	4.945	4.903	4.995
	[2.791]	[2.587]	[2.489]	[5.405]	[5.406]	[5.918]	[4.830]	[4.806]	[4.670]	[6.994]	[7.059]	[6.862]
R^2	0.190	0.239	0.238	0.265	0.309	0.307	0.503	0.505	0.506	0.518	0.526	0.522

Monthly time-series regressions of the form:

$$r_{t+k,k}^{(LT)} = a + b\,MWD/GDP + c(y_t^{(LT)} - y_t^{(1)}) + e\Delta W_t^{Arb} MWD/GDP_t + f\Delta W_t^{Arb}(y_t^{(LT)} - y_t^{(1)}) + g y_t^{(1)} + u_{t+k}$$

The dependent variable is the one-or three-year return of the long-term bond. The independent variables include the spread between the long-term and the one-year yield, MWD/GDP, and interactions between these variables and changes in arbitrageur wealth. We use two measures of arbitrageur wealth. The first measure, ΔW_1^{Arb}, is the product of the one-year lagged spread between the long-term and the one-year yield, times the subsequent one-year excess return of the long-term bond. The second measure, ΔW_2^{Arb}, is the product of the one-year lagged MWD/GDP, times the subsequent one-year excess return of the long-term bond. *t*-statistics, reported in parentheses, are based on Newey-West standard errors, with 36 lags in the case of the one-year return regression, and 54 lags in the case of three-year return regression.

0.0015. Therefore, a one-standard-deviation movement in the wealth measure changes the coefficient of maturity-weighted debt to GDP by $4.436 \times 0.0015 = 0.0067$. This is approximately a one-quarter percentage change, since the coefficient is 0.026 (Tables 2 and 7).

Table 7 assumes that arbitrageur risk aversion is influenced by trading performance over a one-year horizon. The relevant horizon might be different, however, and is influenced by the speed at which fresh capital can enter in the arbitrage industry. If entry is fast, then capital losses over the distant past do not affect current capital or risk aversion because the lost capital is replenished quickly. Our analysis provides an estimate of the speed of entry, which might be relevant for theories of slow-moving capital.[10] In Table C.7 in online Appendix C we measure the change in arbitrageur wealth by

$$\Delta W_{1t}^{Arb} = \sum_{j=1}^{k} (y_{t-j}^{(\tau)} - y_{t-j}^{(1)}) \cdot rx_{t-j+1}^{(\tau)},$$

$$\Delta W_{2t}^{Arb} = \sum_{j=1}^{k} \left(\frac{MWD}{GDP}\right)_{t-j} \cdot rx_{t-j+1}^{(\tau)},$$

that is, the sum of wealth changes over the past k years, where the change in wealth over any given year is measured as in the baseline case (Equations (22) and (23)). The interaction term is economically and statistically significant at horizons of one and two years. These findings suggest that capital losses in term-structure arbitrage take two to three years to be offset by inflows of fresh capital.

4. Calibration

In this section we calibrate our model to the data. We estimate parameters for the supply-factor process (5) and the short-rate process (6). We also estimate the sensitivity of supply at each maturity to changes in the supply factor. These parameters, together with the arbitrageur risk-aversion coefficient a, fully determine the effects of supply on yields and expected returns within the model. Since these effects are an increasing function of a, there exists a unique value of a that equates the average effect of supply on yields and expected returns in the model and in the data. We compute this value and compare it with estimates of risk aversion used in the literature. We also examine whether the value of a that matches the average effect of supply can also match relative effects—for example, on yields relative to expected returns and on long- relative to short-term bonds.

---

[10] See, for example, Duffie's (2010) presidential address to the American Finance Association for a model of slow-moving capital and a survey of the theoretical and empirical work in that area. See also Gromb and Vayanos (2010) for a survey of the theoretical literature on the limits of arbitrage.

To estimate parameters $(\kappa_\beta, \kappa_r, \gamma, \sigma_\beta, \sigma_r, \sigma_{r\beta})$ for the supply-factor process (5) and the short-rate process (6), we discretize these processes and perform a vector auto-regression (VAR) on monthly data in the 1952–2007 sample. We use maturity-weighted debt to GDP as our proxy for the supply factor β_t, and the one-year yield $y_t^{(1)}$ as our proxy for the short rate r_t.[11] The details of the vector auto-regressions and of the remaining steps in our calibration are in Appendix B. Our estimate of the mean-reversion parameter κ_r for the short rate is larger than its counterpart κ_β for the supply factor: the short rate mean-reverts at business-cycle frequency, while movements in maturity-weighted debt to GDP occur at a lower frequency, as shown in Figure 2. Our estimate of γ is positive, meaning that a shock to the supply factor moves expected future short rates, holding the current short rate constant, and this movement is in the direction opposite to the shock. Our estimate of $\sigma_{r\beta}$ is positive, meaning that shocks to the supply factor and to the short rate are positively correlated. The correlation is small, however.

To estimate the sensitivity of supply at any given maturity τ to changes in the supply factor, we regress supply at that maturity scaled by GDP on maturity-weighted debt to GDP. This yields an estimate for the function $\theta(\tau)$, which we plot in Figure C.2 in online Appendix C. The function $\theta(\tau)$ is positive, meaning that an increase in the supply factor raises supply at each maturity.

Given our estimates for $(\kappa_\beta, \kappa_r, \gamma, \sigma_\beta, \sigma_r, \sigma_{r\beta}, \theta(\tau))$ and a value for the arbitrageur risk-aversion coefficient a, we compute an equilibrium in our model using Theorem B.2 in Appendix B. We then compute the average coefficient of supply across the 28 yield and return regressions reported in Table 3, which concern zero-coupon bonds with maturities from two to fifteen years. The value of a that renders this average coefficient in the model equal to that in the data is $a = 57$. Using a different set of regressions has a small effect on a: for example, using only the fourteen yield regressions we find $a = 42$, and using only the fourteen return regressions we find $a = 64$.

We can compute a standard error for our estimates of a using Monte-Carlo simulation. We generate artificial samples by simulating the VAR equations with the parameters computed from our actual sample and reported in Table B1 in Appendix B. For each artificial sample, we re-estimate the VAR, and compute a new value for a using the procedure described in this section. With 10,000 samples, the standard error for the estimate $a = 57$ is 13.7, and the [5%, 95%] confidence interval is [37.1, 82.6].

---

[11] An alternative proxy for β_t could be constructed based on the total supply of debt, i.e., including corporate and mortgage-backed debt in addition to government debt. This proxy captures the idea that arbitrageurs care about the total duration risk that they bear, whether it comes from government or non-government debt. If shocks to the supply of government and non-government debt are positively correlated, then our estimate of arbitrageur risk aversion a under the alternative proxy will be lower than under our original proxy. Indeed, since positive correlation implies larger shocks to total duration risk, explaining a given effect of supply requires lower arbitrageur risk aversion. The opposite will be true if the two sets of shocks are negatively correlated. Constructing the alternative proxy requires time-series data on the maturity composition of non-government debt, which are not available for much of our sample.

The coefficient *a* measures the absolute risk aversion of arbitrageurs. To convert *a* into a coefficient of relative risk aversion (CRRA), we must multiply it by arbitrageur wealth. This wealth must be expressed as a fraction of GDP since debt supply is expressed in the same manner. A natural interpretation of arbitrageurs in our model is as hedge funds and proprietary-trading desks: these agents typically have short horizons as our arbitrageurs. Hedge Fund Research reports that the capital controlled by hedge funds in 2007 was 13.3% of GDP. Assuming that this is representative of the level of arbitrage capital in our 1952–2007 sample, we estimate the CRRA to be $57 \times 13.3\% = 7.6$.

Our estimate of CRRA assumes that arbitrageurs are the only agents to absorb shocks to the supply of government debt. This is because we proxy the supply available to arbitrageurs, which we do not observe, by the supply of bonds issued by the government. Arbitrageurs, however, are likely to be absorbing only a fraction of supply shocks, with the rest being absorbed by investors such as pension funds, insurance companies, and mutual funds. Such investors typically have longer horizons than arbitrageurs. Our model treats their demand as exogenous and part of the net supply available to arbitrageurs. As a crude way to adjust for the presence of these investors, we add the capital that they control to arbitrage capital. According to the Flow of Funds tables, the capital controlled by private pension funds in 2007 was 45.7% of GDP, that by insurance companies (life and property casualty) was 45.2%, and that by mutual funds was 55.8%. Adding these to hedge-fund capital, we estimate the CRRA to be $57 \times (13.3\% + 45.7\% + 45.2\% + 55.8\%) = 91.2$.

Our second estimate of the CRRA assumes that investors can respond to supply shocks in the same manner as arbitrageurs. If, however, their response is slow or limited by constraints related to market segmentation (e.g., pension funds must keep a stable bond-stock mix), then the estimate would be smaller. The estimate would also be smaller if we include additional risk factors in the model—for example, allow for variation in investor demand. Yet, our finding that supply effects in the bond market can be consistent with CRRA values that are large relative to typical values used in the literature is worthy of further investigation. This is especially so since estimates of supply effects from recent quantitative easing programs are somewhat larger than ours (Section 3.1), hence implying even larger values of risk aversion.

We next examine whether the risk-aversion coefficient that matches the average effect of supply in the model and in the data can also match relative effects. Figure 6 plots the coefficients of supply in the 28 yield and return regressions in Table 3 as a function of maturity, and compares with the coefficients derived from the model for $a = 57$. The spread between the returns and yields coefficients is smaller in the model than in the data. These discrepancies are, however, mostly within confidence intervals. Both sets of coefficients are increasing with maturity in the model as in the data. The effect of maturity is larger in the model than in the data for the yields coefficients and smaller for the returns coefficients.
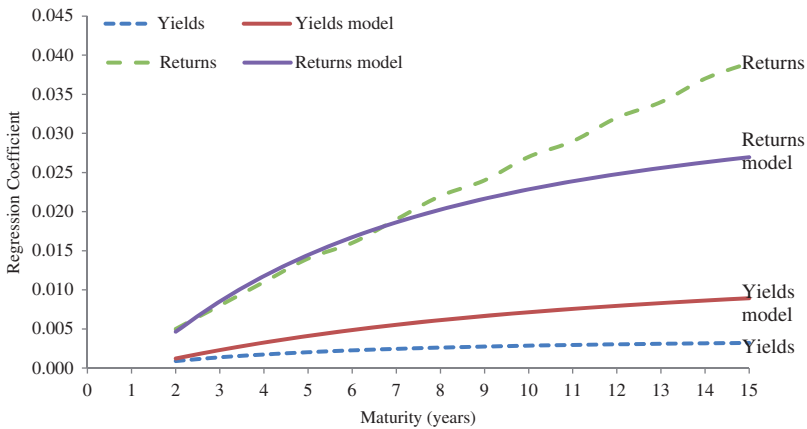
Figure 6
Model calibration
The solid and dashed lines plot the coefficients b_τ from the monthly time-series regressions:

$$y_t^{(\tau)} = a + b_\tau (MWD/GDP)_t + c y_t^{(1)} + u_t, \text{ and}$$

$$r_{t+k,k}^{(\tau)} = a + b_\tau (MWD/GDP)_t + c y_t^{(1)} + u_{t+k},$$

as a function of maturity τ ranging from two to fifteen years. In the first equation the yield of a zero-coupon bond with maturity τ is regressed on the maturity-weighted-debt-to-GDP ratio, controlling for the one-year yield. In the second equation the dependent variable is the bond's one-year return. The solid lines are the coefficients derived from the model. The dashed lines are the coefficients when the regressions are performed on actual data during the 1971–2007 period. The risk aversion coefficient of arbitrageurs in the model is chosen so that the average across all points in the two solid lines is equal to the average across all points in the two dashed lines.

5. Conclusion

The supply and maturity structure of government debt play no role in standard term-structure theories. Yet, their effects on bond yields and expected returns are the subject of numerous policy debates, ranging from debt management by treasury departments to quantitative easing by central banks. Given the importance of these debates, it is surprising how little empirical evidence there is correlating supply and maturity structure to bond yields and returns in long time series. This paper is an attempt to fill that gap.

We organize our empirical investigation around a term-structure model in which risk-averse arbitrageurs absorb shocks to the demand and supply for bonds of different maturities. The model predicts that an increase in supply should raise bond yields and expected returns, holding the short rate constant, and these effects should be stronger for longer-maturity bonds and during times when arbitrageurs are more risk averse. The model also suggests that the empirically relevant measure of supply is maturity-weighted debt, which captures the duration risk that arbitrageurs must bear. Using U.S. data, we find support for the model's predictions. In particular, an increase in our supply measure by one standard deviation, holding the one-year rate constant, raises

the yield on a long-term government bond with an approximate maturity of twenty years by 40 basis points and its expected return over a one-year horizon by 259 basis points. We use our empirical estimates of supply effects to calibrate the model and infer arbitrageur risk aversion.

Appendix

A. Proofs of Theoretical Results

Proof of Lemma 1: Using (8), we can write (2) as

$$dW_t = \left(W_t r_t + \int_0^T x_t^{(\tau)}(\mu_t^{(\tau)} - r_t)d\tau\right)dt - \left(\int_0^T x_t^{(\tau)} A_r(\tau)d\tau\right)\sigma_r dB_{r,t} - \left(\int_0^T x_t^{(\tau)} A_\beta(\tau)d\tau\right)\sigma_\beta dB_{\beta,t},$$

and (3) as

$$\max_{\{x_t^{(\tau)}\}_{\tau \in (0,T]}} \left[\int_0^T x_t^{(\tau)}(\mu_t^{(\tau)} - r_t)d\tau - \frac{a\sigma_r^2}{2}\left(\int_0^T x_t^{(\tau)} A_r(\tau)d\tau\right)^2 - \frac{a\sigma_\beta^2}{2}\left(\int_0^T x_t^{(\tau)} A_\beta(\tau)d\tau\right)^2\right].$$

$$(A.1)$$

Point-wise maximization of (A.1) yields (10). ∎

We next show two useful lemmas.

Lemma A.1. If a function $g(\tau)$ is positive and increasing, then $\int_0^T g(\tau)\theta(\tau)d\tau > 0$.

Proof. We can write the integral $\int_0^T g(\tau)\theta(\tau)d\tau$ as

$$\int_0^T g(\tau)\theta(\tau)d\tau = \int_0^{\tau^*} g(\tau)\theta(\tau)d\tau + \int_{\tau^*}^T g(\tau)\theta(\tau)d\tau$$

$$> g(\tau^*)\int_0^{\tau^*} \theta(\tau)d\tau + g(\tau^*)\int_{\tau^*}^T \theta(\tau)d\tau$$

$$= g(\tau^*)\int_0^T \theta(\tau)d\tau \geq 0,$$

where the second step follows from Part (ii) of Assumption 1 and because $g(\tau)$ is increasing, and the last step follows from Part (i) of Assumption 1 and because $g(\tau)$ is positive. ∎

Lemma A.2. The functions $A_r(\tau)$ and $A_\beta(\tau)$, given by (13) and (14), respectively, are positive and increasing. For $A_\beta(\tau)$, this holds for any value of $\hat{\kappa}_\beta$, and not only for $\hat{\kappa}_\beta$ solutions to (16).

Proof. Equation (13) implies that $A_r(\tau)$ is positive and increasing. Therefore, Lemma A.1 implies that $I_r > 0$ and $Z > 0$. To show that $A_\beta(\tau)$ is positive and increasing, we write it as

$$A_\beta(\tau) = Z \int_0^\tau \frac{1 - e^{-\kappa_r \hat{\tau}}}{\kappa_r} e^{-\hat{\kappa}_\beta(\tau - \hat{\tau})} d\hat{\tau}. \tag{A.2}$$

Since $Z > 0$, (A.2) implies that $A_\beta(\tau)$ is positive. Differentiating (A.2), we find

$$A'_\beta(\tau) = Z\left(\frac{1 - e^{-\kappa_r \tau}}{\kappa_r} - \hat{\kappa}_\beta \int_0^\tau \frac{1 - e^{-\kappa_r \hat{\tau}}}{\kappa_r} e^{-\hat{\kappa}_\beta(\tau - \hat{\tau})} d\hat{\tau}\right). \tag{A.3}$$

If $\hat{\kappa}_\beta \leq 0$, (A.3) implies that

$$A'_\beta(\tau) \geq Z\frac{1 - e^{-\kappa_r \tau}}{\kappa_r} > 0.$$

If $\hat{\kappa}_\beta > 0$, (A.3) implies that

$$A'_\beta(\tau) > Z\frac{1 - e^{-\kappa_r \tau}}{\kappa_r}\left(1 - \hat{\kappa}_\beta \int_0^\tau e^{-\hat{\kappa}_\beta(\tau - \hat{\tau})} d\hat{\tau}\right) = Z\frac{1 - e^{-\kappa_r \tau}}{\kappa_r} e^{-\hat{\kappa}_\beta \tau} > 0,$$

since $A_r(\tau)$ is increasing in τ. Therefore, in both cases, $A_\beta(\tau)$ is increasing in τ. ∎

Proof of Theorem 1: Substituting $x_t^{(\tau)}$ from (4) and (12) into (11), we find

$$\lambda_{i,t} = a\sigma_i^2 \int_0^T [\zeta(\tau) + \theta(\tau)\beta_t] A_i(\tau) d\tau. \tag{A.4}$$

Substituting $\mu_t^{(\tau)}$ and $\lambda_{i,t}$ from (9) and (A.4) into (10), we find an affine equation in (r_t, β_t). Identifying terms in r_t yields

$$\kappa_r A_r(\tau) + A_r'(\tau) - 1 = 0, \tag{A.5}$$

identifying terms in β_t yields

$$\kappa_\beta A_\beta(\tau) + A_\beta'(\tau) = a\sigma_r^2 A_r(\tau) \int_0^T A_r(\tau)\theta(\tau) d\tau + a\sigma_\beta^2 A_\beta(\tau) \int_0^T A_\beta(\tau)\theta(\tau) d\tau, \tag{A.6}$$

and identifying constant terms yields

$$C'(\tau) - \kappa_r \bar{r} A_r(\tau) + \frac{\sigma_r^2}{2} A_r(\tau)^2 + \frac{\sigma_\beta^2}{2} A_\beta(\tau)^2 = a\sigma_r^2 A_r(\tau) \int_0^T A_r(\tau)\zeta(\tau) d\tau + a\sigma_\beta^2 A_\beta(\tau) \int_0^T A_\beta(\tau)\zeta(\tau) d\tau. \tag{A.7}$$

The ordinary differential equations (ODEs) (A.5)–(A.7) must be solved with the initial condition $A_r(0) = A_\beta(0) = C(0) = 0$. The solution to (A.5) is (13). Using (13) and the definitions of I_r and Z, we can write (A.6) as

$$\hat{\kappa}_\beta A_\beta(\tau) + A_\beta'(\tau) = Z \frac{1 - e^{-\kappa_r}}{\kappa_r}, \tag{A.8}$$

where

$$\hat{\kappa}_\beta \equiv \kappa_\beta - a\sigma_\beta^2 \int_0^T A_\beta(\tau)\theta(\tau) d\tau. \tag{A.9}$$

The solution to (A.8) is (14). Substituting into (A.9), we find that $\hat{\kappa}_\beta$ is given by (16). The solution to (A.7) is

$$C(\tau) = Z_r \int_0^\tau A_r(\tau') d\tau' + Z_\beta \int_0^\tau A_\beta(\tau') d\tau' - \frac{\sigma_r^2}{2} \int_0^\tau A_r(\tau')^2 d\tau' - \frac{\sigma_\beta^2}{2} \int_0^\tau A_\beta(\tau')^2 d\tau', \tag{A.10}$$

where

$$Z_r \equiv \kappa_r \bar{r} + a\sigma_r^2 \int_0^T \frac{1 - e^{-\kappa_r}}{\kappa_r} \zeta(\tau) d\tau,$$

$$Z_\beta \equiv Z a\sigma_\beta^2 \int_0^T \frac{1}{\kappa_r} \left(\frac{1 - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta} - \frac{e^{-\kappa_r \tau} - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta - \kappa_r} \right) \zeta(\tau) d\tau.$$

To complete the proof of the theorem, we must show that (16) has a solution for a below a threshold $\bar{a} > 0$. Since the function $A_\beta(\tau)$ is positive and increasing from Lemma A.2, Lemma A.1 implies that the second term in the right-hand side of (16) is positive. Therefore, any solution to (16) must satisfy $\hat{\kappa}_\beta < \kappa_\beta$. When $\hat{\kappa}_\beta = \kappa_\beta$, the left-hand side of (16) is equal to κ_β and the right-hand side is smaller than κ_β. When instead $\hat{\kappa}_\beta$ goes to $-\infty$, both the left- and the right-hand side go to $-\infty$, but the right-hand side converges at the rate $-\frac{e^{-\hat{\kappa}_\beta T}}{\hat{\kappa}_\beta^2}$, which is faster than the rate $\hat{\kappa}_\beta$ at which the left-hand side converges. Therefore, (16) has an even number of solutions, possibly zero. A sufficient condition for (16) to have a solution is that the left-hand side is smaller than the right-hand side for $\hat{\kappa}_\beta = 0$, that is,

$$\kappa_\beta - a^2 \sigma_r^2 \sigma_\beta^2 I_r \int_0^T \frac{1}{\kappa_r} \left(\tau - \frac{1 - e^{-\kappa_r \tau}}{\kappa_r} \right) \theta(\tau) d\tau > 0$$

$$\Leftrightarrow a < \sqrt{\frac{\kappa_\beta}{\sigma_r^2 \sigma_\beta^2 I_r \int_0^T \frac{1}{\kappa_r} \left(\tau - \frac{1 - e^{-\kappa_r \tau}}{\kappa_r} \right) \theta(\tau) d\tau}}. \tag{A.11}$$

Thus, (16) has a solution when a is smaller than the threshold $\bar{a} > 0$ defined as the right-hand side of (A.11). ∎

Proof of Proposition 1: Since the yield of the bond with maturity τ at time t is

$$y_t^{(\tau)} \equiv -\frac{\log\left(P_t^{(\tau)}\right)}{\tau} = \frac{A_r(\tau)r_t + A_\beta(\tau)\beta_t + C(\tau)}{\tau}, \tag{A.12}$$

the effect of a shock to β_t is

$$\frac{\partial y_t^{(\tau)}}{\partial \beta_t} = \frac{A_\beta(\tau)}{\tau}. \tag{A.13}$$

Since the function $A_\beta(\tau)$ is positive, the effect in (A.13) is positive. To determine how the effect depends on maturity, we differentiate with respect to τ:

$$\frac{d\left(\frac{A_\beta(\tau)}{\tau}\right)}{d\tau} = \frac{\tau A'_\beta(\tau) - A_\beta(\tau)}{\tau^2}.$$

The function $\tau \to \tau A'_\beta(\tau) - A_\beta(\tau)$ is zero for $\tau = 0$. Its derivative is $\tau A''_\beta(\tau)$. Differentiating (A.3), we find

$$A''_\beta(\tau) = Z\left(e^{-\kappa_r\tau} - \hat{\kappa}_\beta\frac{1-e^{-\kappa_r\tau}}{\kappa_r} + \hat{\kappa}_\beta^2\int_0^\tau \frac{1-e^{-\kappa_r\hat{t}}}{\kappa_r}e^{-\hat{\kappa}_\beta(\tau-\hat{t})}d\hat{t}\right).$$

The function $A''_\beta(\tau)$ has the same sign as

$$F(\tau) \equiv e^{(\hat{\kappa}_\beta-\kappa_r)\tau} - \hat{\kappa}_\beta\frac{e^{\hat{\kappa}_\beta\tau} - e^{(\hat{\kappa}_\beta-\kappa_r)\tau}}{\kappa_r} + \hat{\kappa}_\beta^2\int_0^\tau \frac{1-e^{-\kappa_r\hat{t}}}{\kappa_r}e^{\hat{\kappa}_\beta\hat{t}}d\hat{t}.$$

Since $F(0) > 0$ and

$$F'(\tau) = (\hat{\kappa}_\beta - \kappa_r)e^{(\hat{\kappa}_\beta-\kappa_r)\tau} - \hat{\kappa}_\beta\frac{\hat{\kappa}_\beta e^{\hat{\kappa}_\beta\tau} - (\hat{\kappa}_\beta-\kappa_r)e^{(\hat{\kappa}_\beta-\kappa_r)\tau}}{\kappa_r} + \hat{\kappa}_\beta^2\frac{1-e^{-\kappa_r\tau}}{\kappa_r}e^{\hat{\kappa}_\beta\tau}$$

$$= -\kappa_r e^{(\hat{\kappa}_\beta-\kappa_r)\tau} < 0,$$

the function $F(\tau)$ is either positive or positive and then negative, and the same is true for $A''_\beta(\tau)$. Therefore, the function $\tau A'_\beta(\tau) - A_\beta(\tau)$ is either increasing, or increasing and then decreasing. Since it is zero for $\tau = 0$, it is either positive, or positive and then negative. Hence, the function $A_\beta(\tau)/\tau$ is either increasing, or increasing and then decreasing, which means that the effect of a shock to β_t on yields is either increasing or hump-shaped across maturities. ∎

Proof of Proposition 2: Equations (10) and (A.4) imply that the effect of a shock to β_t on the instantaneous expected return of the bond with maturity τ at time t is

$$\frac{\partial \mu_t^{(\tau)}}{\partial \beta_t} = a\sigma_r^2 A_r(\tau)\int_0^T A_r(\tau)\theta(\tau)d\tau + a\sigma_r^2 A_\beta(\tau)\int_0^T A_\beta(\tau)\theta(\tau)d\tau. \tag{A.14}$$

Since the functions $A_r(\tau)$ and $A_\beta(\tau)$ are positive and increasing from Lemma A.2, Lemma A.1 implies that the two integrals in (A.14) are positive. This property, together with $A_r(\tau)$ and $A_\beta(\tau)$ being positive and increasing, imply that the effect in (A.14) is positive and increasing across maturities. ∎

Proof of Proposition 3: Equations (A.6) and (A.14) imply that

$$A_\beta(\tau) = \int_0^\tau \frac{\partial \mu_t^{(\hat{t})}}{\partial \beta_t}e^{-\kappa_\beta(\tau-\hat{t})}d\hat{t} < \frac{\partial \mu_t^{(\tau)}}{\partial \beta_t}\frac{1-e^{-\kappa_\beta\tau}}{\kappa_\beta}, \tag{A.15}$$

where the second step follows because Proposition 2 implies that $\partial \mu_t^{(\tau)}/\partial \beta_t$ is increasing in τ. Combining (A.13) and (A.15), we find

$$\frac{\partial y_t^{(\tau)}}{\partial \beta_t} < \frac{\partial \mu_t^{(\tau)}}{\partial \beta_t}\frac{1-e^{-\kappa_\beta\tau}}{\kappa_\beta\tau} < \frac{\partial \mu_t^{(\tau)}}{\partial \beta_t},$$

where the second step follows because $\kappa_\beta > 0$. Hence, the effect of a shock to β_t on yields is smaller than on instantaneous expected returns. ∎

Proof of Proposition 4: Since the right-hand side of (16) is decreasing in a, the largest solution for $\hat{\kappa}_\beta$ is decreasing in a. Equation (A.2) then implies that the function $A_\beta(\tau)$ is increasing in a, and (13) and (A.14) imply that the effect of β_t on instantaneous expected returns is increasing in a. ∎

B. Extensions and Calibration

B.1 Hedging demand

In this section we modify arbitrageurs' preferences to introduce a hedging demand. We replace the optimization problem (3) by

$$\max_{\{x_t^{(\tau)}\}_{\tau \in (0,T]}} \left[E_t(dW_t) - \frac{a}{2} [Var_t(dW_t) + \Psi Cov_t(dW_t, dR_t)] \right], \tag{B.1}$$

where dR_t is a portfolio return with loadings Λ_r and Λ_β, respectively, on the shocks $dB_{r,t}$ and $dB_{\beta,t}$.

The covariance term can be given multiple interpretations. For example, arbitrageurs could be asset managers tracking a benchmark portfolio. Their optimization problem would then be

$$\max_{\{x_t^{(\tau)}\}_{\tau \in (0,T]}} \left[E_t(dW_t) - \frac{a}{2} Var_t(dW_t - W_t dR_t) \right], \tag{B.2}$$

where R_t is the return on the benchmark portfolio. Alternatively, arbitrageurs could be pension-fund managers hedging a fixed-term liability. Their optimization problem would then be (B.2), where W_t is replaced by the market value L_t of the liability and R_t is the return on L_t. These optimization problems are equivalent to (B.1), provided that $\Psi_t = -2W_t$ in the first case and $\Psi = -2L_t$ in the second case. If W_t and L_t are constant over time because managers form overlapping generations that start with the same wealth and liabilities, then Ψ is constant. Moreover, if dR_t is a portfolio of zero-coupon bonds with weights that are constant over time, then Λ_r and Λ_β are constant.

Introducing a hedging demand affects only the function $C(\tau)$ in the equilibrium derived in Section 1, but not the functions $A_r(\tau)$ and $A_\beta(\tau)$. Therefore, Propositions 1–4 continue to hold.

Theorem B.1. The functions $A_r(\tau)$ and $A_\beta(\tau)$ are given by (13) and (14), respectively. The function $C(\tau)$ is given by (A.10), with Z_r and Z_β given by (B.9) and (B.10), respectively.

Proof of Theorem B.1: With a hedging demand, the arbitrageurs' optimization problem (A.1) is replaced by

$$\max_{\{x_t^{(\tau)}\}_{\tau \in (0,T]}} \left[\int_0^T x_t^{(\tau)}(\mu_t^{(\tau)} - r_t)d\tau - \frac{a\sigma_r^2}{2}\left(\int_0^T x_t^{(\tau)} A_r(\tau)d\tau\right)^2 - \frac{a\sigma_\beta^2}{2}\left(\int_0^T x_t^{(\tau)} A_\beta(\tau)d\tau\right)^2 \right.$$

$$\left. - \frac{a\Psi\sigma_r\Lambda_r}{2}\int_0^T x_t^{(\tau)} A_r(\tau)d\tau - \frac{a\Psi\sigma_\beta\Lambda_\beta}{2}\int_0^T x_t^{(\tau)} A_\beta(\tau)d\tau \right]. \tag{B.3}$$

Point-wise maximization of (B.3) yields the first-order condition (10), with (11) replaced by

$$\lambda_{r,t} \equiv a\sigma_r^2 \int_0^T x_t^{(\tau)} A_r(\tau)d\tau + \frac{a\Psi\sigma_r\Lambda_r}{2}, \tag{B.4}$$

$$\lambda_{\beta,t} \equiv a\sigma_\beta^2 \int_0^T x_t^{(\tau)} A_\beta(\tau)d\tau + \frac{a\Psi\sigma_\beta\Lambda_\beta}{2}. \tag{B.5}$$

Substituting $x_t^{(\tau)}$ from (4) and (12) into (B.15) and (B.16), we find

$$\lambda_{r,t} \equiv a\sigma_r^2 \int_0^T [\zeta(\tau) + \theta(\tau)\beta_t] A_r(\tau)d\tau + \frac{a\Psi\sigma_r\Lambda_r}{2}, \tag{B.6}$$

$$\lambda_{\beta,t} \equiv a\sigma_\beta^2 \int_0^T [\zeta(\tau) + \theta(\tau)\beta_t] A_\beta(\tau)d\tau + \frac{a\Psi\sigma_\beta\Lambda_\beta}{2}. \tag{B.7}$$

Substituting $\mu_t^{(\tau)}$, $\lambda_{r,t}$, and $\lambda_{\beta,t}$ from (9), (B.6), and (B.7) into (10), we find an affine equation in (r_t, β_t). Identifying terms in r_t and β_t yields (A.5) and (A.6), respectively. Identifying constant terms yields

$$C'(\tau) - \kappa_r \bar{r} A_r(\tau) + \frac{\sigma_r^2}{2} A_r(\tau)^2 + \frac{\sigma_\beta^2}{2} A_\beta(\tau)^2$$

$$= a\sigma_r^2 A_r(\tau) \int_0^T A_r(\tau)\zeta(\tau)d\tau + a\sigma_\beta^2 A_\beta(\tau) \int_0^T A_\beta(\tau)\zeta(\tau)d\tau + \frac{a\Psi\sigma_r\Lambda_r}{2} A_r(\tau) + \frac{a\Psi\sigma_\beta\Lambda_\beta}{2} A_\beta(\tau). \tag{B.8}$$

The solutions to (A.5) and (A.6) are (13) and (14), respectively. The solution to (B.8) is (A.10) with

$$Z_r \equiv \kappa_r \bar{r} + a\sigma_r^2 \int_0^T \frac{1 - e^{-\kappa_r}}{\kappa_r} \zeta(\tau)d\tau + \frac{a\Psi\sigma_r\Lambda_r}{2}, \tag{B.9}$$

$$Z_\beta \equiv Z a\sigma_\beta^2 \int_0^T \frac{1}{\kappa_r} \left(\frac{1 - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta} - \frac{e^{-\kappa_r \tau} - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta - \kappa_r} \right) \zeta(\tau)d\tau + \frac{a\Psi\sigma_\beta\Lambda_\beta}{2}. \tag{B.10}$$

Hence, the functions $A_r(\tau)$, $A_\beta(\tau)$ and $C(\tau)$ are as in the theorem's statement. ∎

B.2 Correlated short rate and supply factor
In this section we study the case where the short rate r_t and the supply factor β_t are correlated. Correlation affects the functions $A_\beta(\tau)$ and $C(\tau)$ in the equilibrium derived in Section 1, but not the function $A_r(\tau)$.

Theorem B.2. The functions $A_r(\tau)$ and $A_\beta(\tau)$ are given by (13) and (14), respectively, where

$$Z \equiv \frac{a(\sigma_r^2 + \sigma_{r\beta}^2)I_r - \kappa_r\gamma}{1 - a\sigma_{r\beta}\sigma_\beta \int_0^T \frac{1}{\kappa_r} \left(\frac{1 - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta} - \frac{e^{-\kappa_r \tau} - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta - \kappa_r} \right) \theta(\tau)d\tau},$$

I_r is given by (15), and $\hat{\kappa}_\beta$ solves

$$\hat{\kappa}_\beta = \kappa_\beta - a\sigma_{r\beta}\sigma_\beta I_r - \frac{\left(a(\sigma_r^2 + \sigma_{r\beta}^2)I_r - \kappa_r\gamma \right) a\sigma_\beta^2 \int_0^T \frac{1}{\kappa_r} \left(\frac{1 - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta} - \frac{e^{-\kappa_r \tau} - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta - \kappa_r} \right) \theta(\tau)d\tau}{1 - a\sigma_{r\beta}\sigma_\beta \int_0^T \frac{1}{\kappa_r} \left(\frac{1 - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta} - \frac{e^{-\kappa_r \tau} - e^{-\hat{\kappa}_\beta \tau}}{\hat{\kappa}_\beta - \kappa_r} \right) \theta(\tau)d\tau}. \tag{B.11}$$

Eq. (B.11) has a solution if a is below a threshold $\bar{a} > 0$. The function $C(\tau)$ is given by (B.23).

Proof of Theorem B.2: In the correlated case, (8) is replaced by

$$\frac{dP_t^{(\tau)}}{P_t^{(\tau)}} = \mu_t^{(\tau)}dt - A_r(\tau)\sigma_r dB_{r,t} - (A_r(\tau)\sigma_{r\beta} + A_\beta(\tau)\sigma_\beta)dB_{\beta,t}, \tag{B.12}$$

where

$$\mu_t^{(\tau)} = A_r'(\tau) r_t + A_\beta'(\tau) \beta_t + C'(\tau) + A_r(\tau)\kappa_r(r_t - \bar{r} + \gamma\beta_t) + A_\beta(\tau)\kappa_\beta\beta_t$$

$$+ \frac{1}{2} A_r(\tau)^2(\sigma_r^2 + \sigma_{r\beta}^2) + \frac{1}{2} A_\beta(\tau)^2 \sigma_\beta^2 + A_r(\tau)A_\beta(\tau)\sigma_{r\beta}\sigma_\beta. \tag{B.13}$$

The arbitrageurs' optimization problem (A.1) is replaced by

$$\max_{\{x_t^{(\tau)}\}_{\tau\in(0,T]}} \left[\int_0^T x_t^{(\tau)}(\mu_t^{(\tau)} - r_t)d\tau - \frac{a\sigma_r^2}{2}\left(\int_0^T x_t^{(\tau)}A_r(\tau)d\tau\right)^2 - \frac{a}{2}\left(\int_0^T x_t^{(\tau)}(\sigma_{r\beta}A_r(\tau) + \sigma_\beta A_\beta(\tau))d\tau\right)^2 \right]. \tag{B.14}$$

Point-wise maximization of (B.14) yields the first-order condition (10), with (11) replaced by

$$\lambda_{r,t} \equiv a\sigma_r^2 \int_0^T x_t^{(\tau)}A_r(\tau)d\tau + a\sigma_{r\beta}\int_0^T x_t^{(\tau)}(\sigma_{r\beta}A_r(\tau) + \sigma_\beta A_\beta(\tau))d\tau, \tag{B.15}$$

$$\lambda_{\beta,t} \equiv a\sigma_\beta \int_0^T x_t^{(\tau)}(\sigma_{r\beta}A_r(\tau) + \sigma_\beta A_\beta(\tau))d\tau. \tag{B.16}$$

Substituting $x_t^{(\tau)}$ from (4) and (12) into (B.15) and (B.16), we find

$$\lambda_{r,t} \equiv a\sigma_r^2 \int_0^T [\zeta(\tau) + \theta(\tau)\beta_t]A_r(\tau)d\tau + a\sigma_{r\beta}\int_0^T [\zeta(\tau) + \theta(\tau)\beta_t](\sigma_{r\beta}A_r(\tau) + \sigma_\beta A_\beta(\tau))d\tau, \tag{B.17}$$

$$\lambda_{\beta,t} \equiv a\sigma_\beta \int_0^T [\zeta(\tau) + \theta(\tau)\beta_t](\sigma_{r\beta}A_r(\tau) + \sigma_\beta A_\beta(\tau))d\tau. \tag{B.18}$$

Substituting $\mu_t^{(\tau)}$, $\lambda_{r,t}$, and $\lambda_{\beta,t}$ from (9), (B.17), and (B.18) into (10), we find an affine equation in (r_t, β_t). Identifying terms in r_t yields (A.5), identifying terms in β_t yields

$$\kappa_r \gamma A_r(\tau) + \kappa_\beta A_\beta(\tau) + A_\beta'(\tau)$$

$$= a\sigma_r^2 A_r(\tau)\int_0^T A_r(\tau)\theta(\tau)d\tau + a(\sigma_{r\beta}A_r(\tau) + \sigma_\beta A_\beta(\tau))\int_0^T (\sigma_{r\beta}A_r(\tau) + \sigma_\beta A_\beta(\tau))\theta(\tau)d\tau, \tag{B.19}$$

and identifying constant terms yields

$$C'(\tau) - \kappa_r\bar{r}A_r(\tau) + \frac{\sigma_r^2 + \sigma_{r\beta}^2}{2}A_r(\tau)^2 + \frac{\sigma_\beta^2}{2}A_\beta(\tau)^2 + \sigma_{r\beta}\sigma_\beta A_r(\tau)A_\beta(\tau)$$

$$= a\sigma_r^2 A_r(\tau)\int_0^T A_r(\tau)\zeta(\tau)d\tau + a(\sigma_{r\beta}A_r(\tau) + \sigma_\beta A_\beta(\tau))\int_0^T (\sigma_{r\beta}A_r(\tau) + \sigma_\beta A_\beta(\tau))\zeta(\tau)d\tau. \tag{B.20}$$

The solution to (A.5) is (13). Using (13) and the definition of I_r, we can write (A.6) as

$$\hat{\kappa}_\beta A_\beta(\tau) + A_\beta'(\tau) = \left[a(\sigma_r^2 + \sigma_{r\beta}^2) + a\sigma_{r\beta}\sigma_\beta\int_0^T A_\beta(\tau)\theta(\tau)d\tau - \kappa_r\gamma\right]\frac{1 - e^{-\kappa_r}}{\kappa_r}, \tag{B.21}$$

where

$$\hat{\kappa}_\beta \equiv \kappa_\beta - a\left(\sigma_{r\beta}\sigma_\beta I_r + \sigma_\beta^2 \int_0^T A_\beta(\tau)\theta(\tau)d\tau\right). \tag{B.22}$$

The solution to (A.8) is (14), with Z given by

$$Z = a(\sigma_r^2 + \sigma_{r\beta}^2) + Z\sigma_\beta\sigma_{r\beta}I_\beta - \kappa_r\gamma$$

and hence as in the theorem's statement. Substituting into (B.22), we find that $\hat{\kappa}_\beta$ is given by (B.11). The solution to (B.20) is

$$C(\tau) = Z_r \int_0^\tau A_r(\tau')d\tau' + Z_\beta \int_0^\tau A_\beta(\tau')d\tau'$$

$$- \frac{\sigma_r^2 + \sigma_{r\beta}^2}{2}\int_0^\tau A_r(\tau')^2 d\tau' - \frac{\sigma_\beta^2}{2}\int_0^\tau A_\beta(\tau')^2 d\tau' - \frac{\sigma_{r\beta}\sigma_\beta}{2}\int_0^\tau A_r(\tau')A_\beta(\tau')d\tau', \quad \text{(B.23)}$$

where

$$Z_r \equiv \kappa_r\bar{r} + a(\sigma_r^2 + \sigma_{r\beta}^2)\int_0^T \frac{1-e^{-\kappa_r}}{\kappa_r}\zeta(\tau)d\tau + Za\sigma_{r\beta}\sigma_\beta\int_0^T \frac{1}{\kappa_r}\left(\frac{1-e^{-\hat{\kappa}_\beta\tau}}{\hat{\kappa}_\beta} - \frac{e^{-\kappa_r\tau} - e^{-\hat{\kappa}_\beta\tau}}{\hat{\kappa}_\beta - \kappa_r}\right)\zeta(\tau)d\tau,$$

$$Z_\beta \equiv a\sigma_{r\beta}\sigma_\beta\int_0^T \frac{1-e^{-\kappa_r}}{\kappa_r}\zeta(\tau)d\tau + Za\sigma_\beta^2\int_0^T \frac{1}{\kappa_r}\left(\frac{1-e^{-\hat{\kappa}_\beta\tau}}{\hat{\kappa}_\beta} - \frac{e^{-\kappa_r\tau} - e^{-\hat{\kappa}_\beta\tau}}{\hat{\kappa}_\beta - \kappa_r}\right)\zeta(\tau)d\tau.$$

To complete the proof of the theorem, we must show that (B.11) has a solution for a below a threshold $\bar{a} > 0$. When κ_β goes to ∞, the left-hand side of (B.11) goes to ∞ and the right-hand side goes to the finite limit $\kappa_\beta - a\sigma_{r\beta}\sigma_\beta I_r$. A sufficient condition for (B.11) to have a solution is that (i) the left-hand side is smaller than the right-hand side for $\hat{\kappa}_\beta = 0$, and (ii) the denominator in the right-hand side is bounded away from zero for $\hat{\kappa}_\beta \in [0,\infty)$, which implies that the right-hand side is a continuous function of $\hat{\kappa}_\beta$ over $[0,\infty)$. Since (i) is satisfied for $a = 0$, it is also satisfied for a below a threshold $\bar{a}_1 > 0$. Since the function

$$\hat{\kappa}_\beta \longrightarrow \int_0^T \frac{1}{\kappa_r}\left(\frac{1-e^{-\hat{\kappa}_\beta\tau}}{\hat{\kappa}_\beta} - \frac{e^{-\kappa_r\tau} - e^{-\hat{\kappa}_\beta\tau}}{\hat{\kappa}_\beta - \kappa_r}\right)\theta(\tau)d\tau$$

converges to zero when $\hat{\kappa}_\beta$ goes to ∞, it has a finite supremum on $[0,\infty)$. Therefore, (ii) is satisfied for a below a threshold $\bar{a}_2 > 0$. ∎

B.3 Calibration

We compute maturity-weighted debt to GDP in face-value terms—that is, use our main measure of supply. Since our supply factor β_t has zero mean but our supply measure does not, we de-mean our measure. We discretize (5) and (6) with step Δt as follows:

$$\beta_{t+\Delta t} = (1 - \kappa_\beta\Delta t)\beta_t + \sigma_\beta(B_{\beta,t+\Delta t} - B_{\beta,t}), \quad \text{(B.24)}$$

$$r_{t+\Delta t} = \kappa_r\bar{r}\Delta t + (1 - \kappa_r\Delta t)r_t - \kappa_r\gamma\Delta t\beta_t + \sigma_r(B_{r,t+\Delta t} - B_{r,t}) + \sigma_{r\beta}(B_{\beta,t+\Delta t} - B_{\beta,t}), \quad \text{(B.25)}$$

and perform the vector auto-regression

$$\beta_{t+\Delta t} = c_\beta\beta_t + \epsilon_{\beta,t+\Delta t}, \quad \text{(B.26)}$$

$$r_{t+\Delta t} = c + c_r r_t + c_{r\beta}\beta_t + \epsilon_{r,\tau+\Delta t}. \quad \text{(B.27)}$$

The regression results are in Table B1.

Comparing (B.26) and (B.27) with (B.24) and (B.25), we find

$$\kappa_\beta = \frac{1 - c_\beta}{\Delta t}, \quad \text{(B.28)}$$

$$\kappa_r = \frac{1 - c_r}{\Delta t}, \quad \text{(B.29)}$$

$$\gamma = -\frac{c_{r\beta}}{\kappa_r\Delta t}, \quad \text{(B.30)}$$

Table B1
Results from the vector auto-regression (B.26) and (B.27)

	c_β	c_r	$c_{r\beta}$
Coefficient	0.99824	0.98322	–0.00025
Standard error	0.02264	0.00655	0.00019

Covariance matrix of residuals	$\epsilon_{\beta,t+\Delta t}$	$\epsilon_{r,t+\Delta t}$
$\epsilon_{\beta,t+\Delta t}$	0.003412	0.000008
$\epsilon_{r,t+\Delta t}$	0	0.000023

Table B2
Estimated parameters for the supply-factor process (5) and the short-rate process (6)

κ_β	κ_r	γ	σ_β	σ_r	$\sigma_{r\beta}$
0.021	0.201	0.015	0.202	0.017	0.001

$$\sigma_\beta = \sqrt{\frac{\text{Var}(\epsilon_{\beta,t})}{\Delta t}}, \tag{B.31}$$

$$\sigma_{r\beta} = \frac{\text{Cov}(\epsilon_{\beta,t}, \epsilon_{r,t})}{\sigma_\beta \Delta t}, \tag{B.32}$$

$$\sigma_r = \sqrt{\frac{\text{Var}(\epsilon_{r,t})}{\Delta t} - \sigma_{r\beta}^2}. \tag{B.33}$$

Substituting the results from Table B1 into (B.28)–(B.33), and setting the discretization step Δt to 1/12 because we use monthly data, we can compute $(\kappa_\beta, \kappa_r, \gamma, \sigma_\beta, \sigma_r, \sigma_{r\beta})$. The results are in Table B2.

To estimate the function $\theta(\tau)$, we express supply at any given maturity in face-value terms (as we do for maturity-weighted debt to GDP). We plot the function $\theta(\tau)$ in Figure C.2 in online Appendix C.

Proposition B.1 computes the model-implied coefficients of supply in our basic yield and return regressions. These coefficients are not identical to the effects of supply derived in Propositions 1 and 2 because (i) the short rate and the supply factor can be correlated, (ii) we are controlling for the one-year yield rather than for the instantaneous short rate, and (iii) we are regressing one-year returns rather than instantaneous returns.

Proposition B.1. The regression (19) of the τ-year yield $y_t^{(\tau)}$ on the supply factor β_t and the one-year yield $y_t^{(1)}$ produces a coefficient

$$b = \frac{A_\beta(\tau)}{\tau} - \frac{A_r(\tau)A_\beta(1)}{\tau A_r(1)} \tag{B.34}$$

on the supply factor. The regression (20) of the future one-year return $r_{t+1}^{(\tau)}$ of the τ-year bond on the supply factor β_t and the one-year yield $y_t^{(1)}$ produces a coefficient

$$b = A_\beta(\tau) - e^{-\kappa_\beta} A_\beta(\tau-1) - A_\beta(1) + \kappa_r \gamma \frac{e^{-\kappa_\beta} - e^{-\kappa_r}}{\kappa_r - \kappa_\beta} A_r(\tau-1). \tag{B.35}$$

on the supply factor.

Proof of Proposition B.1: Using (A.12), we can write (19) as

$$\frac{A_r(\tau)r_t + A_\beta(\tau)\beta_t + C(\tau)}{\tau} = a + b\beta_t + c\left(A_r(1)r_t + A_\beta(1)\beta_t + C(1)\right) + u_t. \tag{B.36}$$

The two independent variables account for all the variation in the dependent variable, and hence $u_t = 0$. Identifying terms in r_t and β_t, we find

$$\frac{A_r(\tau)}{\tau} = cA_r(1) \Rightarrow c = \frac{A_r(\tau)}{\tau A_r(1)},$$

$$\frac{A_\beta(\tau)}{\tau} = b + cA_\beta(1) \Rightarrow b = \frac{A_\beta(\tau)}{\tau} - \frac{A_r(\tau)A_\beta(1)}{\tau A_r(1)}.$$

Since returns are computed in logs,

$$r_{t+1}^{(\tau)} = \log\left(\frac{P_{t+1}^{(\tau-1)}}{P_t^{(\tau)}}\right)$$

$$= A_r(\tau)r_t + A_\beta(\tau)\beta_t + C(\tau) - \left(A_r(\tau-1)r_{t+1} + A_\beta(\tau-1)\beta_{t+1} + C(\tau-1)\right), \tag{B.37}$$

where the second step follows from (7). Using (A.12) and (B.37), we can write (20) as

$$A_r(\tau)r_t + A_\beta(\tau)\beta_t + C(\tau) - \left(A_r(\tau-1)r_{t+1} + A_\beta(\tau-1)\beta_{t+1} + C(\tau-1)\right)$$

$$= a + b\beta_t + c\left(A_r(1)r_t + A_\beta(1)\beta_t + C(1)\right) + u_{t+1}. \tag{B.38}$$

Because the processes (5) and (6) are linear, we can compute the conditional expectations of r_{t+1} and β_{t+1} as of date t by omitting the Brownian terms in (5) and (6), and solving the resulting ODEs from date t onward. The solution to these ODEs is

$$\beta_{t'} = \left(1 - e^{-\kappa_\beta(t'-t)}\right)\bar{\beta} + e^{-\kappa_\beta(t'-t)}\beta_t, \tag{B.39}$$

$$r_{t'} = \left(1 - e^{-\kappa_r(t'-t)}\right)\bar{r} + e^{-\kappa_r(t'-t)}r_t - \kappa_r\gamma\frac{e^{-\kappa_\beta(t'-t)} - e^{-\kappa_r(t'-t)}}{\kappa_r - \kappa_\beta}\beta_t, \tag{B.40}$$

for $t' > t$. Using (B.39) and (B.40), we can write (B.38) as

$$A_r(\tau)r_t + A_\beta(\tau)\beta_t + C(\tau)$$

$$- A_r(\tau-1)\left(\left(1 - e^{-\kappa_r}\right)\bar{r} + e^{-\kappa_r}r_t - \kappa_r\gamma\frac{e^{-\kappa_\beta} - e^{-\kappa_r}}{\kappa_r - \kappa_\beta}\beta_t\right)$$

$$- A_\beta(\tau-1)\left(\left(1 - e^{-\kappa_\beta}\right)\bar{\beta} + e^{-\kappa_\beta}\beta_t\right) - C(\tau-1) + v_{t+1}$$

$$= a + b\beta_t + c\left(A_r(1)r_t + A_\beta(1)\beta_t + C(1)\right) + u_{t+1}, \tag{B.41}$$

where v_{t+1} has zero conditional expectation as of date t. Subtracting (B.41) from its conditional expectation as of date t, we find $u_{t+1} = v_{t+1}$. Identifying terms in r_t, we find

$$A_r(\tau) - e^{-\kappa_r}A_r(\tau-1) = cA_r(1) \Rightarrow c = 1,$$

where the second step follows from (13). Identifying terms in β_t, we find

$$A_\beta(\tau) - e^{-\kappa_\beta}A_\beta(\tau-1) + \kappa_r\gamma\frac{e^{-\kappa_\beta} - e^{-\kappa_r}}{\kappa_r - \kappa_\beta}A_r(\tau-1) = b + cA_\beta(1)$$

$$\Rightarrow b = A_\beta(\tau) - e^{-\kappa_\beta}A_\beta(\tau-1) - A_\beta(1) + \kappa_r\gamma\frac{e^{-\kappa_\beta} - e^{-\kappa_r}}{\kappa_r - \kappa_\beta}A_r(\tau-1).$$

Therefore, the coefficients are as in the proposition's statement. ∎

References

Baker, M., R. Greenwood, and J. Wurgler. 2003. The maturity of debt issues and predictable variation in bond returns. *Journal of Financial Economics* 70:261–91.

Baker, M., and J. Wurgler. 2012. Comovement and predictability relationships between bonds and the cross-section of stocks. *Review of Asset Pricing Studies* 2:57–87.

Barro, R. 1974. Are government bonds net wealth? *Journal of Political Economy* 82:1095–117.

Bekaert, G., and R. Hodrick. 2001. Expectations hypotheses tests. *Journal of Finance* 56:1357–94.

Bekaert, G., R. Hodrick, and D. Marshall. 2001. Peso problem explanations for term structure anomalies. *Journal of Monetary Economics* 48:241–70.

Bernanke, B., V. Reinhart, and B. Sack. 2004. Monetary policy alternatives at the zero bound: An empirical assessment. *Brookings Papers on Economic Activity* 2:1–78.

Bohn, H. 2008. The sustainability of fiscal policy in the United States. In R. Neck and J. Sturm (Eds.), *Sustainability of Public Debt*, 15-49. Cambridge, MA: MIT Press.

Cochrane, J. 2008. Comments on "Bond supply and excess bond returns" by Robin Greenwood and Dimitri Vayanos. Unpublished notes, University of Chicago.

Culbertson, J. 1957. The term structure of interest rates. *Quarterly Journal of Economics* 71:485–517.

Dai, Q., and T. Philippon. 2006. Fiscal policy and the term structure of interest rates. Working Paper, New York University.

D'Amico, S., W. English, D. Lopez-Salido, and E. Nelson. 2012. The Federal Reserve's large-scale asset purchase programs: Rationale and effects. *Economic Journal* 122:F415–46.

D'Amico, S., and T. King. 2013. Flow and stock effects of large-scale Treasury purchases: Evidence on the importance of local supply. *Journal of Financial Economics* 108:425–48.

DeLong, B., A. Shleifer, L. Summers, and R.Waldmann. 1990. Noise trader risk in financial markets. *Journal of Political Economy* 98:703–38.

Doepke, M., and M. Schneider. 2006. Inflation and the redistribution of nominal wealth. *Journal of Political Economy* 114:1069–97.

Duffee, G. 1996. Idiosyncratic variation of Treasury bill yields. *Journal of Finance* 51:527–52.

Duffie, D. 2010. Presidential address: Asset price dynamics with slow-moving capital. *Journal of Finance* 65:1237–67.

Fama, E., and R. Bliss. 1987. The information in long-maturity forward rates. *American Economic Review* 77:680–92.

Fernald, J., P. Mosser, and F. Keane. 1994. Mortgage security hedging and the yield curve. *Federal Reserve Bank of New York Quarterly Review*, Summer–Fall. 92–100.

Ferson, W., and C. Harvey. 1991. The variation of economic risk premiums. *Journal of Political Economy* 99:385–415.

Fleming, M. 2002. Are larger issues more liquid? Evidence from bill reopenings. *Journal of Money, Credit and Banking* 34:707–35.

Fleming, M., and J. Rosenberg. 2007. How do Treasury dealers manage their positions? Staff Report 299, Federal Reserve Bank of New York.

Gagnon, J., M. Raskin, J. Remache, and B. Sack. 2011. The financial market effects of the Federal Reserve's large-scale asset purchases. *International Journal of Central Banking* 7:3–43.

Garbade, K. 2007. The emergence of "regular and predictable" as a Treasury debt management strategy. *Federal Reserve Bank of New York Economic Policy Review*, March. 53–71.

Garbade, K., and M. Rutherford. 2007. Buybacks in Treasury cash and debt management. Staff Report 304, Federal Reserve Bank of New York.

Greenwood, R., S. Hanson, and J. Stein. 2010. A gap-filling theory of corporate debt maturity choice. *Journal of Finance* 65:993–1028.

Greenwood, R., and D. Vayanos. 2010. Price pressure in the government bond market. *American Economic Review, Papers and Proceedings*, 585–90.

Gromb, D., and D. Vayanos. 2010. Limits of arbitrage. *Annual Review of Financial Economics* 2:251–75.

Guibaud, S., Y. Nosbusch, and D. Vayanos. 2013. Preferred habitat and the optimal maturity structure of government debt. *Review of Financial Studies* 26:1914–61.

Gurkaynak, R., B. Sack, and J. Wright. 2007. The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics* 54:2291–304.

Hamilton, J., and C. Wu. 2012. The effectiveness of alternative monetary policy tools in a zero lower bound environment. *Journal of Money, Credit, and Banking* 44:3–46.

Hanson, S. 2012. Mortgage convexity. Working Paper, Harvard Business School.

Hetzel, R., and R. Leach. 2001. The Treasury-Fed accord: A new narrative account. *Federal Reserve Bank of Richmond Economic Quarterly* 87:33–55.

Hong, H., D. Sraer, and J. Yu. 2013. Reaching for maturity. Working Paper, Princeton University.

Joyce, M., A. Lasaosa, I. Stevens, and M. Tong. 2011. The financial market impact of quantitative easing in the United Kingdom. *International Journal of Central Banking* 7:113–61.

Joyce, M., D. Myles, A. Scott, and D. Vayanos. 2012. Quantitative easing and unconventional monetary policy: An introduction. *Economic Journal* 122:F271–88.

Kambhu, J., and P. Mosser. 2001. The effect of interest rate options hedging on term-structure dynamics. *Federal Reserve Bank of New York Economic Policy Review* 7:51–70.

Koijen, R., O. Van Hemert, and S. Van Nieuwerburgh. 2009. Mortgage timing. *Journal of Financial Economics* 93:292–324.

Krishnamurthy, A., and A. Vissing-Jorgensen. 2011. The effects of quantitative easing on interest rates: Channels and implications for policy. *Brookings Papers on Economic Activity* 2:215–87.

———. 2012. The aggregate demand for Treasury debt. *Journal of Political Economy* 120:233–67.

Kuttner, K. 2006. Can Central Banks Target Bond Prices? Working Paper 12454, National Bureau of Economic Research.

Li, C., and M. Wei. 2012. Term structure modeling with supply factors and the Federal Reserve's large scale asset purchase programs. Working Paper, Board of Governors of the Federal Reserve System.

Longstaff, F. 2004. The flight-to-liquidity premium in U.S. Treasury bond prices. *Journal of Business* 77:511–526.

Lou, D., H. Yan, and J. Zhang. 2013. Anticipated and repeated shocks in liquid markets. *Review of Financial Studies* 26:1891–912.

Ludvigson, S., and S. Ng. 2009. Macro factors in bond risk premia. *Review of Financial Studies* 22:5027–67.

Malkhozov, A., P. Mueller, A. Vedolin, and G. Venter. 2013. Mortgage hedging in fixed income markets. Working Paper, McGill University.

Mankiw, G., and M. Shapiro. 1986. Do we reject too often? Small sample properties of tests of rational expectations models. *Economics Letters* 20:139–45.

Modigliani, F., and R. Sutch. 1966. Innovations in interest-rate policy. *American Economic Review* 56:178–97.

Newey, W., and K. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–8.

Ogden, J. 1987. The end of the month as a preferred habitat: A test of operational efficiency in the money market. *Journal of Financial and Quantitative Analysis* 22:329–43.

Park, S.-Y., and M. Reinganum. 1986. The puzzling price behavior of Treasury bills that mature at the end of calendar months. *Journal of Financial Economics* 16:267–83.

Politis, D., and J. Romano. 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89:1303–13.

Reinhart, V., and B. Sack. 2000. The economic consequences of disappearing government debt. *Brookings Papers on Economic Activity* 31:163–220.

Ross, M. 1966. Operation Twist': A mistaken policy? *Journal of Political Economy* 74:195–9.

Sierra, J. 2010. International capital flows and bond risk premia. Working Paper, Bank of Canada.

Simon, D. 1991. Segmentation in the Treasury bill market: Evidence from cash management bills. *Journal of Financial and Quantitative Analysis* 26:97–108.

——. 1994. Further evidence on segmentation in the Treasury bill market. *Journal of Banking and Finance* 18:139–51.

Spiegel, M. 1998. Stock price volatility in a multiple security overlapping generations model. *Review of Financial Studies* 11:419–47.

Stambaugh, R. 1986. Bias in regressions with lagged stochastic regressors. Working Paper, University of Chicago.

Swanson, E. 2011. Let's twist again: A high-frequency event-study analysis of Operation Twist and its implications for QE2. *Brookings Papers on Economic Activity* 1:151–88.

Tobin, J. 1958. Liquidity preference as behavior towards risk. *Review of Economic Studies* 25:124–31.

——. 1969. A general equilibrium approach to monetary theory. *Journal of Money, Credit, and Banking* 1:15–29.

Vayanos, D., and J.-L. Vila. 2009. A preferred-habitat model of the term structure of interest rates. Working Paper, London School of Economics.

Wallace, N. 1967. The term structure of interest rates and the maturity composition of the Federal Debt. *Journal of Finance* 22:301–12.